On the Detection of Direct Directed Information Flow in fMRI

Wolfgang Mader, David Feess, Rüdiger Lange, Dorothee Saur, Volkmar Glauche, Cornelius Weiller, Jens Timmer, and Björn Schelter

Abstract—To infer interactions from functional magnetic resonance imaging (fMRI) data, structural equation modeling (SEM) as well as dynamic causal modeling (DCM) has been suggested. Directed partial correlation (dPC) is a measure which detects Granger causality in multivariate systems. To demonstrate the strengths as well as the limitations of directed partial correlation we first applied it to simulated data tailored to the problem at hand. Second, after dPC has proven to be usefull for fMRI data analysis, we applied it to actual fMRI data.

Index Terms—Directed partial correlation, fMRI, Granger causality, instantaneous interactions, VAR-processes.

I. INTRODUCTION

INFERENCE of the brain network structure while performing specific tasks presents a hot topic in neuroscience research. The network structure contains information about the functioning or dysfunctioning of the brain. Several researchers, therefore, address the challenge to infer the network structure from measured signals. While concepts have originally been developed for electroencephalography or magnetoencephalography data, recently particular emphasis has been laid on functional magnetic-resonance imaging (fMRI) data. For the latter, the temporal resolution has been increased tremendously during the past years although it is still much lower than for electroencephalography or magnetoencephalography data. The spatial resolution is, however, a unique advantage of fMRI data.

The comparably low temporal resolution of fMRI data poses challenges to the applied analysis techniques. Moreover,

W. Mader and D. Feess are with the Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany, and theDepartment for Neurology, University Medical Center of Freiburg, Freiburg, Germany (e-mail: wolfgang.maders@fdm.uni-freiburg.de; david.feess@fdm.uni-freiburg.de).

R. Lange, D. Saur, V. Glauche, and C. Weiller are with the Department for Neurology, University Medical Center of Freiburg, Freiburg, Germany.

J. Timmer is with Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany, and the Institute of Physics, University of Freiburg, Freiburg, Germany, and the Freiburg Institute for Advanced Studies, 79104 Freiburg, Germany.

B. Schelter is with Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany, and the Institute of Physics, University of Freiburg, Freiburg, Germany.

neuronal activity is not observed directly but through the blood oxygen level dependency (BOLD). Concepts like dynamic causal modeling are based on certain prior assumptions about the investigated interaction structure [1]. We concentrate on the analysis of the fMRI signals directly without trying to access the underlying neuronal activity.

Although dealing with the BOLD signals directly, analyzing multivariate time series as obtained from fMRI still poses certain challenges to the techniques applied. Inference of the interaction structure implies that direct and indirect connections should be distinguished by the analysis technique. A bivariate analysis of pairwise interactions is therefore not expected to be able to provide a reasonable estimate of the underlying interaction structure. To avoid false positive conclusions about the interaction structure, multivariate techniques that are able to distinguish direct and indirect interdependencies should be applied as already done in other fields of research [2]–[4]. Another challenge which should be taken care of by the analysis techniques is the direction of information flow, which might be of particular interest. To this aim, Bayesian networks have been suggested [5], which allow detection of directed acyclic graphs. Since the presence of cycles in the human brain cannot be ruled out in the first place, Bayesian network models providing directed acyclic graphs likely produce false positive or negative results in this case. Recently, [6] suggested an extension of Baysian networks, the so-called dynamic Bayesian networks that allow for certain variations of interactions with time. Using the temporal information, cycles in the graphs can be revealed indirectly.

Analysis techniques which are supposed to detect the direction of information flow and, thus, causality usually base on Granger causality [7], [8]. Granger causality in turn is based on the common sense conception that causes do need to precede their effects in time. Vector autoregressive modeling is commonly used to implement Granger causality. A VAR process aims to explain as much as possible of its current state-vector from its own past state-vectors.

One might speculate whether or not the temporal information contained in fMRI data is too poor to allow for the inference of causal interdependencies between processes. First, this does not hold true when it comes to analyze the contact point of an external stimulus on the network under investigation, because the BOLD response is delayed for approximately 2 s with respect to the onset time of the stimulus, and this time is resolveable by fMRI. As shown in this manuscript, the temporal information is crucial in this case. Second, because a VAR process uses the knowledge of its own past to explain its current state the covariance matrix of the driving noise reflects the instantaneous interactions which can not be explained by the own past. This

covariance matrix is much more accurate than the information provided by pure correlation approaches like partial correlation [9].

The investigation of directed direct interaction structures in the human brain by directed partial correlation utilizes the temporal signals of certain hotspots, so-called seeds, which have to be identified for instance using the statistical parametric mapping (SPM) package first [1], [10]. Additional *a priori* assumptions are not necessary.

The mathematical background of dPC is described in [11]. This paper also introduces dPC for analyzing fMRI data, but the performance of dPC is only demonstrated on data simulated using a VAR[2] process. The manuscript at hand aims to close this gap by presenting a simulation study with simulated data which is designed to be closer to actual data.

After introducing the mathematical concept underlying directed partial correlation, we present its abilities and limitations first in an application to data simulated by an autoregressive model and second in an application to simulated fMRI signals. In an application to a language comprehension task we investigate functional interactions in a real-world experiment.

II. MATHEMATICAL BACKGROUND

In this section, we briefly describe the analysis techniques used. For a more in detail description see [11].

A. Granger Causality and Autoregressive Modeling

When analyzing real-world data, one is often interested in inference of the underlying network structure. Bivariate measures are not capable of providing the true interaction structure as they cannot distinguish direct and indirect interactions [12]. Moreover, when it comes to analyze the direction of information transfer leading to directed networks, a technique is mandatory that is capable of providing the direction of information flow.

Commonly the direction of information flow is based on the notion of causality. Granger [7], [8] introduced his concept of causality based on the common sense conception that causes necessarily precede their effects in time. In terms of predictability this leads to the following definition: a process X_j is Granger-noncausal to process X_i if the knowledge of the past of X_j does not improve the information about X_i. This idea can be readily extended to the multivariate case. In an n-dimensional multivariate system X_V $V = \{1, 2, \ldots n\}$ the process X_j is called Granger-noncausal for X_i if the knowledge about X_i based on the past of $X_{V|j}$, i.e. all processes but j, is the same as if the past of the entire X_V is used.

The concept of Granger-causality or Granger-noncausality well transfers into the notion of vector-autoregressive processes (VAR[p])

$$\mathbf{X}(t) = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}(t) = \sum_{j=1}^{p} \mathbf{A}(j)\mathbf{X}(t-j) + \varepsilon(t). \quad (1)$$

Here, the matrix $\mathbf{A}(j)$ is the j-th $n \times n$ coefficient matrix and $\varepsilon(t)$ is the n-dimensional white noise term with covariance matrix Σ.
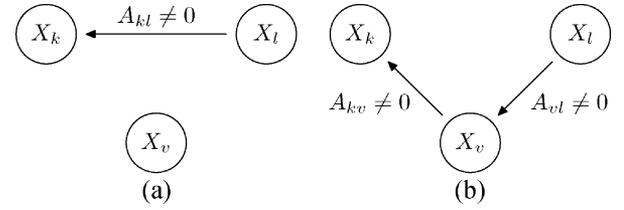

Fig. 1. (a) Shows a direct connection between X_l and X_k, (b) an indirect connection between X_l and X_k. It is important to distinguish direct and indirect connections to avoid false positive conclusions.

Imagine now that certain off-diagonal entries of the matrices $\mathbf{A}(j)$ are nonzero, say $A_{kl}(j) \neq 0$. Then, process X_l is Granger-causal for X_k given all the information from the past of X_V. Furthermore, it makes a difference if $A_{kl}(j) \neq 0$ (Fig. 1(a)) or if, both $A_{kv}(j) \neq 0$ and $A_{vl}(j) \neq 0$ and $A_{kl}(j) = 0$ [Fig. 1(b)]. The first refers to a direct directed interaction while the latter illustrates an indirect directed interaction from process X_l onto X_k as the entire information is mediated by process X_v. Conclusively, the coefficient matrices contain the information of direct directed signal transfer in multivariate autoregressive systems. If an indirect interaction is present for instance between process X_l and X_k mediated by process X_v, process X_l still influences process X_k, but if process X_v ceased to exist, the interaction between X_l and X_k would also cease to exist.

A nonzero entry in the covariance matrix Σ of the noise term $\varepsilon(t)$ also correlates processes. By this, two processes can be correlated even if there is no directed, causal connection. Since the influence is instantaneous in time, this correlation should be referred to as *Instantaneous Interaction* although it is sometimes misleadingly called *Instantaneous Causality* [13].

In cases where measured signals are to be analyzed, vector autoregressive models should be fitted to the raw data. For this purpose, different estimation techniques are known [14]. For nonlinear systems it is not clear in the first place if autoregressive modeling is a reasonable approach. However, extensive simulation studies have shown that autoregressive modeling and the concept of Granger-causality works well also in nonlinear systems [12], [15].

B. Measures Based on Autoregressive Modeling

Several measures based on Granger causality applying autoregressive modeling are conceivable. The most naïve one would be to check the specific entries of the matrix $A_{kl}(j)$ for all j. If for one of those j entries A_{kl} is significantly different from zero, one would claim a Granger-causal influence from process X_l onto process X_k. This approach, however, is inferior to other approaches as it does not provide an intuitive way to present the results for high autoregressive model orders p.

Recently, a time domain measure called directed partial correlation has been suggested by [11], [13]. The advantage of this technique is that it treats both the Granger-causal influences as well as instantaneous interactions. Roughly, directed partial correlation presents the multivariate cross-correlation function conditioning on third processes without the disadvantages of correlated errors as known for the cross-correlation analysis.

Mathematically, directed partial correlation π_{ij} is defined for the time lag τ by

$$\pi_{ij}(\tau) = \frac{A_{ij}(\tau)}{\sqrt{\Sigma_{ii}\rho_{ij}(\tau)}} \qquad (2)$$

with

$$\rho_{ij}(\tau) = K_{ij} + \sum_{v=1}^{\tau-1}\sum_{k,l\in V} A_{kj}(v)K_{kl}A_{lj}(v) + \frac{A_{ij}^2(\tau)}{\Sigma_{ii}}. \qquad (3)$$

Here, the matrix $\mathbf{K}$ is the inverse of the covariance matrix of the noise Σ, $\mathbf{K} = \Sigma^{-1}$. For $\tau = 0$, we do have

$$\pi_{ij}(0) = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}, \qquad (4)$$

which is the cross-correlation of the noise ε. In contrast to ordinary cross-correlation analysis, the directed partial correlation at time lag zero contains the entries of the covariance matrix of the driving noise in the autoregressive model. Thus, $\pi_{ij}(0)$ describes the direct interactions that cannot be explained by the past of the multivariate process.

When applying this to real-world data the estimates of the entries of coefficient matrices $\hat{A}_{kl}(j)$ of $A_{kl}(j)$ and the estimates of the covariance matrix entries $\hat{\Sigma}_{kl}$ of Σ_{kl} replace their true values in the above equations. For estimation procedures, please refer to [14], as described above.

The statistics for directed partial correlation is also known [13]. This allows a rigorous assessment of statistical significance of interactions at certain time lags. For time lag zero the instantaneous direct correlations are provided.

To demonstrate the abilities and limitations of this technique, a simulation study is performed in the following two sections before it is applied to real-world data.

III. SIMULATION STUDY

To illustrate the performance of directed partial correlation as introduced above, we apply it to simulated data of a vector autoregressive model first. Since the technique has been developed for autoregressive models, directed partial correlation is expected to detect the true underlying network structure. Afterwards it is also applied to simulated signals obtained from a model which is supposed to be a model for fMRI signals. This is an important step to show that dPC can meaningfully be applied to actual fMRI data even though this data surely violates the assumptions made by dPC.

A. Autoregressive Model

An autoregressive process

$$\mathbf{X}(t) = \sum_{j=1}^{2} \mathbf{A}(j)\mathbf{X}(t-j) + \varepsilon(t) \qquad (5)$$

of order 2 has been simulated for 256 data points. The following entries
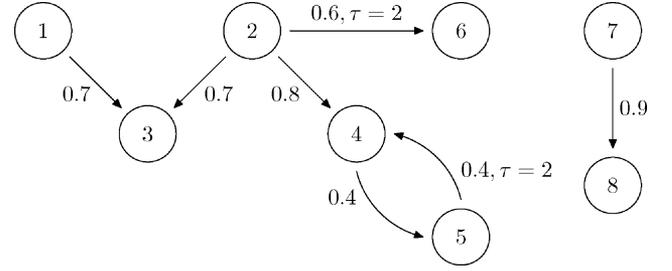
$$A_{11}(1) = \ldots = A_{88}(1) = 0.4$$

Fig. 2. Graph summarizing the simulated network for VAR data generation. The connections not labeled with τ are at time lag $\tau = 1$.

$$
\begin{aligned}
A_{31}(1) &= 0.7, & A_{32}(1) &= 0.7 \\
A_{42}(1) &= 0.8, & A_{54}(1) &= 0.4 \\
A_{87}(1) &= 0.9, & & \\
A_{62}(2) &= 0.6, & A_{45}(2) &= 0.4
\end{aligned} \qquad (6)
$$

of the matrices

$$\mathbf{A}(\tau) = \begin{pmatrix} A_{11}(\tau) & \ldots & A_{1n}(\tau) \\ \vdots & \ddots & \vdots \\ A_{n1}(\tau) & \ldots & A_{nn}(\tau) \end{pmatrix} \qquad (7)$$

have been chosen nonequal to zero.

The resulting interaction structure is shown in Fig. 2. The coefficient quantifying the interaction and the corresponding time lag is attached to the edges between the nodes. An edge between two nodes, thus, indicates a direct interaction between the corresponding nodes. The network is designed to include indirect interactions, loops, processes that influence many other processes, and independent sub-networks. This should represent typical scenarios one has to face in real-world applications.

B. Result for the Autoregressive Model

First, a VAR[2] has been fitted to the data. This is supposed to give optimal results as the simulated process is a VAR[2] as well. Fig. 3 shows the correlations as well as the directed partial correlations between the processes estimated from the VAR model that has been fitted to the data. The upper triangular part of Fig. 3 shows the bivariate correlations for each combination of the eight processes. The lower triangular part shows the result of the multivariate directed partial correlation analysis.

In Fig. 3 on the abscissa, time lags are shown and on the ordinate the correlation coefficients. The vertical dashed line in every subplot marks the time lag $\tau = 0$. The bar presented there quantifies the instantaneous interaction caused by non-zero off-diagonal entries in the covariance matrix Σ of the driving noise ε. The bars on the right side of this line correspond to information which is transferred from processes with lower number to processes with higher numbers with increasing time lag; the bars on the left side denote the opposite direction. In other words, positive time lags denote an interaction from the process in the i-th column to the j-th row of the figure, while the negative lags correspond to an influence from the process in the j-th row to the i-th column. The horizontal dotted line in every diagram is the 95% significance level. It varies with the time lag under investigation.
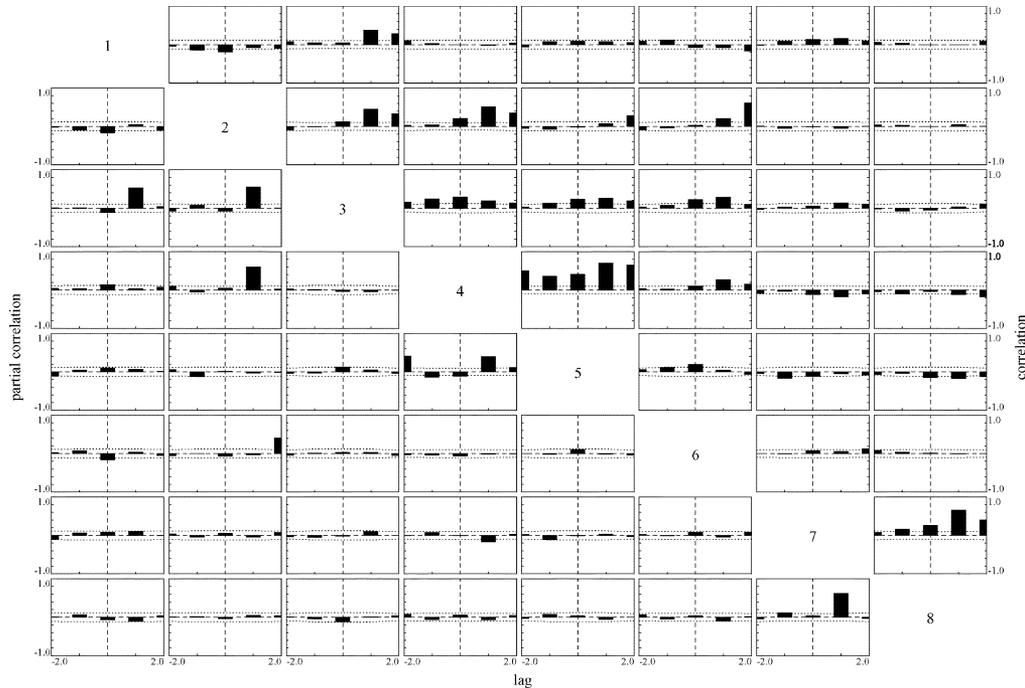
Fig. 3. Result for data simulated with a VAR[2] model and without observational noise. The order of the fitted $\mathrm{VAR}[p]$ also was $p = 2$. This is expected to give the best results. The lower triangular part of the plot show the directed partial correlation values, while the upper triangular part shows the estimates for the bivariate correlation analysis. Concerning the directed partial correlation, bars at positive time lags, right from the vertical dashed line, denote an interaction from the process in the ith column to the jth row of the figure, while the negative lags correspond to an influence from the process in the jth row to the ith column. The dotted horizontal line marks the 95% significance level.

The bivariate correlation analysis leads to false positive conclusions about the interaction between processes if the information is transmitted by a third process. This is demonstrated by the processes 2 and 5. Process 5 is only connected with process 2 via process 4 (see Fig. 2). The bivariate analysis shows an influence from process 2 to 5 with a time lag $\tau = 2$, which is caused by the mediated influence through process 4. This false positive conclusion is, however, prevented by the multivariate analysis as the corresponding directed partial correlation is compatible with zero.

Summarizing, the results from the directed partial correlation analysis lead to the true underlying network structure shown in Fig. 2.

To show the ability of directed partial correlation when the assumptions are slightly violated, we simulated the more realistic setting from which the results are displayed in Fig. 4. The simulated processes for this analysis of again 256 data points each have been contaminated with observational noise. The signal-to-noise ratios vary between 1:1 and 3:1 for the individual processes. The directed partial correlations gathered in Fig. 4 have been estimated using a $\mathrm{VAR}[p]$ model of the order $p = 7$ to fit the data, which is higher than the true order $p = 2$. Even in this situation the underlying network is revealed correctly.

C. Simulated fMRI Signals

To simulate signals which are closely related to actual fMRI-signals, we utilized a model for fMRI data here. To motivate this model, the mechanisms underlying fMRI signals are briefly summarized.

Oxygenated hemoglobin is diamagnetic and deoxyhemoglobin is paramagnetic. If deoxyhemoglobin is present the magnetic field of the tomograph is distorted and the MR signal is altered. This alteration is observable.

If a population of neurons is active, its consumption of oxygen increases. To deliver more oxygen to the population the blood flow in this brain region is increased. At the same time the local oxygen extraction fraction decreases due to the smaller transit time of the blood through the tissue. Therefore, locally the blood is more oxygenated and less deoxyhemoglobin is present; the activity of neurons leads to less deoxyhemoglobin. This change is observable and is called BOLD signal [16].

To generate data that resembles brain data as well as possible, dynamic causal modeling [1] as implemented in SPM5 was used as forward-model. Input signals have been generated with random stimulus onset times and durations. Dynamic causal modeling is based on a purely deterministic model, which is hardly expected to be present in real-world data. To overcome this limitation of DCM, independent Gaussian noise was added to the input stimulus time series for those nodes which do receive an input stimulus and pure Gaussian noise serves as input for those regions with no stimulus input. Thus, all regions receive an input signal, either pure Gaussian noise or randomized stimulus boxcar functions with additional Gaussian noise (cf. Fig. 6). Presence and absence of stimuli are binary-coded, i.e. input signals in dynamic causal modeling have the values 0 or 1. In this arbitrary units, we added driving Gaussian noise of standard deviation 0.25. Finally, observational Gaussian noise of two different variances was added to the time series.

D. The Model

The network structure used for this simulation is shown in Fig. 5. Its topology has been chosen similar to the topology of
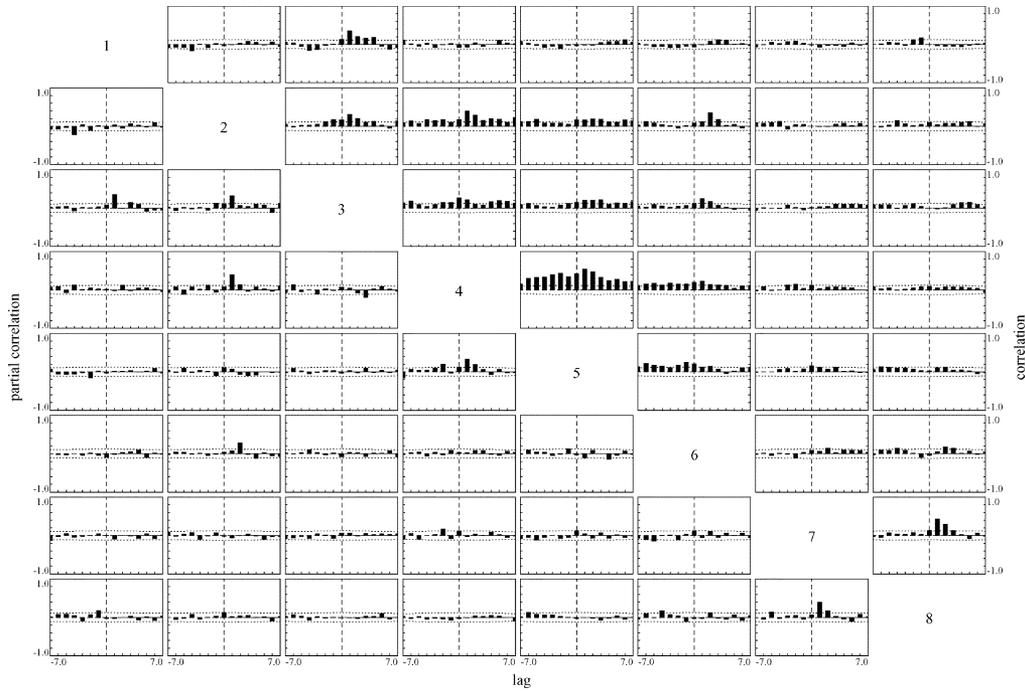
Fig. 4. Result for data simulated with a VAR[2] model with observational noise. The results are presented in the same way as in Fig. 3. The signal-to-noise ratio differs between 1:1 and 3:1 depending on the individual process. The order of the fitted $\mathrm{VAR}[p]$ was $p = 7$. Even if the order of the fit is chosen higher than the true one and the data is contaminated with observational noise, the revealed network is the same as the simulated one.
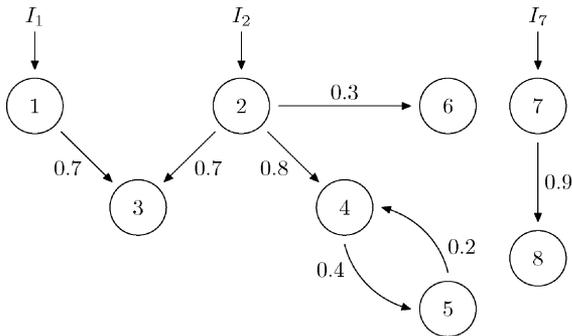

Fig. 5. Graph summarizing the simulated network for dynamic causal modeling data generation. Each input I_1, I_2 and I_7 is composed of boxcar functions plus independent Gaussian noise (cf. Fig. 6). All other nodes do also have input, but pure Gaussian noise.

the network for VAR simulation. Due to the different approaches of $\mathrm{VAR}[p]$ and dynamic causal modeling, input stimuli were added and some connection coefficients were altered this time.

The underlying activity of neuronal populations is modeled by

$$\dot{z}(t) = \left(\mathbf{A} + \sum_j u_j(t) \mathbf{B}^j \right) z(t) + \mathbf{C}u(t). \qquad (8)$$

The nonzero entries of the intrinsic connection matrix A are

$$
\begin{aligned}
A_{11} &= \ldots = A_{88} = -1 \\
A_{31} &= 0.7, \quad A_{32} = 0.7 \\
A_{42} &= 0.8, \quad A_{45} = 0.2 \\
A_{54} &= 0.4, \quad A_{62} = 0.3 \\
A_{87} &= 0.9.
\end{aligned} \qquad (9)
$$

The matrix B quantifying the effects of input signals on intrinsic connections is set to zero. The matrix C is, here, the identity matrix assigning the noisy inputs to every node. The nodes 1, 2 and 7 are chosen to receive boxcar stimuli. 1000 realizations of 256 data points were computed, where the intrinsic noise, observational noise, stimulus onsets and durations were drawn randomly every time. In accordance with data obtained from actual experiments the sampling time was set to 2.2 s. Stimulus lengths are drawn from a Gaussian distribution with a mean of 7 data points and a standard deviation of 5 data points, rounded to the next integer; for negative stimulus lengths the absolute value is taken. The stimulus onset times are equally distributed over the 256 data points. Fig. 6 shows an exemplary stimulus input function and two simulated time series generated as described above.

E. Signal-to-Noise Ratio in Simulated fMRI Signals

Signal-to-noise ratio (SNR) is defined as the quotient of signal variance divided by observational noise variance. It can be seen as measure for how grave the impact of the noise on the data is. We assume the absolute level of observational noise to be roughly constant in actual fMRI experiments over channels and thus also over the data of all regions under investigation. Therefore we added independent Gaussian noise each of the same variance to the data of every node in our simulations. As the signals of each node have different variances themselves, adding noise of constant variance leads to different SNRs for every node in the system. Additionally, as the input time series and the intrinsic noise are redrawn in every realization, the signal variance of each individual node itself as well varies in between realizations. Hence, it is not possible to declare an exact global SNR.
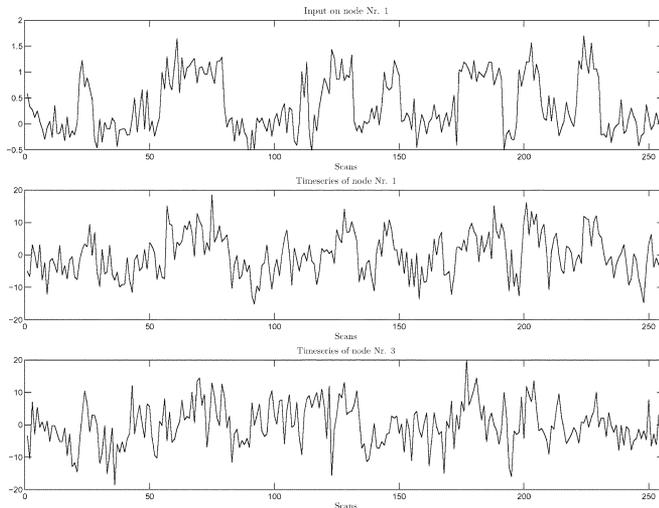
Fig. 6. Exemplary time series from the fMRI signal simulation. The upper plot shows an input time series of an stimulus contaminated with Gaussian noise. Every stimulus consists of a few boxcar functions. This is the input for node 1 of Fig. 5. The plot in the second row shows the resulting output of node 1. The plot at the bottom shows the output of node 3. Node 3 is driven by the output of node 1, but also by its own noise and by the output of node 2. The y-axis are in arbitrary units.

Two sets of simulated data have been generated, one with observational noise variance 6.25, the other with variance 20.25. These values *per se* are not meaningful, as the scale of the signals that DCM produces is rather arbitrary. Still, choosing 6.25 as noise variances leads to SNRs in the range of 1:1 to 4:1. We refer to this set of simulations as SNRa. Noise variance 20.25 leads to SNRs in the range of 0.4:1 to 0.9:1, this data set shall be called SNRb. Fig. 7 illustrates the actual SNR distributions for the SNRa data set. The distributions for the SNRb data set look similar, but shifted towards smaller SNRs. For SNRb the two outliers, node 5 and 6, have SNR distributions around 0.5:1 and 0.3:1 respectively.

F. Results for Simulated fMRI Signals

The connections have been estimated for every realization. The order of the fitted VAR[p] process was $p = 7$. To obtain group results we divide each directed partial correlation value by its significance level and then average this fractions over the group. This leads to values indicating the likeliness of the presence of a connection: values significantly higher than 1 indicate that, on average, the estimated direct partial correlation is higher than its significance level, i.e., a connection has been detected. All values regarded are those belonging to timelag $\tau = 0$, because connections are assumed to be fast in relation to the sampling time of 2.2 s used both in the simulation and the fMRI experiment below.

Tables I and II show the results of the group analysis with different levels of observational noise. The results presented in Table I correspond to data set SNRa, while Table II corresponds to SNRb. In both scenarios the strongest connections $A_{31}, A_{32}, A_{42}, A_{87}$ are correctly identified. The 2σ confidence interval of every result presented in the Tables I, II and III ranges from 0.03 to 0.04. The connection between nodes 4 and 5, actually a weak loop, is found only in the case of low observational
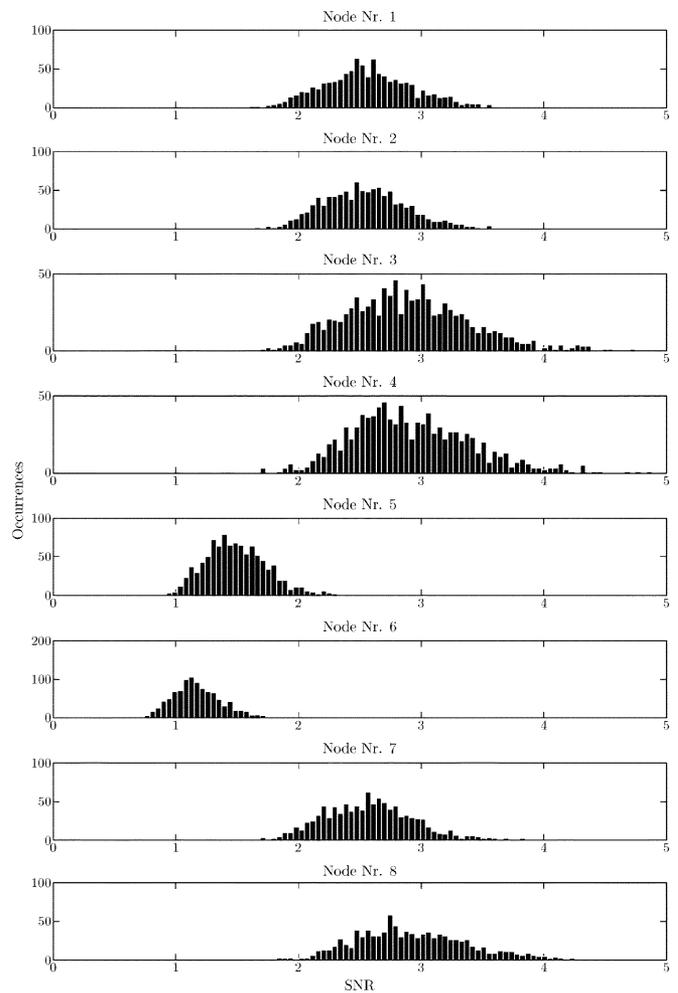

Fig. 7. Histograms of the signal-to-noise ratios from simulation set SNRa. Observational noise variance is 6.25. For most of the nodes the SNR ranges from 2:1 to 4:1. Nodes 5 and 6, however, show noticeable worse SNRs. As the observational noise variance is constant, the signal variance in this nodes has to be smaller than in the other nodes. This is consistent with the fact that nodes 5 and 6 receive the weakest stimulation, cf. model parameters in Fig. 5.

TABLE I
RESULTS OF THE NETWORK ESTIMATION WITH SIMULATED fMRI SIGNALS, DATA SET SNRa. THE LIKELINESS OF THE PRESENCE OF A CONNECTION IS CONSIDERED AS THE ESTIMATED DIRECTED PARTIAL CORRELATION DIVIDED BY ITS SIGNIFICANCE LEVEL. VALUES LARGER THAN ONE THEREFORE INDICATE TRUE DIRECT CONNECTIONS. THE SHOWN VALUES ARE THE MEANS OF 1000 INDIVIDUAL REALIZATIONS. THEY ARE MEANT TO DENOTE AN INTERACTION BETWEEN THE NODE IN COLUMN i AND THE NODE IN ROW j OF THE TABLE. THE 2σ CONFIDENCE INTERVAL OF THE MEAN RANGES BETWEEN 0.03 AND 0.04 FOR EACH CASE

	1	2	3	4	5	6	7
2	-0.5						
3	**2.6**	**1.7**					
4	-0.2	**2.4**	0.7				
5	0	0.3	0.1	**1.6**			
6	-0.1	1.0	0.3	0.4	0.1		
7	0	0	0	0	0	0	
8	0	0	0	0	0	0	**3.1**

noise. The weak connection between nodes 2 and 6 cannot be detected in either setting. We want to point out that the presence

TABLE II
RESULTS OF THE NETWORK ESTIMATION WITH SIMULATED fMRI SIGNALS FROM DATA SET SNRb (CF. TABLE I). THE 2σ CONFIDENCE INTERVAL OF THE MEAN IS 0.04 IN EACH CASE. THE WEAK CONNECTION BETWEEN NODES 4 AND 5 COULD NOT BE REVEALED IN THIS CASE

	1	2	3	4	5	6	7
2	-0.2						
3	**1.4**	**1.1**					
4	-0.2	**1.6**	0.7				
5	0	0.4	0.2	1.0			
6	0	0.7	0.3	0.4	0.1		
7	0	0	0	0	0	0	
8	0	0	0	0	0	0	**2.0**

TABLE III
RESULTS OF THE ESTIMATION WITH SIMULATED fMRI SIGNALS FROM DATA SET SNRa WITH ALL 3 INPUTS INCLUDED TO THE ANALYSIS. THE TABLE SHOWS ONLY THE RESULTS FOR THE INPUTS ACTING ON THE REGIONS TO TIME LAG $\tau = -1$. WE FIND A HIGH LIKELINESS OF THE INPUTS ACTING ON THE CORRECT NODES, INPUT 1 ON NODE 1, INPUT 2 ON NODE 2 AND INPUT 7 ON NODE 7. STILL, WE ALSO FIND SIGNIFICANT EVIDENCE FOR INPUT 2 ACTING ON NODE 4 AND INPUT 7 ACTING ON NODE 8. THE 2σ CONFIDENCE INTERVAL OF THE MEAN IS SETTLED BETWEEN 0.03 AND 0.04 IN EACH CASE

	1	2	3	4	5	6	7	8
i1	**2.3**	0	0.9	0	0	0	0	0
i2	0	**2.3**	0.9	**1.2**	0.3	0.6	0	0
i7	0	0	0	0	0	0	**2.3**	**1.2**

of observational noise always reduces the accuracy of estimation. Still, only the very weak connections could not be recovered in our simulation studies. However, we want to emphasize that dPC does not draw false positive conclusions. Therefore, if dPC does not detect a connection between two regions this does not necessarily imply that the regions in the underlying network are disconnected. The data, in this case, simply does not carry enough information to reveal the connection. If, in contrary, a connection is detected, dPC provides a statistical significant and therefore a reliable estimate.

G. Integrating the Stimuli

Both in simulations and in real fMRI studies, the experimenter has good knowledge about onsets and durations of stimuli. In order to learn more about the levels of signal processing within the set of regions, one might want to know which regions the stimulus initially acts on. In terms of dPC this would be represented by causal relationships between the stimuli and the network.

To preliminarily investigate if dPC is capable of drawing such inferences, we have chosen a very naïve approach. The inputs to our simulations are modeled as noisy boxcar functions, cf. Section III-C. Here, we segregate the boxcar functions from the noise and regard the clean boxcar function as the external stimulation time series. These time series may now be included in the dPC estimation. The dimension of the network thereby increases by the number of input time series included.

As mentioned above the time resolution of an fMRI experiment is around 2 s. Relative to the stimulus onset the corresponding BOLD response is also delayed by 1 to 2 s [16]. Therefore, an interaction between stimulus and network is expected to be found at time lag $\tau = -1$.

Fig. 8 shows the result of a dPC estimation of a single realization of data from the SNRa set with one input included. Note that the integration of binary boxcar functions clearly violates the assumptions for AR processes. Therefore, inferences on the actual network structure should not be drawn from the estimations with inputs included. Looking at the bottom row, Fig. 8 delivers strong evidence for input 1 acting on region 1 to time lag $\tau = -1$. Comparing with the simulated network topology in Fig. 5 we find that this is correct.

As for the analyses above, we can compute group results based on all realizations from both data sets. This time, though, group results are computed from dPCs to timelag $\tau = -1$. Table III shows the results for data set SNRa with all 3 inputs included simultaneously. The inputs are linked to the correct nodes with the highest likeliness. The two false positives result from the fact that the nodes receive the stimulation indirectly: input 2 is propagated to node 4 by node 2 and input 7 is propagated to node 8 by node 7. It appears that, in this setting of fast signal propagation, slow sampling frequency and the naïve input integration approach, this difference between direct and indirect influence can not be resolved.

The group results for the SNRb data set, however, do not reveal false positive conclusions but only the correct assignment of inputs to nodes. The likeliness values are 1.5 with a 2σ confidence interval of the mean of 0.04 in each case. The decrease of the values is an expected result of the lower SNR.

Altogether, the promising results of this naïve integration of the inputs suggest further development in this direction.

IV. ANALYSIS OF fMRI DATA

A. The Data

In an fMRI event-related experiment we aurally presented 90 sentences of each normal speech, pseudo speech and reversed speech, resulting in 270 stimuli distributed to three sessions with a total of 690 scans. The sampling time was 2.2 s. From seed regions, which were defined by the peak voxels of activated clusters on group level, time series were extracted. The data was fully preprocessed with slice time correction, motion correction and smoothing with a 9 mm Gaussian smoothing kernel over the entire brain. For all corrections SPM5 was used.

To get rid of the scanner drift we fitted a polynomial of degree 3 to each time series and subtracted its function values from it. A small simulation study has shown that polynomials are less critical concerning AR-fitting than e.g. low-pass filters.

B. The Results

Here, we present the result (see Fig. 9) of a single subject (see Fig. 10). The five temporal and frontal seed regions are functionally interconnected directly and indirectly. From the posterior temporal seed (T2p) an indirect functional connection to the frontal lobe (F3tri) via the fusiform gyrus (FUS) and anterior temporal lobe (T2a) as well as a direct functional connection was found. This supports current theories on language processing suggesting parallel and serial information flow. However, these findings need to be evaluated in a large-scale group analysis.
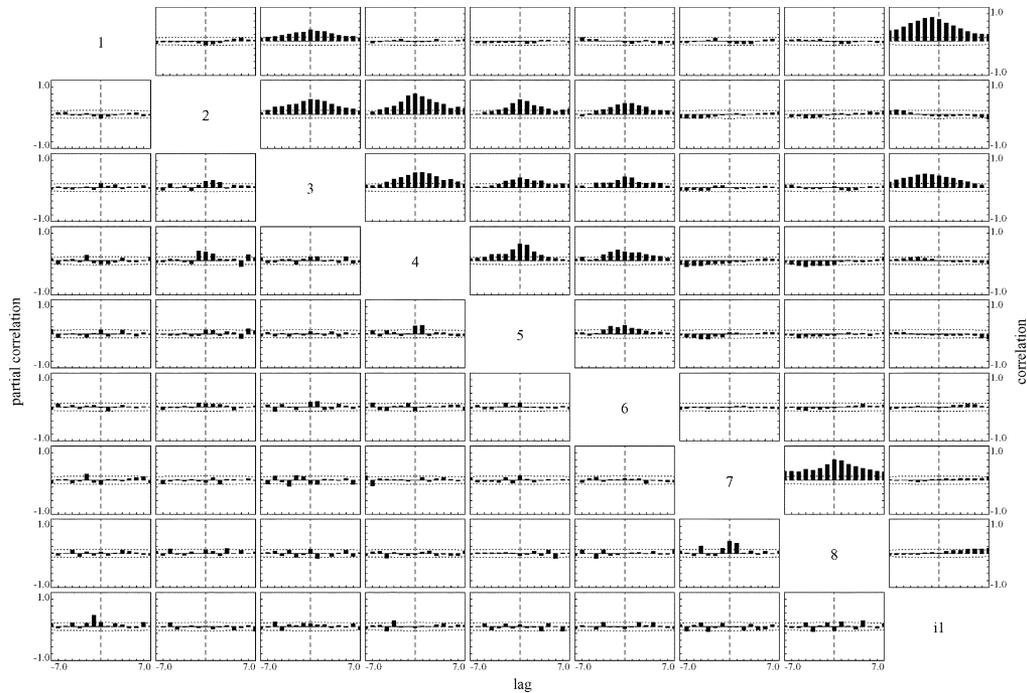
Fig. 8.   Results of the dPC estimation of one realization from the SNRa data set with input 1 included. In the bottom line we see that input 1 affects node 1 to time lag $\tau = -1$. This is the correct assignment. No other node is influenced by input 1 with such high significance and to a reasonable time lag. Regarding the estimated network topology, one can see that some connections are still estimated correctly. A_{32}, A_{42}, A_{54} and A_{87} are found significant at time lag $\tau = 0$. Still, as the inclusion of the nonstochastic input time series violates the requirements for AR modeling, the network topology should be inferred from analyzes without included inputs.
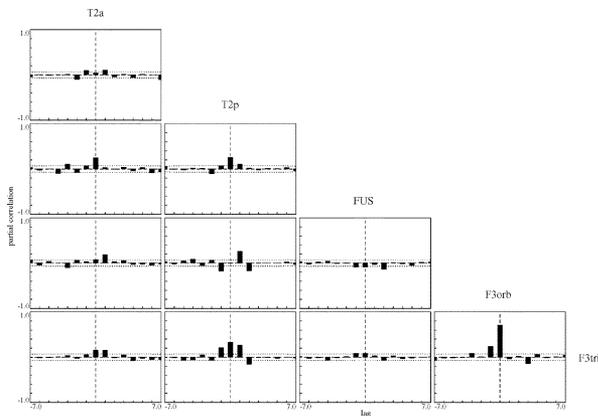

Fig. 9.   Estimated dPC for the presented single subject. Due to the low time resolution of the fMRI technique only dPCs for timelag $\tau = 0$ are of concern. This values are represented by the bars directly on the dotted vertical line in every subplot. Fig. 10 is a graphical representation of this result.
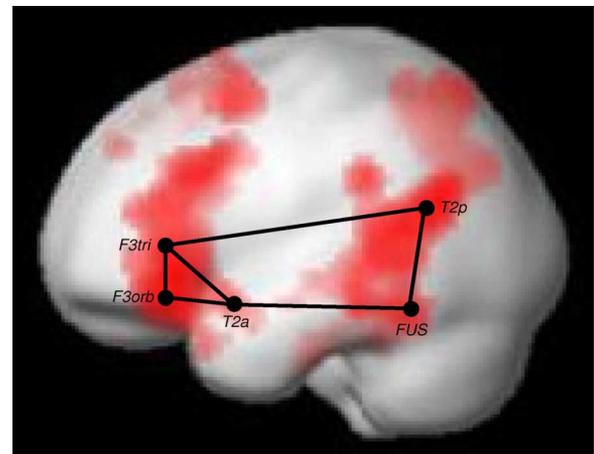

Fig. 10.   fMRI single subject data for the contrast of speech compared with meaningless pseudo speech ($p < 0.05$ corrected) are superimposed on a canonical brain. Peak activations are indicated with black dots and were located in the anterior (T2a) and posterior temporal lobe (T2p), in the fusiform gyrus (FUS) and the inferior frontal gyrus, pars orbitalis (F3orb) and triangularis (F3tri). Significant functional connections between seed regions are indicated with black lines. For the estimates see Fig. 9.

V. DISCUSSION AND CONCLUSION

This manuscript presents an overview on directed partial correlation and a simulation which shows, that dPC can be applied to fMRI data. Based on this simulation and an application to real-world data, we present dPC's advantages and limitations.

In fMRI experiments the seed points, and therefore the timeseries which are taken from these seed points are selected in the first place. To this aim, the seed points contributing to the network have to be identified. To address the issue of how these seed points should be detected and which seed points should be included in the analysis is beyond the scope of this manuscript.

We mention though that the preselection of certain seed points will influence the result of the analysis, as directed partial correlation detects the interaction structure of the underlying networks based on the observed time series. We recommend performing simulation studies tailored to the problem at hand prior to the analysis.

If directed partial correlation is confronted with data which holds too poor information about the network structure, dPC is

not able to estimate the underlying network structure. Such situations are likely to appear if the measured network consists of too many nodes. With an increasing number of nodes the information on the network must increase, as the number of possible connections does. But because dPC does not tend to infer false positive conclusions it provides the researcher with a good hint if the data is good enough for this kind of investigation. If dPC ends up with all nodes unconnected and the network is expected to have connections, for instance learned from bivariate analysis, the data is not able to reveal this network. Analyses towards subnetworks are a reasonable way out of this dilemma [17]. In cases with too poor signal-to-noise ratios, directed partial correlation analysis again ends up with isolated nodes. In any case of our simulations directed partial correlation analysis did not draw false positive conclusions about the network structure.

Additionally, directed partial correlation is model independent: No anatomical or functional a priori knowledge is needed. The method is entirely data driven; this way it is possible to analyze networks which have not been under investigation so far. It is also possible to use directed partial correlation as a pre-analysis for other techniques like DCM which do need constraint about the underlying network.

By analyzing actual data from an auditory paradigm, we demonstrated the performance of directed partial correlation in a real-world scenario. Further analyses for a collective of subjects will be performed in the near future.

REFERENCES

[1] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *Neuroimage*, vol. 19, no. 4, pp. 1273–1302, Aug. 2003.

[2] M. Eichler, R. Dahlhaus, and J. Sandkühler, "Partial correlation analysis for the identification of synaptic connections," *Biol. Cybern.*, vol. 89, pp. 289–302, 2003.

[3] R. Dahlhaus and M. Eichler, "Causality and graphical models for time series," in *Highly Structured Stochastic Systems*, P. Green, N. Hjort, and S. Richardson, Eds. Oxford, U.K.: Oxford Univ. Press, 2003, pp. 115–137.

[4] B. Schelter, M. Winterhalder, B. Hellwig, B. Guschlbauer, C. Lücking, and J. Timmer, "Direct or indirect? Graphical models for neural oscillators," *J. Physiol. Paris* vol. 152, no. 1, pp. 210–219, Jan. 2006 [Online]. Available: http://www.dx.doi.org/10.1016/j.jphysparis.2005.06.006

[5] X. Zheng and J. C. Rajapakse, "Learning functional structure from fMR images," *Neuroimage* vol. 31, no. 4, pp. 1601–1613, Jul. 2006 [Online]. Available: http://www.dx.doi.org/10.1016/j.neuroimage.2006.01.031

[6] J. C. Rajapakse and J. Zhou, "Learning effective brain connectivity with dynamic Bayesian networks," *Neuroimage*, vol. 37, no. 3, pp. 749–760, Sep. 2007.

[7] J. Granger and M. Hatanaka, *Spectral Analysis of Economic Time Series*. Princeton, NJ: Princeton Univ. Press, 1964.

[8] J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.

[9] G. Marrelec, B. Horwitz, J. Kim, M. Plgrini-Issac, H. Benali, and J. Doyon, "Using partial correlation to enhance structural equation modeling of functional MRI data," *Magn. Reson. Imag.* vol. 25, no. 8, pp. 1181–1189, Oct. 2007 [Online]. Available: http://www.dx.doi.org/10.1016/j.mri.2007.02.012

[10] K. J. Friston, A. P. Holmes, C. J. Price, C. Bchel, and K. J. Worsley, "Multisubject fMRI studies and conjunction analyses," *Neuroimage* vol. 10, no. 4, pp. 385–396, Oct. 1999 [Online]. Available: http://www.dx.doi.org/10.1006/nimg.1999.0484

[11] M. Eichler, "A graphical approach for evaluating effective connectivity in neural systems," *Philos. Trans. Roy. Soc. Lond. B, Biol. Sci.* vol. 360, no. 1457, pp. 953–967, May 2005 [Online]. Available: http://www.dx.doi.org/10.1098/rstb.2005.1641

[12] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistritz, D. Klan, R. Bauer, J. Timmer, and H. Witte, "Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems," *Signal Process.*, vol. 85, pp. 2137–2160, 2005.

[13] M. Eichler, "Graphical modeling of dynamic relationships in multivariate time series," in *Handbook of Time Series Analysis*, B. Schelter, M. Winterhalder, and J. Timmer, Eds. New York: Wiley-Vch, 2006, ch. 14, pp. 335–372.

[14] H. Lütkepohl, *Introduction to Multiple Time Series Analysis*. New York: Springer, 1993.

[15] D. Smirnov, B. Schelter, M. Winterhalder, and J. Timmer, "Revealing direction of coupling between neuronal oscillators from time series: Phase dynamics modeling versus partial directed coherence," *Chaos* vol. 17, no. 1, p. 013111, 2007 [Online]. Available: http://www.link.aip.org/link/?CHA/17/013111/1

[16] R. B. Buxton, K. Uluda, D. J. Dubowitz, and T. T. Liu, "Modeling the hemodynamic response to brain activation," *Neuroimage* vol. 23, no. Suppl 1, pp. S220–S233, 2004 [Online]. Available: http://www.dx.doi.org/10.1016/j.neuroimage.2004.07.013

[17] B. Schelter, B. Hellwig, B. Guschlbauer, C. Lücking, and J. Timmer, "Application of graphical models in bilateral essential tremor," in *Proc. IFMBE (EMBEC)*, 2002, vol. 2, pp. 1442–1443.

Wolfgang Mader is currently pursuing the physics degree at the University of Freiburg, Freiburg, Germany, working on his thesis on detecting information flow in fMRI data.

He is a member of the research group "Freiburg Brain Imaging" at the Department for Neurology, University Medical Center of Freiburg, and the group Data Analysis and Modeling of Dynamic Processes in the Life Sciences, Center for Data Analysis and Modeling, Institute of Physics, University of Freiburg. His main research interest is the development of multivariate time series analysis techniques and their application to fMRI data.

David Feess is pursuing the degree in physics and working on his thesis at the Department for Neurology, University Medical Center of Freiburg, Freiburg, Germany.

He is a Researcher in the "Freiburg Brain Imaging" group at the Department for Neurology, University Medical Center of Freiburg, and the group Data Analysis and Modeling of Dynamic Processes in the Life Science, Center for Data Analysis and Modeling, Institute of Physics, University of Freiburg. His main research interest is the development of multivariate time series analysis techniques and their application to fMRI data.

Rüdiger Lange is a Neurologist at the Department of Neurology, University Medical Center, Freiburg, Germany. His research focuses on the investigation and modulation of motor behavior using trancranial magnetic stimulation and functional magnetic resonance imaging.

Dorothee Saur is a Neurologist at the Department of Neurology, University Medical Center, Freiburg, Germany. Her research focuses on the investigation of language processing in the healthy brain and reorganization of language functions in the lesioned brain using neuroimaging techniques like fMRI and DTI.

Jens Timmer received the Ph.D. degree from the University of Freiburg, Freiburg, Germany, in 1994.

Since 2005, he has been a Professor in Theoretical Physics at the University of Freiburg. His research interests are the development and interdisciplinary application of mathematical methods to analyse and model dynamic processes in the life sciences.

Volkmar Glauche is coordinating the research groups in "Freiburg Brain Imaging" at the University of Freiburg Medical Center, Freiburg, Germany. His research mainly focuses on the integration of functional and structural MRI analysis methods.

Björn Schelter received the Ph.D. degree from the Institute of Physics, University of Freiburg, Freiburg, Germany.

He is a member of the research group Data Analysis and Modeling of Dynamic Processes in the Life Sciences at the Center for Data Analysis and Modeling and the Institute of Physics of the University of Freiburg, Germany. His research mainly focuses on the development and interdisciplinary application of multivariate time series analysis techniques.

Cornelius Weiller is Head of the Department of Neurology at the University of Freiburg, Freiburg, Germany. He researches anatomy of language and motor function and the adaptation to lesions during treatment using *in-vivo* imaging techniques as MRI or PET.