CONNECTIVITY BASED FEATURE-LEVEL FILTERING FOR SINGLE-TRIAL EEG BCIS

Dominic Heger, Emiliyana Terziyska, Tanja Schultz

Karlsruhe Institute of Technology (KIT), Cognitive Systems Lab

ABSTRACT

EEG-based Brain Computer interfaces (BCIs) often rely on power spectral density features to represent relevant aspects of brain activity. The information flow within human brain networks and the corresponding connectivity patterns may contain useful information to improve BCI performance, however they are typically not leveraged in current systems. In this paper, analyzes of information flow between independent sources of brain activity have been incorporated into the feature extraction stage of a BCI. For this purpose, connectivity measures based on multivariate autoregressive models have been estimated and are applied as filters to power spectral density based features. Two publicly available data sets have been used to evaluate the proposed feature extraction method: a two-back task and a motor imagery task. The results demonstrate significant performance improvements of the proposed method over band-power features and indicate that connectivity in brain networks can be used as powerful feature-level filters for BCIs.

***Index Terms*—** Connectivity, Granger causality, direct directed transfer function, brain-computer interfaces, electroencephalography

1. INTRODUCTION

The brain's neural network is a huge information processing system, in which multiple brain areas interact with each other in order to perform cognitive tasks. Recent progress in neuroengineering has shown that the pathways of information flow between different brain activity sources are of particular importance when analyzing brain function [1]. Granger Causality (GC) based connectivity measures are an established means to analyze the information flow among different activity sources in the brain. Multiple GC based connectivity measures have been proposed that can be calculated using multivariate autoregressive (MVAR) models, including partial directed coherence [2] and direct directed transfer function [3].

Non-invasive EEG-based Brain-Computer Interfaces (BCIs) measure brain activity using multiple electrodes attached to the scalp and attempt to decode the underlying patterns into machine understandable commands. Calculating the power within different frequency bands of EEG data is one of the standard approaches for feature extraction within the pattern recognition pipelines of BCIs. However, such power spectral density (PSD) based features do not explicitly take connectivity patterns into account. To incorporate connectivity information into the feature extraction of BCIs we propose to integrate connectivity estimates based on direct directed transfer functions as a linear filter at feature-level. Connectivity estimates between different activity sources are used as an importance reweighting and feature selection mechanism for PSD features. The idea of the proposed feature-level connectivity filtering is that estimates of connectivity between activity sources may help to model the flow of information that is relevant for the particular task more precisely and may therefore improve BCI performance.

Instead of using the raw EEG, we derive GC based connectivity measures from multivariate autoregressive (MVAR) models that are calculated from independent sources of the EEG using Independent Component Analysis (ICA). Calculating connectivity measures in the source space avoids spurious connectivity estimates that occur due to volume conduction effects [4, 5]. In our approach, feature-level filters based on the connectivity information are calculated from the training trials only and applied to the features for single-trial prediction of the test data. Therefore, connectivity estimates can be robustly calculated on a sufficient amount of data and feature-level filtering can be applied during the application of the BCI system by a simple linear transform.

1.1. Related Work

The idea to include connectivity patterns into feature extraction methods for BCIs has recently been addressed by a few studies:

In [6], the authors evaluated features derived from the directed transfer function (DFT) connectivity measure for motor imagery BCIs and compared them with power spectral density (PSD) and phase locking value based features. They could show significant improvements in classification accuracy for combined feature sets including DFT features and the conventional features over the conventional features alone. In [7], a nonlinear Granger Causality (GC) measure of functional connectivity was applied to a BCI for decoding different intended arm reaching movements (left, right and forward). The directional flow measure was based on nonlinear predictive models using radial basis functions. A threshold was set through a spatial statistical process so that the top 20% of pathways ranked according to their GC values are considered. The results of the study showed that directional flow patterns are distinct with respect to the intended arm movement direction. In [8], the authors considered the dynamics of communication between activity sources in the brain during cognitive processing as a control feature for BCIs. The subjects in this experiment were asked to perform or imagine finger taps. Phase synchronization was used to map inter-channel relationships. Hidden Markov models were applied to describe the dynamics of the resulting complex networks. The results show that functional connectivity dynamics can offer additional information to improve BCI classification accuracies. In [9], the authors proposed feature extraction incorporating source analysis and brain network dynamics. The data set 2 from BCI Competition IV was used as a testing set. The authors developed a Bayesian spatiotemporal model for source-localized EEG and then modeled a dynamic causality brain network. A classification accuracy of 91.25% was reached. The results show that the source analysis and brain network dynamics methods can improve BCI performance and are suitable for use in practical applications. In [10], the authors used different connectivity measures as features for a motor imagery BCI. They applied Independent Components Analysis (ICA) to the EEG data and estimated connectivity measures using multivariate autoregressive models. In their ex-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

periments, they could show that features based on effective connectivity estimated from single-trial data allow a reliable classification and that their performance is similar to using traditional band-power features. The EEGLAB-compatible toolbox SIFT [11, 12] provides a variety of techniques to analyze source connectivity dynamics of electrophysiological brain activity. It provides methods to calculate different GC based connectivity measures using multivariate autoregressive (MVAR) models.The toolbox contains functions to estimate and validate MVAR models using statistical analyzes, provides a graphical user interface and visualization of the estimated connectivity. The SIFT implementations for model fitting, validation and connectivity estimation were used for the evaluation in this paper.

This work aims to explore the applicability of effective brain connectivity patterns as a filter mechanism during the feature extraction of BCIs. In comparison to previous work, we do not use the estimated connectivity measures as features for a BCI directly, but apply them as filters to standard PSD based features. In the evaluation of two different data sets (two-back and motor imagery), we show that our approach can significantly improve BCI performance. Although connectivity estimates have been used in BCIs before, to the best of our knowledge, they have never been applied as a feature-level filtering mechanism.

2. CONNECTIVITY ESTIMATION

2.1. Granger Causality and Autoregressive Models

When determining paths of information flow within the brain (effective connectivity), Granger Causality (GC) is one of the most popular and important tools. It was introduced by Clive Granger [13] and originally applied to financial data analyzes. A time series X Granger-causes another time series Y, if it can be shown that the values of X provide statistically significant information about future values of Y. GC has become a widely used technique in neuroscientific research, where it has often been used to assess causal connectivity relations in brain activity, for example [14, 15, 16].

GC can be calculated using autoregressive modeling. The basic GC is only applicable to bivariate time series. In multivariable situations the connections between time series can be direct or indirect. Repeated bivariate analysis can lead to the conclusion that one time series is causal to another time series, when in fact both were influenced by a third one but with different time delays. This problem was addressed by Geweke, who introduced conditional GC [17]. Compared to pairwise GC that is based on bivariate autoregressive models, conditional GC is based on multivariate autoregressive (MVAR) models, which are suitable to express dependencies between the individual signals considering all channels together.

Given $V = \{1,2,...,d\}$ time series (e.g. activity from d relevant sources in the brain) and model order p, a prediction of the vector x_t by the MVAR model is the linear combination of the p previous values

$$x_t^{(V)} = \sum_{k=1}^{p} A_k x_{t-k}^{(V)} + u_t, \tag{1}$$

where $x_t^{(V)}$=[$x_t(1)$, $x_t(2)$, ..., x_t(d)] is the t-th sample of the time series and A_k is a d-by-d matrix of the MVAR model coefficients (weights) at time lag k and u_t=[$u_t(1)$, $u_t(2)$, ..., u_t(d)] is the zero-mean residual.

2.2. MVAR model based connectivity measures

Based on equation (1), we can assess GC based connectivity measures in the frequency domain. If we transform the MVAR model to the z-domain, we obtain

$$X(f) = A(f)^{-1}U(f). \tag{2}$$

with a transfer function of the system $H(f) = A^{-1}(f)$. The spectral density $S(f)$ is given by

$$S(f) = H(f)\Sigma H^*(f). \tag{3}$$

$H^*(f)$ is the transposed complex conjugate of $H(f)$ and Σ is the covariance matrix of the noise $U(f)$.

Given the quantities $A(f)$, $H(f)$, $S(f)$ and Σ multiple connectivity measures can be calculated [18]. The most traditional approach for detecting cooperative neuronal activity in electrophysiological signals is coherence, which describes the linear relationship in the frequency domain between channels [19]. The ordinary coherence with frequency f is defined as

$$C_{ij}(f) = \frac{|S_{ij}(f)|}{\sqrt[2]{S_{ii}(f)S_{jj}(f)}}, \tag{4}$$

where the values $S_{ij}(f)$ represent the cross-spectrum between channels i and j at frequency f (autospectra for i=j).

When using non-invasive EEG measurements, volume conduction effects can not be avoided. Partialing out the influence of other channels is an option to mitigate volume conduction effects [20]. The partial coherence (pCOH) between i and j represents the residual coherence, where the common combinations to any other channels are removed [21]:

$$pCOH_{ij}(f) = \frac{S_{ij}^{-1}(f)}{\sqrt[2]{S_{ii}^{-1}(f)S_{jj}^{-1}(f)}}. \tag{5}$$

A connectivity measure that reveals direct and indirect directional connections is the directed transfer function (DTF) [3]:

$$DTF_{ij}(f) = \frac{H_{ij}(f)}{\sqrt[2]{\sum_{k=1}^{n}|H_{ik}(f)|^2}}. \tag{6}$$

One extension of DTF is the full frequency directed transfer function (ffDTF), in which the summation over the whole frequency range removes the dependence of the denominator on frequency [22]:

$$ffDTF_{ij}(f) = \frac{|H_{ij}(f)|^2}{\sum_f \sum_{k=1}^{n}|H_{ik}(f)|^2}. \tag{7}$$

Another extension of DTF is dDTF, which describes only direct connections. It results from multiplying $pCOH_{ij}(f)$, as defined in equation (5), by $ffDTF_{ij}(f)$, as defined in equation (7):

$$dDTF_{ij}(f) = pCOH_{ij}(f) \cdot ffDTF_{ij}(f). \tag{8}$$

We decided to use the direct Directed Transfer Function (dDTF) as connectivity measure for the evaluations in this paper, as preliminary tests showed that extensions of DTF had consistently superior performance than multiple other connectivity measures when used within the feature extraction of BCIs. This is also supported by the findings of Billinger and collaborators [10].

3. DATA CORPORA

Two publicly available data sets have been used to evaluate the proposed method for connectivity based feature-level filtering:

1. **2BACK**: This data set is provided by the Swartz Center for Computational Neuroscience[1] along with the sample material of the SIFT Toolbox [11]. It consists of EEG data of one subject performing a n-back task with feedback (n=2), where a

[1] http://sccn.ucsd.edu/wiki/SIFT

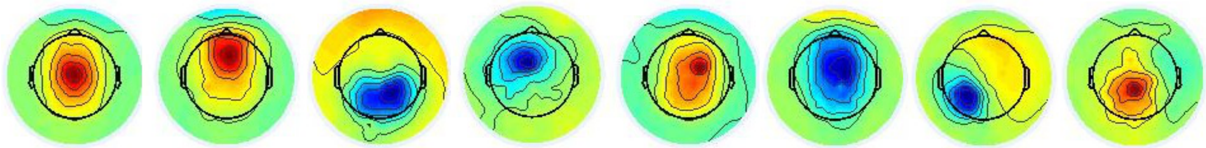

Fig. 1. Topographical scalp maps of the eight selected ICA components for the **2BACK** data set.

continuous stream of letters, separated by ~1500 ms, is presented. The subject is instructed to press a button with the right thumb if the current letter matches the one presented twice earlier in the sequence and press with the left thumb if the letter is not a match [18]. Correct and erroneous responses are indicated by two different kinds of auditory feedback. The data consists of 123 correct and 123 error trials. EEG data was recorded using 64 channels at a sampling rate of 256 Hz. In the evaluations in section 5, we will discriminate trials with correct responses from trials with incorrect responses.

2. **BCI3IVa**: This data set is provided by the Berlin BCI group and is available as data set IVa from the BCI Competition 3 website[2]. It contains EEG data recorded from 5 subjects performing motor imagery of 2 different classes: right hand, right foot [23]. A particular challenge of this competition is that for some subjects only few training data is available. The data has been recorded using 118 EEG channels and was downsampled to 100 Hz. In the evaluations in section 5, we discriminate the different classes of motor imagery and apply same conditions as in the competition.

4. SIGNAL PROCESSING AND FEATURE EXTRACTION

The following section describes the signal processing and feature extraction pipeline used for the evaluation of the two data sets. Fig. 2 gives an overview of the processing steps involved.

4.1. Pre-processing of the data

The two data sets described in section 3 were bandpass filtered to broadly cover the relevant EEG activity. The **2BACK** data set was filtered between 2 Hz and 50 Hz, **BCI3IVa** between 8 Hz and 30 Hz. An Extended Infomax Independent Component Analysis (ICA) [12] was calculated to transform the signals into sources space. ICA components were selected in order to remove artifacts and irrelevant brain activity, but also to reduce the dimensionality of the MVAR models. For **2BACK**, 8 ICA components were selected that are suggested in [18] based on neurophysiological expert knowledge. Fig. 1 shows the corresponding 2D topographical scalp maps of the selected components. For **BCI3IVa** the ADJUST toolbox [24] was used to identify ICA components automatically that are artifact free. Between $d = 79$ and $d = 116$ of the 118 components were selected for the different subjects.

4.2. MVAR model fitting and validation

For analyzing multivariate causality and information flow between sources of EEG activity we estimated multivariate autoregressive (MVAR) models using the Vieira-Morf algorithm [25] on the training trials data. The suitability of the fitted model has to be proven using statistical testing for model consistency and stability. Therefore, we applied the percent consistency tests [26] that calculate the

[2]http://www.bbci.de/competition/iii/

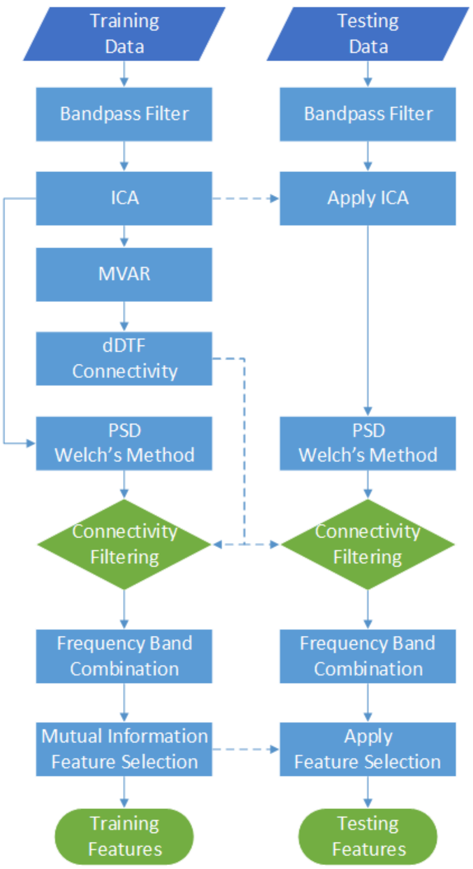

Fig. 2. Processing steps for signal processing and feature extraction including feature-level connectivity filtering.

amount of correlation structure of the original data that is mapped by the autoregressive model. In addition to that, stability tests ensure the stationarity of the model, i.e. to ensure that the model will not diverge to infinity [18]. Stable MVAR models with high percent consistency around 95% were built upon **2BACK** using a model order of $p = 15$. For **BCI3IVa**, the validation tests confirmed the stability of the MVAR models for each subject and indicated percent consistencies of 62.4% on average using model order $p = 20$.

4.3. Feature-level connectivity filtering

Using the fitted autoregressive models (see 4.2), Granger Causality based connectivity measures between the activity sources can be estimated. For the proposed approach we calculated direct Directed Transfer Fuction (dDTF) according to equation (8). For each frequency band f the dDTF based connectivity estimate between ICA sources i and j can be represented in a matrix $\mathbf{K}$ with elements $dDTF_{ij}(f)$. To emphasize the influence of features that are involved in strong connectivity links, all but the k largest elements in $\mathbf{K}$ were set to zero and the values were multiplied with quadratically decreasing weighting coefficients.

For each trail t, power spectral density (PSD) feature vectors $\mathbf{p_t}$ with elements $p_t^j(f)$ were extracted using Welch's method [27] (activity sources $j \in \{1,\ldots,d\}$). A logarithm was applied to $\mathbf{p_t}$ to make the features approximately Gaussian.

Feature-level connectifity filtering was performed by the matrix-vector product of the connectivity matrix $\mathbf{K}$ and the features $\mathbf{p_t}$:

$$\hat{\mathbf{p}}_\mathbf{t}(f) = \mathbf{K} \cdot \mathbf{p_t} = \sum_{j=1}^{d} dDTF_{nj}(f) \cdot p_t^j(f). \tag{9}$$

Hence, the n-th coefficient of the resulting vector $\hat{\mathbf{p}}_\mathbf{t}(f)$ can be interpreted as the weighted sum of the information flowing to activity source n from the other sources

4.4. Dimensionality reduction

EEG-based BCIs often use the power in different frequency bands as features. Besides the classical frequency bands, such as $\delta, \theta, \alpha, \beta$, smaller (usually 2-4 Hz wide) consecutive frequency bands are typically chosen, which can precisely model the relevant EEG rhythms and can be estimated robustly from few seconds of EEG data. To represent frequency bands in the feature vectors, we merged coefficients in $\hat{\mathbf{p}}_\mathbf{t}(f)$ that correspond to adjacent frequencies by averaging.

To further reduce the dimensionality of the feature space, feature selection based on mutual information was performed (c.f. [28]). We selected the l features that contained most information with the ground truth class distribution in the training data. For this purpose, we calculated the mutual information between the continuous training features and the corresponding discrete class labels estimated by kernel density estimation. The number of features l was selected to include more than 50% of the total mutual information between the features and the class labels. For **2BACK**, $l = 30$ features were chosen. For all subjects of **BCI3IVa**, $l = 7$ features were chosen.

5. EVALUATION AND RESULTS

To evaluate the proposed feature-level filtering approach, we evaluated the two data sets using a Naïve-Bayes classifier based on kernel density estimation (Parzen windows). The classifier and the relevant parameters were estimated on the training data and evaluated on the test data as indicated by Fig. 2.

5.1. 2BACK data set

A 10x10-fold cross-validation (repeated pseudo-random sampling) was performed to estimate the classification rates for **2BACK**. The results of the proposed feature extraction method were compared with PSD based features without feature-level connectivity filtering as baseline. Fig. 3 shows the classification accuracies for both approaches. The error bars indicate average standard deviations across the 10 cross-validation folds. The average recognition accuracy using connectivity based feature-level filtering was 64.8%, which compares to 59.7% for the baseline performance without filtering (otherwise identical processing and parameters). A paired t-test on the 10 cross-validation runs shows that incorporating connectivity estimates as feature-level filters results in significantly better performance ($p<0.0001$, $t=4.21$, $df=99$).

5.2. BCI3IVa data set

For the evaluation of **BCI3IVa** training and testing was performed comparable to BCI Competition III [23]. Please note that our aim here was not to tune the system towards maximum performance in the competition, but to compare the proposed method with a generic BCI pipeline that is applicable to many BCI problems and to evaluate it on a publicly available data set. The results of the proposed feature-level filtering approach are presented in Fig. 4. The average recognition rate over the five subjects was 79.5%. The baseline performance for PSD based features without feature-level connectivity filtering was 73.7%. Comparing the classification rates of the proposed feature-level filtering with the corresponding ones based only on PSD based features (otherwise identical processing and parameters) shows significantly higher performance for the proposed approach (paired t-test, $p<0.05$, $t=2.97$, $df=4$). The results of both approaches are shown as confusion matrices in Table 1.

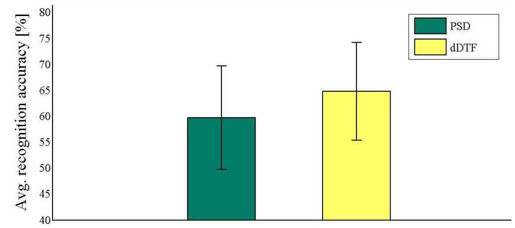

Fig. 3. Comparison of the classification accuracies of **2BACK** for features using connectivity based feature-level filtering (dDTF) and power spectral density based features (PSD).

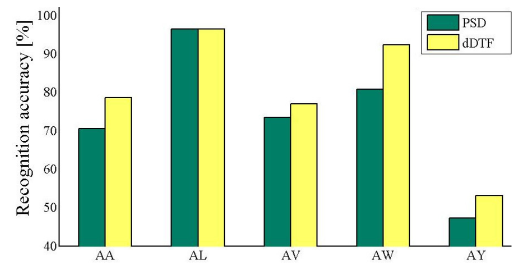

Fig. 4. Comparison of the classification accuracies of the 5 subject in **BCI3IVa** for features using connectivity based feature-level filtering (dDTF) and power spectral density based features (PSD).

		dDTF actual		**PSD** actual	
		Foot	**Hand**	**Foot**	**Hand**
predicted	**Foot**	**85.4%**	14.6%	**73.4%**	26.6%
	Hand	34.4%	**65.6%**	36.0%	**64.0%**

Table 1. Confusion matrices of the **BCI3IVa** data set using dDTF based feature-level connectivity filtering (left) and baseline power spectral density based features (right) averaged over all subjects.

6. CONCLUSION

In this paper, we proposed a feature extraction method for BCI applications combining connectivity measures based on dDTF as feature-level filters for power spectral density features. The approach was evaluated using data sets of two different publicly available experiments: a two-back task and a two-class motor imagery task. For both data sets, the new approach strongly improved classification rates over the baseline, which shows that dDTF based connectivity measures can successfully be used to select and reweight band-power features for BCIs.

7. REFERENCES

[1] Olaf Sporns, "The human connectome: a complex network," *Annals of the New York Academy of Sciences*, vol. 1224, no. 1, pp. 109–125, 2011.

[2] Luiz A Baccalá and Koichi Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.

[3] MJ Kaminski and KJ Blinowska, "A new method of the description of the information flow in the brain structures," *Biological cybernetics*, vol. 65, no. 3, pp. 203–210, 1991.

[4] Scott Makeig, Anthony J Bell, Tzyy-Ping Jung, Terrence J Sejnowski, et al., "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems*, pp. 145–151, 1996.

[5] Stefan Haufe, Vadim V Nikulin, Klaus-Robert Müller, and Guido Nolte, "A critical assessment of connectivity measures for eeg data: a simulation study," *Neuroimage*, 2012.

[6] J. H. Lim, H. J. Hwang, Y. L. Jung, and C. H. Im, "Feature extraction for brain-computer interface (bci) based on the functional causality analysis of brain signals," in *5th Int. BCI Conference*, 2011, pp. 36–39.

[7] MengTing Liu, Ching-Chang Kuo, and Alan WL Chiu, "Statistical threshold for nonlinear granger causality in motor intention analysis," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 5036–5039.

[8] Ian Daly, Slawomir J Nasuto, and Kevin Warwick, "Brain computer interface control via functional connectivity dynamics," *Pattern recognition*, vol. 45, no. 6, pp. 2123–2136, 2012.

[9] Dongwei Chen, Jie Xiang, Hao Guo, Hongxia Deng, Haifang Li, and Junjie Chen, "Motor imagery classification with source analysis and network dynamics for brain-computer interface," 2012.

[10] Martin Billinger, Clemens Brunner, and Gernot R Müller-Putz, "Single-trial connectivity estimation for classification of motor imagery data," *Journal of neural engineering*, vol. 10, no. 4, pp. 046006, 2013.

[11] Tim Mullen, Arnaud Delorme, Christian Kothe, and Scott Makeig, "An electrophysiological information flow toolbox for eeglab," *Biol. Cybern*, vol. 83, pp. 35–45, 2010.

[12] Arnaud Delorme and Scott Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[13] Clive WJ Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.

[14] Wei Liao, Dante Mantini, Zhiqiang Zhang, Zhengyong Pan, Jurong Ding, Qiyong Gong, Yihong Yang, and Huafu Chen, "Evaluating the effective connectivity of resting state networks using conditional granger causality," *Biological cybernetics*, vol. 102, no. 1, pp. 57–69, 2010.

[15] Alard Roebroeck, Elia Formisano, Rainer Goebel, et al., "Mapping directed influence over the brain using granger causality and fmri," *Neuroimage*, vol. 25, no. 1, pp. 230–242, 2005.

[16] Xue Wang, Yonghong Chen, Steven L Bressler, and Mingzhou Ding, "Granger causality between multiple interdependent neurobiological time series: blockwise versus pairwise methods," *International journal of neural systems*, vol. 17, no. 02, pp. 71–78, 2007.

[17] John F Geweke, "Measures of conditional linear dependence and feedback between time series," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 907–915, 1984.

[18] Tim Mullen, *Source Information Flow Toolbox (SIFT) Theoretical Handbook and User Manual*, Swartz Center for Computational Neuroscience, Institute for Neural Computation, and Department of Cognitive Science University of California, San Diego, release 0.9.7a edition, December 2010.

[19] Guido Nolte, Ou Bai, Lewis Wheaton, Zoltan Mari, Sherry Vorbach, Mark Hallett, et al., "Identifying true brain interaction from eeg data using the imaginary part of coherency," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.

[20] Alois Schlögl and Gernot Supp, "Analzying event-related eeg data with multivariate autoregressive parameters," *Progress in Brain Research*, vol. 159, 2006.

[21] JR Rosenberg, DM Halliday, P Breeze, and BA Conway, "Identification of patterns of neuronal connectivity, partial spectra, partial coherence, and neuronal interactions," *Journal of Neuroscience Methods*, vol. 83, no. 1, pp. 57–72, 1998.

[22] Anna Korzeniewska, Małgorzata Mańczak, Maciej Kamiński, Katarzyna J Blinowska, and Stefan Kasicki, "Determination of information flow direction among brain structures by a modified directed transfer function (ddtf) method," *Journal of neuroscience methods*, vol. 125, no. 1, pp. 195–207, 2003.

[23] Benjamin Blankertz, K-R Müller, Dean J Krusienski, Gerwin Schalk, Jonathan R Wolpaw, Alois Schlögl, Gert Pfurtscheller, Jd R Millan, M Schröder, and Niels Birbaumer, "The bci competition iii: Validating alternative approaches to actual bci problems," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 153–159, 2006.

[24] Andrea Mognon, Jorge Jovicich, Lorenzo Bruzzone, and Marco Buiatti, "Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.

[25] Martin Morf, Augusto Vieira, Daniel TL Lee, and Thomas Kailath, "Recursive multichannel maximum entropy spectral estimation," *Geoscience Electronics, IEEE Transactions on*, vol. 16, no. 2, pp. 85–94, 1978.

[26] Mingzhou Ding, Steven L Bressler, Weiming Yang, and Hualou Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment," *Biological cybernetics*, vol. 83, no. 1, pp. 35–45, 2000.

[27] Peter Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *Audio and Electroacoustics, IEEE Transactions on*, vol. 15, no. 2, pp. 70–73, 1967.

[28] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan, "Mutual information-based selection of optimal spatial–temporal patterns for single-trial eeg-based bcis," *Pattern Recognition*, vol. 45, no. 6, pp. 2137–2144, 2012.