Identifying interactions in mixed and noisy complex systems

Guido Nolte* and Frank C. Meinecke†
Fraunhofer FIRST.IDA, Kekuléstrasse 7, D-12489 Berlin, Germany

Andreas Ziehe‡
Fraunhofer FIRST.IDA, Kekuléstrasse 7, D-12489 Berlin, Germany
and Technical University Berlin, Institute for Software Engineering, Franklinstrasse 28/29, 10587 Berlin, Germany

Klaus-Robert Müller§
Fraunhofer FIRST.IDA, Kekuléstrasse 7, D-12489 Berlin, Germany
and Institut für Informatik, Universität Potsdam, August-Bebel Strasse 89, D-14482 Potsdam, Germany
(Received 5 August 2005; revised manuscript received 16 December 2005; published 23 May 2006)

We present a technique that identifies truly interacting subsystems of a complex system from multichannel data if the recordings are an unknown linear and instantaneous mixture of the true sources. The method is valid for arbitrary noise structure. For this, a blind source separation technique is proposed that diagonalizes antisymmetrized cross-correlation or cross-spectral matrices. The resulting decomposition finds truly interacting subsystems blindly and suppresses any spurious interaction stemming from the mixture. The usefulness of this interacting source analysis is demonstrated in simulations and for real electroencephalography data.

DOI: 10.1103/PhysRevE.73.051913 PACS number(s): 87.19.Nn, 89.75.Hc, 05.40.Ca, 05.45.Tp

I. INTRODUCTION

An important part of the current research in physics, biology, and medical science addresses the analysis and modeling of complex dynamical systems. Often, these systems consist of interacting subsystems; thus, understanding the interaction processes is of fundamental importance. However, when measuring data from complex systems, often only *superimposed* signals with a high noise contamination can be assessed. Superposition effects are dominant where only surface measurements are accessible, while the interesting dynamical processes are hidden "underneath," e.g., in seismology, the study of active stars, or neuroscience [1].

The prime example of such a complex system with interacting components is the human brain. A huge number of neurons form cortical areas that interact dynamically to represent and process information. Although the contribution of this paper is of general interest, our further discussion will focus on the important example of cortical interactions. When measuring brain activity, e.g., with electroencephalography (EEG) or magnetoencephalography (MEG), the sensor information only indirectly reflects the dynamics of the single subsystems involved: volume conduction effects systematically only allow one to measure superpositions with some additional noise.

To study cortical interactions, which manifest in phase synchrony or coherent states in brain activity [2–4], it is of fundamental importance to distinguish interaction-related effects from those that are only induced by superposition and, by this, to avoid measuring spurious interaction. Furthermore, it is typically unclear, what the interesting subsystems are or, in other words, how the whole system should be decomposed into such subsystems.

Recently, there have been several approaches that address parts of this problem from different perspectives, e.g., by proposing a surrogate test [5], analyzing the imaginary part of the coherency function [6], or calculating the phase coherency after a prior source separation step [7].

In this paper, we aim to establish a more general concept for extracting truly interacting components. In contrast to existing methods, this concept is insensitive to mixing effects, can handle both narrow- and wideband interactions, is genuinely multivariate and suited even for situations when there are far more active sources than sensors. Particularly, it can be applied to systems with *arbitrary* noise structure. Although with a different objective, the proposed concept has a close structural relationship with methods from the field of independent component analysis (ICA) [1]; thus, we will give a brief introduction to ICA based on second-order statistics before we derive some fundamental properties of our method. Finally, we will apply it to simulated data and real EEG.

II. THEORY

The fundamental assumption of ICA is that a data matrix X, without loss of generality assumed to be zero mean, originates from a superposition of independent sources S such that

$$X = AS \tag{1}$$

where A is called the mixing matrix which is assumed to be invertible. The task is to find A and, hence, S (apart from trivial ordering and scale transformations of the columns of

*Electronic address: nolte@first.fhg.de
†Electronic address: meinecke@first.fhg.de
‡Electronic address: ziehe@first.fhg.de
§Electronic address: klaus@first.fhg.de

A and the rows of S) by merely exploiting statistical independence of the sources. Since independence implies that the sources are uncorrelated we may choose W, the estimated inverse mixing matrix, such that the covariance matrix of

$$\hat{S} \equiv WX \tag{2}$$

is equal to the identity matrix. This, however, does not uniquely determine W because for any such W also UW, where U is an arbitrary orthogonal matrix, leads to a unit covariance matrix of $\hat{S}$. Uniqueness can be restored if we require that W not only diagonalizes the covariance matrix but also cross-correlation matrices for one [8] or more [9] delays τ, i.e., we require that

$$WC^X(\tau)W^\dagger = \text{diag} \tag{3}$$

with

$$C^X(\tau) \equiv \langle \mathbf{x}(t)\mathbf{x}^\dagger(t+\tau)\rangle \tag{4}$$

where $\mathbf{x}(t)$ is the tth column of X and the angular brackets mean expectation value, which is estimated by the average over t. Although at this stage all expressions are real valued, we introduce a complex formulation for later use.

Note, that since under the ICA assumption the cross-correlation matrices $C^S_{ij}(\tau)=\langle s_i(t)s_j(t+\tau)\rangle$ of the source signals are diagonal and hence the cross-correlation matrices of the mixtures, $C^X(\tau)=AC^S(\tau)A^\dagger$, are symmetric. Consequently, the antisymmetric part of $C^X(\tau)$ can only arise due to meaningless fluctuations and can be ignored. In fact, the TDSEP algorithm [9] uses symmetrized versions of $C^X(\tau)$.

Now, the key and main point of our method is that we will turn the above argument upside down. Since noninteracting sources do not contribute (systematically) to the antisymmetrized correlation matrices

$$D(\tau) \equiv C^X(\tau) - C^{X\dagger}(\tau), \tag{5}$$

any (significant) nonvanishing elements in $D(\tau)$ must arise from interacting sources, and hence, the analysis of $D(\tau)$ is ideally suited to study interactions.

It is now our goal to identify one or many interacting subsystems from a suitable spatial transformation, which corresponds to a demixing of the subsystems rather than individual sources. We will use the technique of simultaneous diagonalization to achieve this goal. We first note that a diagonalization of $D(\tau)$ using a real-valued W is impossible since with $D(\tau)$ also $WD(\tau)W^\dagger$ is antisymmetric and always has vanishing diagonal elements. Hence, $D(\tau)$ can only be diagonalized with a complex-valued W with subsequent interpretation of it in terms of the field patterns of the interacting sources.

We will here discuss the case where all interacting subsystems consist of pairs of (neuronal) sources. Properties of systems with more than two interacting sources will be discussed below. Furthermore, for simplicity we assume an even number of channels. Also, it is understood that for finite data sets the following is only approximately true. Then, a real-valued spatial transformation W_1 exists such that the set of $D(\tau)$ becomes decomposed into $K=N/2$ blocks of size 2×2

$$W_1 D(\tau) W_1^\dagger = \begin{pmatrix} \alpha_1(\tau)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \alpha_K(\tau)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \end{pmatrix}, \tag{6}$$

which can be diagonalized, e.g., with

$$W_2 = \frac{1}{2} id_{K\times K} \otimes \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \tag{7}$$

such that $W_2W_1D(\tau)W_1^\dagger W_2^\dagger = \text{diag}$.

From a simultaneous diagonalization of $D(\tau)$, we obtain $W=W_2W_1$. We are interested in the columns of W_1^{-1}, which correspond to the spatial patterns of the interacting sources. From $W^{-1}=W_1^{-1}W_2^{-1}$ and

$$W_2^{-1} = id_{K\times K} \otimes \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, \tag{8}$$

we observe that the desired spatial patterns are contained in the real and imaginary parts of the columns of W^{-1}. Furthermore, the columns come in pairs (if $\mathbf{w}$ is a column then so is $\mathbf{w}^*$) as well as the corresponding diagonals, which read $\pm i\alpha_k(\tau)$ for the kth pair.

Since diagonalization is invariant with respect to phase transformations, that are independent of channel index and frequency, the real and imaginary parts of a column $\mathbf{w}$ only span the same two-dimensional (2D) subspace as the true field patterns of the sources. Indeed, if $\mathbf{a}$ and $\mathbf{b}$ are these field patterns, then our observables are sums of matrices proportional to $\mathbf{ab}^\dagger - \mathbf{ba}^\dagger$, which is invariant under mixtures *within* the subspace spanned by $\mathbf{a}$ and $\mathbf{b}$.

Instead of diagonalizing the antisymmetrized cross-correlation matrices, it is also possible to diagonalize their respective Fourier-transforms, which are identical to the imaginary parts of the cross-spectra. This variant of the method has the advantage that we can account for largely different signal amplitudes as a function of frequency by an appropriate normalization of the cross-spectra. Specifically, we will diagonalize

$$D(f) \equiv \frac{\text{Im}[C(f)]}{\|C(f)\|} \tag{9}$$

with

$$C(f) \equiv \langle \hat{\mathbf{x}}(f)\hat{\mathbf{x}}^\dagger(f)\rangle, \tag{10}$$

where $\hat{\mathbf{x}}(f)$ is the Fourier-transform of the data and $\|\cdot\|$ is the Frobenius norm. With this normalization, we effectively weight the cross-spectra with respect to a rough estimate of the noise power.

The chosen normalization contains the risk that interesting phenomena are de-emphasized. This would be clearly the case if, e.g., large spectral power always corresponds to interacting systems. On the other hand, higher frequencies contain much less noise than lower ones. An interacting rhythmic system at a high frequency, which is as much smaller in

amplitude as the noise, should be detectable as well as the (hypothetical) low-frequency counterpart. This can only be achieved by a weighting using an estimate of the noise level. Furthermore, it is not clear *in advance* whether a specific rhythm arises from an interacting system or whether a specific interacting system is rhythmic at all. Hence, the normalization provides an unprejudiced view across all frequencies. Specifically, the fact that we observe spectral peaks in the diagonals for the real data example (see Sec. III) although we normalized with respect to power indicates that the observed phenomenon is *not* an artifact. We want to emphasize, however, that the optimal choice of weights depends very much on the specific case and is beyond the scope of this paper.

The general procedure can now be outlined as follows: (a) from the data construct antisymmetrized and normalized cross-spectral matrices as defined in Eq. (9) for a reasonable set of frequencies f, (b) find a complex matrix W such that $WD(f)W^{\dagger}$ is approximately diagonal for all f, and (c) interpret the columns of the demixing matrix and the corresponding diagonal elements as spatial and spectral properties of the interacting systems.

Although (approximate) simultaneous diagonalization of $D(f)$ using complex demixing matrices is always possible with pairwise interactions, we can expect only block-diagonal structure if a larger number of sources are interacting within one or more subsystems. In general, the imaginary cross-spectral matrices are diagonalizable (apart from random fluctuations) if for each interacting system the sources come in pairs. If this is not fulfilled, our method yields, by construction, the dominant part of the interactions, but off-diagonals contain additional information. We also assumed that the effective number of *interacting* sources is smaller than the number of channels. If this is violated, we still observe true interactions, but the decomposition into independent subsystems is incomplete.

Matrices were approximately simultaneously diagonalized with the DOMUNG algorithm [10], which was generalized to the complex domain. Here, an initial choice for the demixing

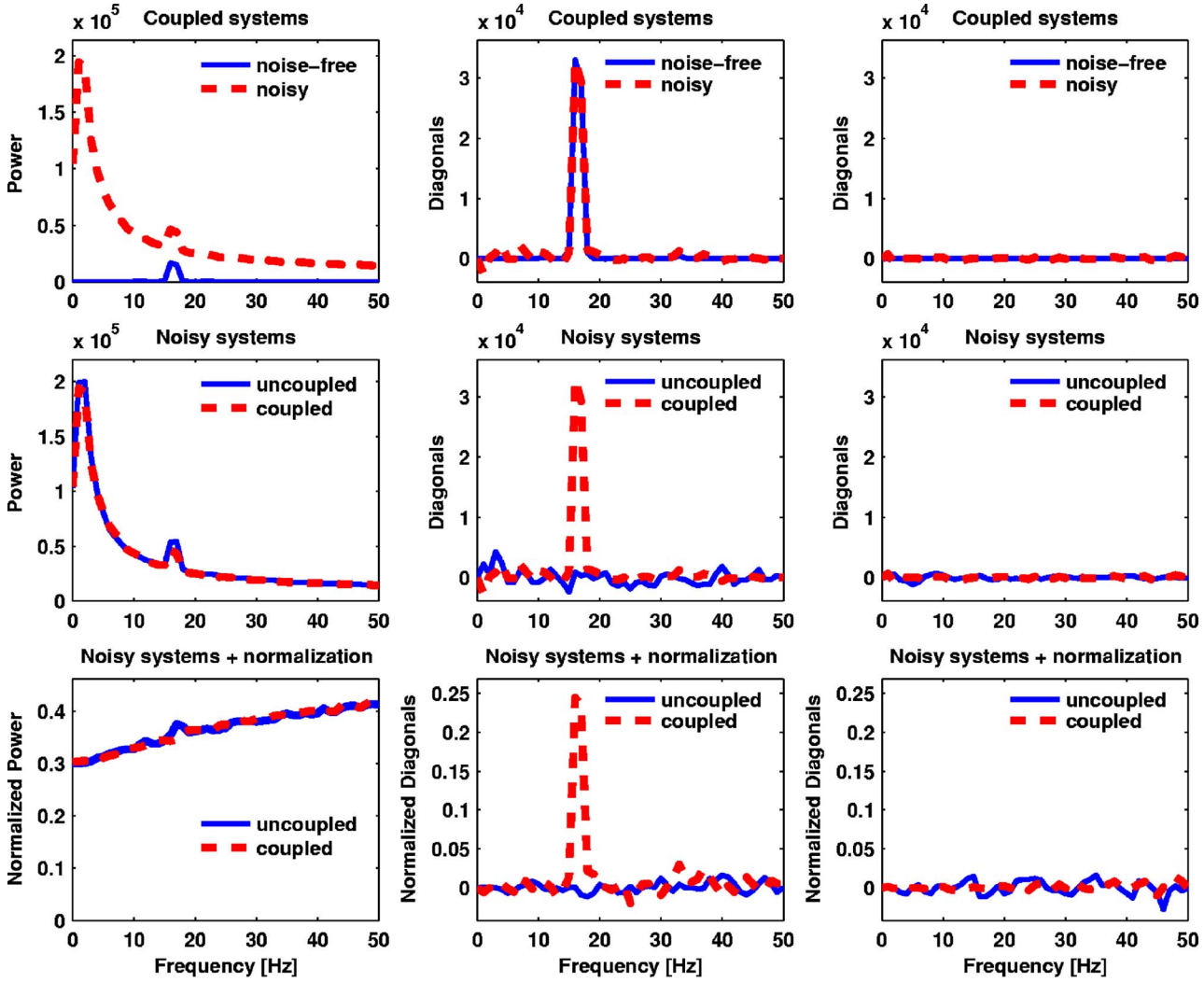

FIG. 1. (Color online) Each row contains results for two different simulations shown as full and dashed lines, respectively. Top: coupled Rössler system for the noisy vs noise-free case; middle: coupled vs uncoupled Rössler system with noise added; bottom: same as middle row using now normalized cross-spectra. The panels in the left column show power averaged over four channels. The panels in the middle (right) column show the corresponding diagonals of the first (second) ISA component after complex diagonalization for each data set.

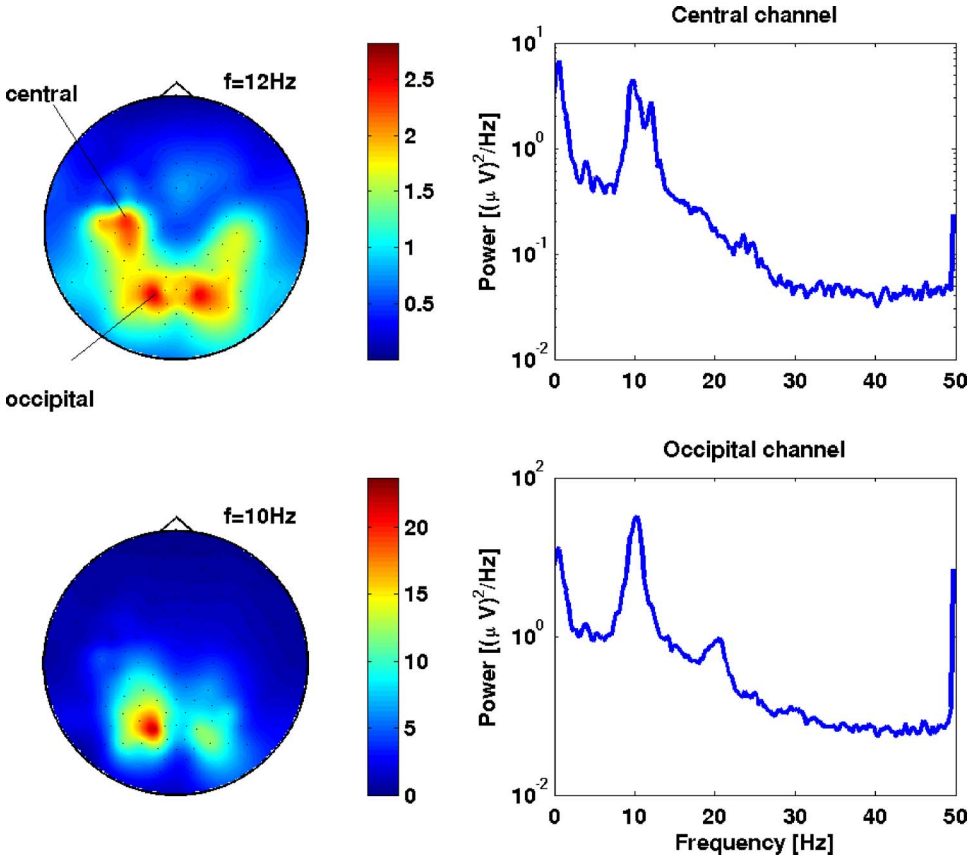

FIG. 2. (Color online) Power as a function of electrode location for two frequencies, and as function of frequency for two selected channels as indicated.

matrix W is successively optimized using a natural gradient approach combined with line search according to the requirement that the off-diagonals are minimal under the constraint $\det(W)=1$. Special care has to be taken in the choice of the initial guess. Because of the complex-conjugation symmetry of our problem (i.e., W^* diagonalizes as well as W) the initial guess may not be set to a real-valued matrix because then the component of the gradient in the imaginary direction will be zero and W will converge to a real-valued saddle point.

III. RESULTS

A. Simulations

We first applied our method to simulated data generated from a coupled Rössler system with the same parameters as in [7]

$$\dot{s}_{1,2} = -\omega_{1,2}u_{1,2} - v_{1,2} + c(s_{2,1} - s_{1,2})$$

$$\dot{u}_{1,2} = \omega_{1,2}s_{1,2} + 0.15u_{1,2}$$

$$\dot{v}_{1,2} = 0.2 + v_{1,2}(s_{1,2} - 10) \qquad (11)$$

with $\omega_1=1.015$, $\omega_2=.985$, and $c=0$ for the uncoupled case and $c=0.04$ for the coupled case. If we define the time unit to be 10 ms, then this system has a spectral peak at 16 Hz. We simulated a 10 min recording with a sampling rate of 100 Hz. We considered s_1 and s_2 as two sources that were randomly mapped into a four-dimensional channel space. To these signals of interest we subsequently added (a) spatially and temporally white noise and (b) the signals of four independent sources containing $1/f$-noise, which were mapped into channel-space by a random mixing matrix with positive elements. The latter was chosen to guarantee additional correlated noise rather than almost spatially white noise due to compensation of correlated and anticorrelated parts. Cross-spectra were calculated on Hanning-windowed overlapping segments of 1 s duration. Each segment was linearly detrended. Diagonalization of the imaginary parts of the cross-spectra was performed both with and without normalization [see Eq. (9)]. The relative contributions of signal and noise parts were defined through the average power across all frequencies and channels. We chose this average for the correlated noise to be twice as big as the average power for the channel noise and the total noise power to be 50 times as big as the average signal power.

The power, without normalization and averaged over all channels, is shown in the upper left panel of Fig. 1 for the coupled system for the noise-free and noisy cases. Diagonalization in four channels leads to two systems for each data set. In the upper middle and upper right panel we show the corresponding diagonals, always divided by i, of the first and second component, respectively. We observe that both for the noisy and noise-free cases only one of the diagonals displays a notable peak, which indicates true interaction, at 16 Hz, showing that the spatial decomposition was essentially error free also for the case where the noise is much larger than the signal of interest.

In the middle row, we compare coupled to uncoupled systems with noise added in both cases. Although the power is

similar, the diagonals indicate a significant interaction only for the coupled case. In the lower panels we show the results for normalized cross-spectra as defined in Eq. (9) for the coupled and uncoupled Rössler system with noise added. We note that for this simulation normalization leads to somewhat pathological results in the noise-free case. The average power is now lower for low frequencies because channels are then more correlated. As in the non-normalized case, the diagonals indicate interaction only for the coupled case.

To evaluate the accuracy of the estimated spatial patterns we constructed the antisymmetric outer product of the real and imaginary parts of the respective ISA-component say **a** and **b**, by

$$C = \frac{\mathbf{a}\mathbf{b}^\dagger - \mathbf{b}\mathbf{a}^\dagger}{\|\mathbf{a}\mathbf{b}^\dagger - \mathbf{b}\mathbf{a}^\dagger\|} \tag{12}$$

where $\|\cdot\|$ denotes Frobenius norm. The distance d in percent between a true pattern C_{true}, constructed as above with the true spatial patterns of the interacting sources, is now defined as

$$d = \min_{\pm}(\|C_{\text{true}} \pm C\|)100\% \tag{13}$$

The minimum with respect to the sign and the normalization is necessary because the spatial pattern $(\mathbf{a}\mathbf{b}^\dagger - \mathbf{b}\mathbf{a}^\dagger)$ can only be reconstructed up to a constant factor that can be absorbed into the respective diagonal. This is analogous to "standard" ICA and, indeed, also to the singular value decomposition, which requires normalization and sign conventions to be unique.

For the noisy and coupled systems, the typical relative error in the spatial structure is around 5%. Simulations with a larger number of channels (and corresponding number of noise sources) showed even better results if we choose the same total power ratios.

B. Real EEG data

After this proof of concept, we applied our method to real data gathered in 118 EEG channels during imagined left or right hand or foot movement from [11]. For each of the randomly mixed conditions, 70 imaginations, each lasting for ~3.5 s, were performed. Cross-spectra were calculated separately for each condition by linear detrending and Hanning-windowing the 3.5 s windows and averaging over the respective 70 trials. To improve the signal-to-noise ratio, we additionally averaged the cross-spectra across neighboring frequencies with a Hanning window of width 11, leading to an effective frequency resolution of ~1 Hz. The imaginary parts of the normalized cross-spectra were then diagonalized simultaneously for frequencies from 0 to 45 Hz.

In Fig. 2, we show EEG power for imagined foot movement as a function of frequency in two selected channels and as function of location for two selected frequencies. We observe peaks at 10 and 12 Hz but only very small ones at 20 or 24 Hz.

We found 4 components containing clear spectral peaks in the diagonals. Two of them are shown in the left and right panels of Fig. 3, respectively. We recall that the spatial pattern of each component corresponds to the two-dimensional subspace spanned by the real and imaginary parts of the respective column of the complex demixing matrix. To visualize the respective subspace, we chose a basis (top-left and middle-left panel for the first component) guided by the idea that each pattern suggests an explanation as simple as possible, preferably in terms of source dipoles (here: radial for top-left and tangential for middle-left panel). The corresponding diagonals are shown in the lower-left panel. This component with a peak at 10 Hz and higher harmonic at 20 Hz clearly corresponds to occipital alpha activity. The second component is shown in the right panels. The spectral content with a peak at 12 Hz and harmonic at 24 Hz clearly represents central mu rhythm. The spatial patterns indicate an interaction between right motor area and deeper brain structures, which has been modeled in [12]. However, the detailed physiological analysis of this phenomenon is beyond the scope of this paper. We finally note that the other two components with clear peaks in the spectrum, which were not shown, roughly correspond to left-right mirror images of the shown components with similar spectral content.

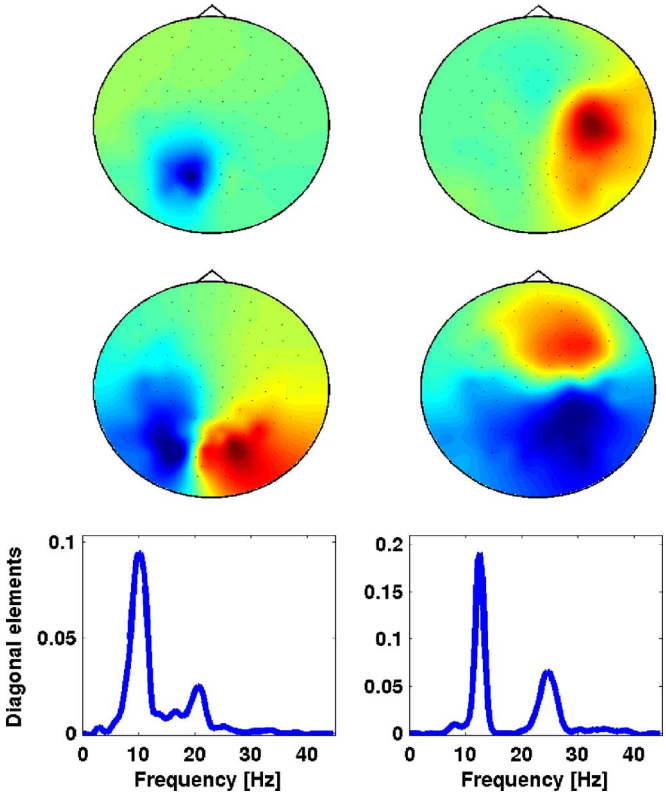

FIG. 3. (Color online) Left: Two basis fields (top and middle) and diagonal spectrum of component corresponding to occipital α. Right: same for central mu rhythm.

To estimate whether a given ISA component truly corresponds to an interacting as opposed to a random fluctuations, we need to test whether a diagonal element at a specific frequency is significantly different from zero. It is a potential risk that the ISA decomposition collects those channels that merely appear to be interacting. Probably the safest way to avoid such a bias in the significance analysis is to divide the data set into two parts and to construct the spatial patterns

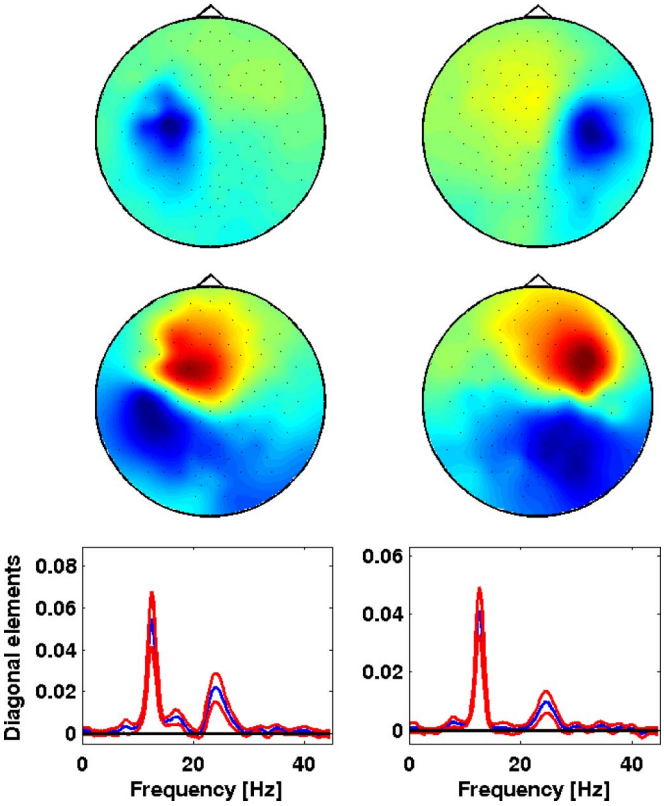

FIG. 4. (Color online) Spatial patterns of μ components calculated from the first half of the data (top and middle row) and spectra±twice the SEM indicated in red calculated from the second half of the data (bottom row). The peaks at 12 Hz differ from zero by 9.42 and 11.43 standard errors of the mean, respectively. The corresponding p values are below 10^{-15}.

from the first part and test the result for significance on the second part. This was done for the real EEG data set with results for the two mu components shown in Fig. 4. The spatial patterns again show activity over central motor areas and the spectra show clear and highly significant peaks around 12 and 14 Hz.

IV. CONCLUSION

In conclusion, when analyzing interaction between sources from macroscopic measurements that are linear mixtures, it is important to distinguish meaningful patterns of interaction from spurious ones. In particular, for EEG/MEG measurements, volume conduction effects make large parts of the human brain seemingly interact although in reality such contributions are purely artifactual. Existing blind source separation (BSS) methods that have been used with success for artifact removal and for estimation of brain sources will by construction fail when attempting to separate interacting i.e. non-independent brain sources. In this work we have proposed a BSS algorithm that uses antisymmetrized cross-correlation or cross-spectral matrices and subsequent diagonalization and can thus reliably extract meaningful interaction while ignoring all spurious effects. We want to emphasize that our method is blind to interactions that do not contain time delays and/or are completely symmetric. However, such interactions are indistinguishable from artifacts of volume conduction, and the partial blindness of our method is the unavoidable price to detect true interactions.

Experiments using our interacting source analysis (ISA) reveal interesting relationships that are found *blindly*. These findings exemplify that ISA is a powerful new technique when analyzing interactions in macroscopic brain measurements. Future studies will therefore apply ISA to other neurophysiological paradigms in order to gain insight into the coherence and synchronicity patterns of cortical dynamics.

ACKNOWLEDGMENTS

We acknowledge partial funding from the Bundesministerium für Bildung und Forschung (Grant No. 01IBE01A and Grant No. BCCNB-A4 01GQ0415). This work has also been supported, in part, by the DFG within the research group on "conflicting rules in cognitive systems," and the IST Programme of the European Community, under the PASCAL Network of Excellence, Grant No. IST-2002-506778. This publication only reflects the authors' views.

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis* (Wiley, New York, 2001).
[2] V. Jirsa, Neuroinformatics **2**, 183 (2004).
[3] W. Singer, Nature (London) **397**, 391 (1999).
[4] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization—A Universal Concept in Nonlinear Sciences Neocortical* (Oxford University Press, London, 2001).
[5] K. T. Dolan and A. Neiman, Phys. Rev. E **65**, 026108 (2002).
[6] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, Clin. Neurophysiol. **115**, 2292 (2004).
[7] F. Meinecke, A. Ziehe, J. Kurths, and K.-R. Müller, Phys. Rev. Lett. **94**, 084102 (2005).
[8] L. Molgedey and H. G. Schuster, Phys. Rev. Lett. **72**, 3634 (1994).
[9] A. Ziehe and K.-R. Müller, in *Proceedings of the 8th International Conference on Artificial Neural Networks, ICANN'98*, edited by L. Niklasson, M. Bodén, and T. Ziemke (Springer Verlag, Berlin, 1998), Perspectives in Neural Computing, pp. 675–680.
[10] A. Yeredor, A. Ziehe, and K.-R. Müller, in *Lecture Notes in Computer Science*, edited by C. G. Puntonet and A. Prieto (Springer-Verlag, Granada, 2004), Vol. 3195, pp. 89–96, Proc. ICA 2004.
[11] B. Blankertz, G. Dornhege, C. Schäfer, R. Krepki, J. Kohlmorgen, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio, IEEE Trans. Neural Syst. Rehabil. Eng. **11**, 127 (2003).
[12] P. Suffczynski, S. Kalitzin, G. Pfurtscheller, and F. Lopes da Silva, Int. J. Psychophysiol **43**, 25 (2001).