Measuring directional coupling between EEG sources

German Gómez-Herrero, [a] Mercedes Atienza, [b]

Karen Egiazarian [a] and Jose L. Cantero* [b]

[a] *Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FIN-33101, Tampere, Finland.*

[b] *Laboratory of Functional Neuroscience, University Pablo de Olavide, Carretera de Utrera, Km. 1, 41013 Seville, Spain.*

* Correspondence should be addressed to:

José L. Cantero, Ph. D.

Laboratory of Functional Neuroscience

University Pablo de Olavide

Ctra. de Utrera, Km. 1

41013-Seville, Spain

Phone: +34-954-977433

Fax: +34-954-349151

E-mail: jlcanlor@upo.es

Running title: Directional coupling and EEG sources

Abstract

Directional connectivity in the brain has been typically computed between scalp electroencephalographic (EEG) signals, neglecting the fact that correlations between scalp measurements are partly caused by volume conduction through the brain tissues and the skull. Although recently proposed techniques are able to identify causality relationships between EEG sources rather than between recording sites, most of them need a priori assumptions about the cerebral regions involved in the EEG generation. We present a novel methodology based on multivariate autoregressive (MVAR) modeling and Independent Component Analysis (ICA) able to determine the temporal activation of the intracerebral EEG sources as well as their approximate locations. The direction of synaptic flow between these EEG sources is then estimated using the directed transfer function (DTF), and the significance of directional coupling strength evaluated with surrogated data. The reliability of this approach was assessed with simulations manipulating the number of data samples, the depth and orientation of the equivalent source dipoles, the presence of different noise sources, and the violation of the non-Gaussianity assumption inherent to the proposed technique. The simulations showed the superior accuracy of the proposed approach over other traditional techniques in most tested scenarios. Its validity was also evaluated analyzing the generation mechanisms of the EEG alpha rhythm recorded from 20 volunteers under resting conditions. Results suggested that the major generation mechanism underlying EEG alpha oscillations consists of a strong bidirectional feedback between thalamus and cuneus. The precuneus also seemed to actively participate in the generation of the alpha rhythm although it did not exert a significant causal influence neither on the thalamus nor on the cuneus. All together, these results suggest that the proposed methodology is a promising non-invasive approach for studying directional coupling between mutually interconnected neural populations.

Key words: Directional coupling, EEG sources, Multivariate Autoregressive Model (MVAR), Independent Component Analysis (ICA), Alpha rhythm.

1 Introduction

The ability of neuronal populations to establish oscillatory synchronization patterns at local and long-range levels has been potentially regarded as one of the brain mechanisms underlying cognition (e.g. Uhlhass and Singer, 2006). But in highly interconnected cerebral systems it is relevant not only to determine neuronal synchronization, but also to identify causal (drive-response) relationships between the studied subsystems. Analysis techniques based on Granger causality criteria provide information about which brain region drives another by measuring how the history of a neural signal predicts the future of another (Granger, 1969; Baccalá and Sameshima, 2001). A closely related concept consists in measuring to what extent an spectral component in a neural signal induces the generation of the same spectral component in another neural signal (Kamiński and Blinowska, 1991; Eichler, 2006). To date, all these information-flow measures have been typically computed between scalp EEG signals recorded in humans (Baccalá and Sameshima, 2001; Kamiński and Blinowska, 1991; Kús et al., 2004) and from intracranial recordings both in humans and animals (Franaszczuk and Bergey, 1998; Freiwald, 1999; Kamiński et al., 2001). More recently, they have started to be applied in brain source space (Bakardjian et al., 2006; Astolfi et al., 2007b; Supp et al., 2007).

A fundamental problem when studying directional connectivity between brain areas is that causality relationships between scalp EEG signals do not imply that the same relationships exist between the underlying neural sources. The

reason is that scalp EEG potentials do not exclusively reveal averaged post-synaptic activity from localized cortical regions beneath one electrode but the superposition of all active coherent neural sources located anywhere in the brain, due to conduction effects in the brain volume (Malmivuo and Plonsey, 1995). This superposition effect inevitably leads to misinterpretations of the synchronization results obtained between scalp EEG signals, specially when subcortical generators are actively involved.

A straightforward solution to these problems consists in computing the coherence between equivalent intracranial current dipoles (Hoechstetter et al., 2004). However, estimating the number, locations and orientations of multiple discrete neuroelectric dipoles is not a trivial issue due to the high number of possible model configurations that fit well the spatial patterns of scalp EEG potentials. A similar approach consists in computing the coherence between brain regions of interest (Lehmann et al., 2006; Gross et al., 2004; Supp et al., 2007), although defining the anatomical boundaries of these areas usually involves a great deal of subjectivity and an *a priori* knowledge about the location of the neural current generators underlying scalp EEG recordings.

An alternative solution to volume conduction effects is based on blind source separation (BSS) techniques. Rather than estimating the locations of the source generators, BSS aims to directly invert the mixing process that the source signals undertook in the brain tissue. To achieve this separation, these methods make mild assumptions about the mixing process, namely that the quasistatic approximation of brain volume conduction is valid (Malmivuo and Plonsey, 1995), and impose strong requirements on the statistical properties of the source signals. A prototypical example of BSS techniques is Independent Component Analysis (ICA), which assumes that the sources are mutually in-

dependent non-Gaussian random variables (see e.g. Hyvärinen et al., 2001, for a review). Multiple studies have shown the usefulness of ICA techniques to remove artifacts (e.g. Jung et al., 2000), to separate physiological sources (e.g. Makeig et al., 2002) and even to study directional connectivity between cortical areas (Bakardjian et al., 2006; Astolfi et al., 2007a). However, if we aim at determining causality relationships between different spatiotemporal EEG sources, we are implicitly assuming that those sources are functionally interrelated, which violates the assumption of independence made by ICA. Several methods have been previously proposed to overcome this pitfall (Meinecke et al., 2005; Nolte et al., 2006), but they are not able to identify directionality, i.e. asymmetric flows of activity between EEG sources.

In this article we propose a novel analysis methodology to overcome volume conduction effects when studying directional connectivity between EEG sources. By carefully combining MVAR modeling and ICA, our approach is able to unravel the mixing caused by volume conduction and estimate the original source signals and their corresponding spatial patterns of scalp potentials. Subsequently, these spatial patterns are localized within the brain volume by the standardized weighted low-resolution brain electromagnetic tomography (swLORETA) (Palmero Soler et al., 2007). Finally, we estimate the direction of synaptic flow between these EEG sources using the directed transfer function (DTF, Kamiński and Blinowska, 1991) and assess the significance of the obtained DTF values using surrogated data. This approach is conceptually simple, it can be applied to high-density EEG datasets, is able to identify causal directionality between EEG sources and does not make any prior assumptions about the locations of the EEG generators. We confirmed the superior accuracy of the proposed methodology over other approaches by

means of simulations, and further assessed its validity through the analysis of the generation mechanisms of the EEG alpha rhythm in 20 volunteer subjects.

2 Materials and methods

2.1 EEG model

The time-varying neural current density responsible for the EEG scalp potentials can be appropriately modeled by a discrete set of K signal generators. Let us denote by $\mathbf{s}(t) = [s_1(t), ..., s_K(t)]^T$ the multivariate activation pattern of those generators at time instant t. Here, as in the following, ^{T} denotes transposition. Let us also assume that those source activation patterns are well described by an MVAR model of order p, at least for a certain temporal range $t = 1, ..., L$. Then, we have that:

$$\mathbf{s}(t) = \sum_{\tau=1}^{p} \mathbf{B_s}(\tau)\mathbf{s}(t-\tau) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{B_s}(\tau)\ \forall\tau = 1, ..., p$ are the coefficient matrices of the MVAR model and $\mathbf{n}(t) = [n_1(t), ..., n_K(t)]^T$ represents the corresponding multivariate *residual process*. We assume that each EEG generator is a source of *independent* activity in the sense that the elements of the residual vector $\mathbf{n}(t)$ behave like mutually independent random variables. Causal relationships between EEG sources are therefore exclusively produced by time-lagged axonal propagation of macroscopic neural behavior among distant regions of the brain (modeled by the coefficient matrices $\mathbf{B_s}(\tau)\ \forall\tau = 1, ..., p$).

The DTF and also several Granger-causality measures can be directly com-

puted from estimates of $\mathbf{B_s}(\tau)$, as explained in the following section. However, scalp EEG recordings do not correspond with the multivariate vector $\mathbf{s}(t)$ but with a (noisy) linear mixture of its components. Therefore, the multichannel EEG recorded at time instant t using M electrodes is a multivariate signal $\mathbf{x}(t)$ satisfying:

$$\mathbf{x}(t) = \mathbf{\Phi}\mathbf{s}(t) + \boldsymbol{\eta}(t) + \mathbf{\Phi}_{noise}\boldsymbol{\epsilon}(t) \tag{2}$$

where $\mathbf{\Phi}$ is an unknown $M \times K$ leadfield matrix modeling volume conduction effects from the location of the sources to the scalp electrodes, $\boldsymbol{\eta}(t) = [\eta_1(t), ..., \eta_M(t)]^T$ denotes the additive noise at each electrode (*measurement noise*), which is assumed to be white and Gaussian, and the term $\mathbf{\Phi}_{noise}\boldsymbol{\epsilon}(t)$ represents the contribution of noisy EEG sources (*biological noise*). By noisy sources we refer to EEG sources unrelated to the EEG features under study (e.g. unrelated to the generation of the EEG-alpha rhythm). We only consider the case when $M \geq K$. Specially for high-density EEG recordings, it is relatively safe to assume that the amount of electrodes is greater than the number of neural sources contributing to the scalp EEG. Furthermore, for the sake of simplicity, we neglect the contribution of measurement and biological noise in the derivations below. By combining Eq. 1 and the noiseless version of Eq. 2, we obtain that the observed EEG follows the MVAR model:

$$\mathbf{x}(t) = \sum_{\tau=1}^{p} \underbrace{\mathbf{\Phi}\mathbf{B_s}(\tau)\mathbf{\Phi}^{+}}_{\mathbf{B_x}(\tau)} \mathbf{x}(t-\tau) + \underbrace{\mathbf{\Phi}\mathbf{n}(t)}_{\mathbf{v}(t)} \tag{3}$$

where $^+$ denotes Moore-Penrose pseudoinversion [1]. Granger causality stud-

[1] If $\mathbf{M}$ is a real matrix and $\left(\mathbf{M}^T\mathbf{M}\right)^{-1}$ exists then $\mathbf{M}^+ = \mathbf{M}^T\left(\mathbf{M}\mathbf{M}^T\right)^{-1}$

ies applied to human EEG signals have typically used estimates of $\mathbf{B_x}(\tau)$ to measure directional flows of macroscopic synaptic activity between *scalp EEG electrodes*, with the implicit assumption that causal relationships between electrodes imply functional connectivity between their respective underlying cortical regions. This is equivalent to assuming that $\mathbf{B_x}(\tau) \approx \mathbf{B_s}(\tau)$. Indeed, this assumption is not valid in general because the MVAR model that best fits the observed EEG data might be strongly affected by volume conduction effects (matrix $\mathbf{\Phi}$ in Eq. 3).

2.2 Analysis procedure: MVAR-EfICA

We propose a novel methodology for measuring directional connectivity that takes into consideration the effects of volume conduction. In order to achieve this goal, the source MVAR coefficient matrices $\mathbf{B_s}(\tau)\ \forall\tau = 1, ..., p$ are estimated based on the observed EEG given in Eq. 3. Our approach, called *MVAR-EfICA*, is depicted in Fig. 1 and consists of the following steps:

2.2.1 Principal Component Analysis (PCA)

We first apply PCA because previous studies have demonstrated its efficacy in integrating brain activity spread across EEG leads, reducing the effects of measurement noise and removing second-order instantaneous cross-correlations caused by volume conduction (see e.g. Richards, 2004). Even more important, PCA reduces the dimensionality of the data and avoids ill-conditioned covariance matrices, which results in a faster and more robust estimation of the MVAR model in the next analysis step.

PCA linearly transforms the scalp EEG signals $\mathbf{x}(t)$ into a set of K mutually uncorrelated principal components. From Eq. 3 follows that the PCA-transformed data is also an MVAR process:

$$\mathbf{x}_{PCA}(t) = \sum_{\tau=1}^{p} \underbrace{\mathbf{C\Phi B_s}(\tau)\left(\mathbf{C\Phi}\right)^{-1}}_{\mathbf{B}_{PCA}(\tau)} \mathbf{x}_{PCA}(t-\tau) + \underbrace{\mathbf{C\Phi n}(t)}_{\mathbf{r}(t)} \tag{4}$$

where $\mathbf{C}$ is a $K \times M$ matrix implementing the PCA transformation. We suggest using as many principal components as necessary to reconstruct most (e.g. 99%) of the variance contained in the EEG signals.

2.2.2 Multivariate Autoregressive (MVAR) modeling

Time-delayed cross-dependencies between the estimated principal components cannot be explained by instantaneous volume conduction in the brain tissue and, therefore, are likely to be of neural origin, i.e. caused by time-delayed axonal propagation of electrical activity between brain generators. To disclose the dynamical characteristics of these time-delayed covariances, an MVAR model is fitted to $\mathbf{x}_{PCA}(t)$ using the algorithm ARfit (Schneider and Neumaier, 2001). ARfit is suitable for the analysis of high-dimensional datasets and, for relatively large samples sizes, similarly accurate to other MVAR estimation algorithms (Schlögl, 2006). We denote by $\hat{\mathbf{B}}_{PCA}(\tau)\ \forall\tau = 1, ..., p$ the estimated model coefficients. The model order is automatically selected using Swartz's Bayesian Criterion (SBC) (Schwarz, 1978). Traditionally, it is determined by locating the minimum of the SBC as a function of model order (p values varied between 2-30 in the present study). Instead, we select the order that guarantees that greater model orders do not significantly reduce the SBC, i.e. the order for which the reduction in the SBC has reached 90% of

the maximum reduction achievable within the tested range of model orders. This modification is motivated by the fact that, for our EEG data, both SBC and Akaike's information criterion (Akaike, 1971) dropped monotonically with increasing model order, which is in close correspondence with previous EEG studies (Brovelli et al., 2004; Supp et al., 2007).

2.2.3 Independent Component Analysis (ICA)

The instantaneous (without delay) higher-order cross-dependencies remaining in the residuals of the MVAR model that were estimated in the previous step are likely to be of non-neural origin, i.e. caused by volume conduction effects [2]. Based on this assumption, we use ICA to estimate a $K \times K$ matrix $\hat{\mathbf{W}} \approx (\mathbf{C}\mathbf{\Phi})^{-1}$ that minimizes the mutual dependencies between the components of the multivariate residual process $\mathbf{r}(t)$ previously obtained [3]. Specifically, we use the algorithm EfICA (Koldovský et al., 2006), which is an asymptotically optimal variant of the popular algorithm FastICA (Hyvärinen, 1999). The idea of applying ICA to the residuals of a linear predictor has been proposed elsewhere (Hyvärinen, 2001; Särelä and Vigário, 2003), although in an univariate context. A combination of MVAR modeling and ICA was previously proposed in the field of audio signal processing (Cheung and Xu, 2003).

[2] Assuming that there are not instantaneous (without delay) flows of activity between brain generators. This is a plausible assumption since instantaneous flows between brain EEG sources located more than few mm apart are unlikely due to axonal propagations delays (Freeman, 2000).

[3] Note that the permutation and sign indeterminacy of ICA becomes irrelevant once the brain sources are localized in section 2.2.5

2.2.4 DTF computation between EEG sources

We use $\hat{\mathbf{W}} \approx (\mathbf{C\Phi})^{-1}$ and Eq. 4 to estimate the MVAR parameters of the underlying EEG sources: $\hat{\mathbf{B}}_{\mathbf{s}}(\tau) = \hat{\mathbf{W}}\hat{\mathbf{B}}_{PCA}(\tau)\hat{\mathbf{W}}^{-1} \approx \mathbf{B}_{\mathbf{s}}(\tau)$. Then, the spatiotemporal spectral properties of the EEG sources can be uncovered by transforming Eq. 1 to the frequency domain:

$$\mathbf{S}(f) = \underbrace{\left(\mathbf{I} - \sum_{\tau=1}^{p} \mathbf{B}_s(\tau) e^{-j2\pi\tau\frac{f}{f_s}}\right)^{-1}}_{\mathbf{H}(f)} \cdot \mathbf{N}(f) \tag{5}$$

where $\mathbf{N}(f)$ is the Fourier transform of the residual process $\mathbf{n}(t)$, f_s is the sampling frequency and $\mathbf{I}$ denotes the identity matrix. The transfer matrix $\mathbf{H}(f)$ describes transfer of spectral properties (coherent links) between EEG sources.

The DTF (Kamiński and Blinowska, 1991) from the jth EEG source to the ith source at certain frequency f is denoted by $\gamma_{ij}(f)$ and is defined as the ratio of influence of $s_j(t)$ on $s_i(t)$, with respect to the combined influence of $s_1(t), ..., s_K(t)$ on $s_i(t)$. The DTF can be compactly expressed in terms of elements of the transfer matrix $\mathbf{H}(f)$ as:

$$\gamma_{ij}(f) = \frac{|H_{ij}(f)|^2}{\sum_{m=1}^{K} |H_{im}(f)|^2} \tag{6}$$

DTF values are in the interval $[0,\ 1]$. The DTF can also be defined for a band of frequencies, e.g. for the alpha-rhythm frequency band ($7.5 - 12.5$ Hz) by integrating the DTF across alpha frequencies: $\gamma_{ij}(7.5 \text{ Hz} < f < 12.5 \text{ Hz}) = \int_{7.5 \text{ Hz}}^{12.5 \text{ Hz}} \gamma_{ij}(f)df$.

Although the DTF is a good indicator of the total spectral influence from one electromagnetic source to another, it is noteworthy mentioning that knowledge of the MVAR model underlying the source signals is all we need to compute other alternative measures of information flow in the frequency domain. An example of such measures is the partial directed coherence (PDC) (Baccalá and Sameshima, 2001) which, contrary to the DTF, allows differentiating between direct and indirect flows between EEG sources. Both DTF and PDC have somewhat complementary advantages and limitations (Kús et al., 2004; Eichler, 2006). Probably, the best approach would be to compute several such complementary measures simultaneously (Eichler, 2006), which is a topic of further research. The DTF was chosen here only for illustration purposes, although any other MVAR-based information flow measure could be used in combination with the proposed MVAR-EfICA methodology.

2.2.5 Intracranial localization of EEG sources

We estimate the leadfield matrix as $\hat{\mathbf{\Phi}} = \left(\hat{\mathbf{W}}\mathbf{C}\right)^{+} \approx \mathbf{\Phi}$. Each of the K columns of the estimated leadfield matrix $\hat{\mathbf{\Phi}}$ determines the brain location where each source signal is generated. Practically any inverse method in the literature could be used to estimate the location of these EEG sources. Given that the generation of human alpha rhythm involves cortical (surface) and thalamic (deep) sources, we applied the swLORETA method, which has shown a better performance when compared with other similar techniques under realistic noisy conditions and for the reconstruction of deep EEG sources (Palmero Soler et al., 2007). In the present study we employed a realistic head model of three layers (scalp, skull and brain with conductivities of 0.33, 0.0042, and 0.33, respectively) created using the boundary element method. Source reconstruc-

tion solutions were projected onto the 3D MR images of the Collin's brain provided by the Montreal Neurological Institute. Probabilities of source activation based on Fisher's F-test were obtained for each independent EEG source.

2.3 Alternative approaches to MVAR-EfICA

Previous studies of directional connectivity have commonly assumed that causality relationships between scalp EEG signals are equivalent to functional connectivity between the cortical regions underlying the corresponding EEG electrodes, completely disregarding volume conduction effects (Bernasconi and König, 1999; Baccalá and Sameshima, 2001; Kamiński et al., 2001; Kús et al., 2004). If the locations of the EEG sources are known a priori, a conceptually equivalent approach consists in fitting an MVAR model only to the signals recorded from *the electrodes closest to the underlying EEG sources*. We refer to this traditional methodology as the *low-dimensional MVAR* approach. Actually, *low-dimensional MVAR* can be regarded as an upper bound for the performance of traditional DTF estimators since the locations of the underlying neuroelectrical sources are typically unknown when dealing with real scalp EEG data.

Astolfi et al., (2007b) used ThinICA to remove volume conduction effects before estimating the PDC. The DTF and the PDC are both based on MVAR modeling and therefore, the same approach can be directly extrapolated to the estimation of the DTF. The approach of Astolfi and colleagues differs from ours in two important facts. First, it applies ThinICA on the principal components directly, rather than on the residuals of the MVAR model best fitting the

principal components. Second, ThinICA is not a traditional ICA algorithm, at least according to the classical definition of ICA. Thus, ThinICA does not require that the sources are approximately non-Gaussian random variables but they are also allowed to be Gaussian processes with temporal structure and zero time-lagged cross-correlations. As will be shown later through simulations, the presence of time-delayed cross-correlations between EEG sources has an important negative effect on the accuracy of ThinICA and, therefore, on the estimation of the DTF. In the simulations we will refer to this method simply as *ThinICA-MVAR*. Since Astolfi et al., (2007b) did not specify the number of time-delayed covariance matrices used by ThinICA, we tested two possibilities: *ThinICA1-MVAR*, which uses only a single covariance matrix for time-lag 0 and *ThinICA5-MVAR*, which uses covariances for time-lags $0, 1, 2, 3, 4$.

2.4 Simulations

The proposed methodology was evaluated by studying the effects of the following simulation parameters: L (number of scalp EEG data samples), d (depth of the source dipoles, using the head radius as reference unit), δ (orientation angle of the source dipoles, in degrees), SNR (signal-to-measurement-noise ratio, in dBs), SBNR (signal-to-biological noise ratio, in dBs) and α (Gaussianity of the residuals). Each tested set of values for those parameters resulted in a simulation instance. In order to assess average performance, each simulation instance was repeated 200 times with random spatial and temporal characteristics of the noise. In addition, any parameter that was not under evaluation during certain simulation instance was either fixed to some favorable values ($L = 38400$, $SNR = 15\ dB$, $SBNR = 15\ dB$) or varied randomly within

certain plausible ranges ($0.2 \leq d \leq 0.8$, $0^\circ \leq \delta \leq 90^\circ$, $1 \leq \alpha \leq 3$). This simulation set-up allowed us to assess the average effects of each tested parameter for an arbitrary combination of the remaining parameters.

We used a one-shell spherical head of normalized radius ($r = 1$). Inside this simulated head we located four *source dipoles* $\vec{\mathbf{q}}_j(t) = s_j(t) \cdot \vec{\mathbf{e}}_j \quad \forall j = 1, 2, 3, 4 \quad \forall t = 1, ..., L$ with orientations $\vec{\mathbf{e}}_j = [e_{j,x}, e_{j,y}, e_{j,z}]^T$ and time-varying activations $s_j(t)$ following an MVAR model as in Eq. 1. Both the residuals and the coefficients of the MVAR model (the coupling strengths between sources) were random in each simulation surrogate, with the only requirement that the resulting MVAR model had to be stable. In order to assess the amount of data needed by different techniques to reach their asymptotic performance, we ran simulations using $1280 < L < 38400$ samples. Fig. 2 (left panel) and Fig. 3 show the positions of the simulated electrodes ($M = 16$) and the positions and orientations of the simulated source dipoles ($K = 4$). As shown in Fig. 3, the spatial characteristics of the source dipoles could be varied using two parameters: the distance to the scalp (d) and the angle that they formed with their radial component (δ). The rotation angles of the dipoles around their radial components also affected the performance of the different methods but the effect of the angle δ was much more pronounced. Because of this, only δ was used to vary the orientation of the dipoles and the rotation angles of the dipoles were random in each simulation surrogate.

The residual processes driving the *source dipoles*, denoted by $n_1(t), ..., n_4(t)$ in Eq. 1, followed a generalized Gaussian distribution with probability density function (pdf): $p(n_i) \propto e^{-\Gamma|n_i|^\alpha}$. By varying α between 1 and 3 we varied the Gaussianity of the residuals. If $\alpha < 2$, they showed a super-Gaussian distribution (acute peak around the mean and "fat" tails). If $\alpha = 2$ they were

exactly Gaussian, and if $\alpha > 2$ they were sub-Gaussian (smaller peak around the mean and "thin" tails).

The signal-to-measurement-noise ratio (SNR) and the signal-to-biological-noise ratio ($SBNR$) were defined as the mean standard deviation of the signal across EEG channels, divided by the standard deviation of the Gaussian thermal noise introduced by the EEG sensors (*measurement noise*) and by the mean standard deviation of the *biological noise* across EEG channels, respectively. The biological noise was simulated using four *noisy dipoles* which were randomly located in each simulation surrogate within the volume depicted in Fig. 2 (right panel). The orientations of the noisy dipoles also varied randomly across simulation surrogates and their individual temporal activations were obtained with an autoregressive (AR) model of order 5 driven by white Gaussian residuals. Both the residuals and the coefficients of these AR models were randomized in each simulation surrogate.

The MATLAB® code necessary to perform the simulations above is freely available in the Internet (Gómez-Herrero, 2008).

2.5 Assessing the accuracy of DTF estimates obtained by different methods

The estimation accuracy for the source DTF between the i^{th} and j^{th} dipole at frequency f was measured using the following index (in percentage):

$$\epsilon = 100 \cdot \frac{1}{K^2 N} \sum_{ijf} \sqrt{\left(\gamma_{ij}(f) - \hat{\gamma}_{ij}(f)\right)^2} \tag{7}$$

where $\gamma_{ij}(f)$ denotes the true DTF between the i^{th} and j^{th} simulated dipoles

(obtained from the true MVAR coefficients in Eq. 1), $\hat{\gamma}_{ij}(f)$ designates the DTF estimation obtained with any of the tested methods, K is the total number of source dipoles (in our case, $K = 4$) and N is the number of frequency bins employed in the DTF analysis (in our simulations, $N = 10$ uniformly distributed bins). Since the DTF at a certain frequency is within the range [0 1], ϵ ranges from 0% (best case, no estimation error) to 100% (worst case, maximum possible estimation error).

As was said in the previous section, each simulation *instance* (corresponding to certain set of values of the simulation parameters) was repeated 200 times. The results obtained from these simulation surrogates were summarized by means of two indexes:

(1) The mean DTF estimation error, defined as $\bar{\epsilon} = \frac{1}{200}\sum_1^{200}\epsilon_i$, where ϵ_i is the DTF error obtained in the ith simulation surrogate. Clearly, $\bar{\epsilon}$ varies between 0% (perfect DTF estimation in every surrogate) and 100% (worst possible estimation in every surrogate).

(2) The p-value of the paired t-test for the null hypothesis $\bar{\epsilon}_0 < \bar{\epsilon}$, where $\bar{\epsilon}_0$ is the mean DTF estimation error obtained with the traditional *low-dimensional MVAR* methodology and $\bar{\epsilon}$ is the mean DTF error for an alternative method. We denoted this performance index by $\bar{\epsilon}_{\text{p-val}}$. This index evaluates the probability (p-value) that the traditional approach performs better than *MVAR-EfICA* and/or *ThinICA-MVAR* in a random simulation surrogate. Therefore, the lower the value of $\bar{\epsilon}_{\text{p-val}}$ the more significant the improvement obtained with any of the tested methods with respect to the traditional approach.

EEG recordings were obtained from 20 cognitively normal volunteers (10 females, mean age: 68.4 ± 6.1 yr) recruited from the local community. Inclusion criteria for this study were (i) no cognitive dysfunctions corroborated by neuropsychological exploration, (ii) clinical dementia rating global score of 0 (no dementia), and (iii) normal independent function both judged clinically and by means of a standardized scale for the activities of daily living. None of them had a history of neurological, psychiatric disorders and/or major medical illness.

EEG recordings were obtained between 9-10 AM in all participants in a relaxed wakefulness state with eyes closed. Vigilance level was constantly controlled. EEG was continuously acquired and referenced to linked mastoids from 59 scalp locations according to the International 10-20 system. Vertical ocular movements were recorded with a pair of electrodes placed above and below the left eye. The horizontal electrooculogram was acquired with another pair 1 cm apart from the outer canthus of each eye. Electrophysiological measurements were recorded with 10 mm diameter gold disk electrodes (Grass, Inc.). Electrode-scalp impedance was always kept below 5 $K\Omega$. All electrophysiological variables were amplified (BrainAmp MR, Brain Vision®), filtered ($0.1 - 100$ Hz bandpass), digitized (250 Hz, 16-bit resolution), and stored in digital format for off-line analysis.

EEG epochs containing prominent ocular, muscular and/or any other type of artifacts were manually identified and eliminated by an experimented researcher (JLC) with expertise in human neurophysiology. Residual ocular ar-

tifacts present in the remaining EEG epochs were corrected by adaptively regressing out the signals recorded at the peri-ocular electrodes (He et al., 2004). A total of 150 s (37500 samples) of artifact-free EEG containing alpha rhythm were then available for each participant. The selected epochs were filtered within $6 - 13$ Hz using a real and phase linear passband filter of order 100. The average DTF for alpha-EEG frequencies was obtained as the average of the DTF values obtained in 10 equally spaced frequency bins within the alpha band (7.5 Hz - 12.5 Hz).

2.7 Finding reliable results with real EEG data

In the analysis of real EEG data, we used ICASSO (Himberg et al., 2004; Himberg and Hyvärinen, 2005) to assess the significance of the ICA estimates obtained in the *MVAR-EfICA* procedure. EfICA, FastICA and most ICA algorithms involve stochastic optimization, which raises concerns about the repeatability and reliability of the analysis when analyzing real data (Särelä and Vigário, 2003; Himberg et al., 2004). ICASSO overcomes these concerns by identifying clusters of ICA estimates that are consistently found across random initializations of the ICA algorithm and across random bootstrap surrogates of the input data.

We introduced three modifications in the standard ICASSO procedure originally proposed by (Himberg et al., 2004). First, we used EfICA as ICA algorithm instead of FastICA, which was motivated by the better performance of the former reported by (Koldovský et al., 2006). Second, ICA estimates were clustered according to the cross-correlation coefficients between their corresponding spatial patterns of scalp potentials. Third, we clustered together

ICA estimates obtained from all available subjects. This latter modification allowed the detection of stable clusters of spatial components across the group of studied subjects. ICASSO was set-up to run EfICA 75 times on each subject's EEG with random initial conditions and with random bootstrap re-sampling. We used ICASSO's default agglomerative clustering with average-linkage criterion. The number of clusters in the data was automatically selected with the R-index (Himberg et al., 2004). Each cluster was uniquely represented by a single *centrotype* ICA-estimate, which is just the estimate in the cluster that has the maximum sum of similarities to other points in the cluster. Only centrotypes of significant clusters were considered as valid ICA-estimates. A cluster was significant if it contained ICA-estimates from at least 50% of the subjects (high inter-subject repeatability) and from at least 50% of the ICA-runs corresponding to those subjects (high intra-subject reliability).

Since the significant centrotype ICA-estimates might have been obtained from different subjects and/or different ICA runs, we had to redefine the computation of the DTF between EEG sources and we had to find a unique approximation of the leadfield matrix for the whole population of subjects. This was done by noting that the 1-dimensional temporal activation of the i^{th} centrotype EEG source, denoted by $s_i^*(t)$, can be retrieved from the EEG measurements of the corresponding subject by:

$$s_i^*(t) = \mathbf{v}_i^* \mathbf{x}_{(i)}(t) \qquad \forall i = 1, ..., K \tag{8}$$

where $\mathbf{v}_i^*$ is a spatial filter (a row vector of scalar coefficients) and $\mathbf{x}_{(i)}(t)$ denotes the column vector containing the EEG measurements at the time instant t for the subject from which the i^{th} centrotype ICA-estimate was obtained. The temporal course of the i^{th} centrotype ICA-estimate can be projected from

brain space back to the scalp EEG sites by:

$$\mathbf{x}_{(i)}^{*i}(t) = \boldsymbol{\phi}_i^* s_i^*(t) \qquad \forall i = 1, ..., K \tag{9}$$

where $\boldsymbol{\phi}_i^*$ is a column vector of scalar coefficients defining the pattern of scalp potentials generated by the i^{th} centrotype EEG source. Therefore, the intracranial localization of the i^{th} centrotype EEG source is fully characterized by $\boldsymbol{\phi}_i^*$ whereas the spatial filter $\mathbf{v}_i^*$ can be used to retrieve its temporal dynamics from the EEG measurements. If we assume that the number of active EEG sources and their localizations do not vary considerably across subjects we can define a common approximation of the leadfield matrix for the whole population: $\hat{\boldsymbol{\Phi}} \approx [\boldsymbol{\phi}_1^*, ..., \boldsymbol{\phi}_K^*]$. This whole-population leadfield matrix can be used to estimate the intracranial localization of each EEG source with swLORETA. Additionally, the temporal dynamics of the centrotype EEG sources can be obtained for an arbitrary subject by means of spatial filtering: $s_i(t) = \mathbf{v}_i^* \mathbf{x}(t)$; $\forall i = 1, ..., K$, where $\mathbf{x}(t)$ is the column vector of EEG measurements. Once we have obtained the time courses of the EEG sources, we estimate the coefficients of the MVAR model that best fit their mutual dynamics and compute the DTF. Lilliefors' test was used to test the null hypothesis that the time courses of the residuals of the EEG sources were normally distributed.

Since the distribution of DTF estimates obtained from an MVAR model is not well established (Kamiński et al., 2001; Kús et al., 2004), we used bootstrap surrogates to determine whether the obtained DTF estimates were statistically significant (Kamiński et al., 2001). As was said before, the DTF measures the ratio between the outflow from the EEG source j to the EEG source

i, in respect to all the inflows to the destination EEG source. Therefore, if $\gamma_{ij}(f) > \gamma_{ik}(f)\ \forall k \neq \{i,j\}$ then we can infer that the i^{th} EEG source is *mainly driven by* the j^{th} EEG source. However, for these inferences to be valid we need to define statistical tests able to reject the null hypothesis that $\gamma_{ij}(f) \leq \gamma_{ik}(f)$ for some $k \neq \{i,j\}$. Approaching this task analytically is complex due to (i) the highly non-linear relationship between data samples and DTF estimates and (ii) the interdependence between DTF values that were obtained from the same subject at the same frequency. Thus, we employed a numerical approach consisting of the following steps:

(1) After the *MVAR-EfICA* analysis, we have estimates of all the unknowns in Eq. 1, that is the order and the parameters of the MVAR model that best fit the centrotype EEG sources: $\hat{\mathbf{B}}_s(\tau) \approx \mathbf{B}_s(\tau) \quad \forall \tau = 1, ..., p$ and $\hat{\mathbf{n}}(t) \approx \mathbf{n}(t) \quad \forall t = 1, ..., L$. By randomly shuffling time instants (columns) of the estimated multivariate residual process $\hat{\mathbf{n}}(t)$ we generated J surrogates of the EEG sources:

$$\mathbf{s}^{(m)}(t) = \sum_{\tau=1}^{p} \hat{\mathbf{B}}_s(\tau)\mathbf{s}^{(m)}(t-\tau) + \hat{\mathbf{n}}^{(m)}(t) \qquad \forall m = 1, ..., J \tag{10}$$

where $^{(m)}$ indexes the surrogates.

(2) We re-estimated the coefficients of the MVAR model for each surrogate: $\hat{\mathbf{B}}_s^{(m)}(\tau)\ ;\ \forall \tau = 1, ..., p\ ;\ \forall m = 1, ..., J$.

(3) Using $\hat{\mathbf{B}}_s^{(m)}$, we re-estimated the DTF in the alpha band for each surrogate: $\gamma_{ij}^{(m)}\ ;\ \forall m = 1, ..., J$.

(4) For each EEG source, we computed the differences between inflows of different origin:

$$\Delta\gamma_{ij,ik}^{(m)} = \gamma_{ij}^{(m)} - \gamma_{ik}^{(m)}\ ;\ \forall i\ ;\ \forall k \neq \{i,j\} \tag{11}$$

(5) We computed the maximum and minimum inflow differences:

$$\begin{aligned} \Delta\gamma_{ij,ik}^{MIN} &= \min_m \left(\Delta\gamma_{ij,ik}^{(m)}\right) \;\; ; \; \forall i \; ; \; \forall k \neq \{i,j\} \\ \Delta\gamma_{ij,ik}^{MAX} &= \max_m \left(\Delta\gamma_{ij,ik}^{(m)}\right) \;\; ; \; \forall i \; ; \; \forall k \neq \{i,j\} \end{aligned} \tag{12}$$

Only if both $\Delta\gamma_{ij,ik}^{MIN}$ and $\Delta\gamma_{ij,ik}^{MAX}$ were positive we were confident that, for the i^{th} EEG source, the inflow from the j^{th} source was larger than the inflow from the k^{th} source since this was true for every surrogate.

3 Results

3.1 Simulations

Fig. 4 shows that *MVAR-EfICA* was reliable even for small data lengths, outperforming the traditional *low-dimensional MVAR* even with as few as 6400 data samples ($\epsilon_{p-val} < 0.001$). The poor performance of *ThinICA-MVAR* was caused by the presence of time-delayed MVAR covariances between the source signals, in violation of the assumptions made by *ThinICA-MVAR*. This explanation is supported by the fact that including more time-lagged covariance matrices did not improve the asymptotic performance of *ThinICA-MVAR* but made it worse.

A major pitfall of the traditional *low-dimensional MVAR* approach is its sensitivity to volume conduction effects. In particular, the accuracy of *low-dimensional MVAR* was seriously compromised even for very shallow cortical dipoles, if they were tangential to the scalp surface. By contrary, *MVAR-EfICA* was considerably more robust to the presence of non-radially oriented

dipoles. Fig. 5 shows that the improvement of *MVAR-EfICA* with respect to the traditional *low-dimensional MVAR* approach was significant for most of the possible depths and orientations of the source dipoles.

Fig. 6 shows that the proposed *MVAR-EfICA* methodology was largely undisturbed by the presence of measurement noise and it performed significantly better than the traditional *low-dimensional MVAR* approach for low levels of biological noise.

MVAR-EfICA was quite robust to mild violations of the assumption of non-Gaussian residuals (Fig. 7). *ThinICA-MVAR* was much more affected by the distribution of the MVAR residuals and performed significantly better than the traditional approach only when they exhibited a highly super-Gaussian distribution ($\alpha < 0.001$; see Fig. 7).

3.2 Real EEG data

Fig. 8 depicts the results of the ICASSO analysis applied to alpha EEG recordings from 20 healthy elderly subjects. The R-index suggested an optimal partition of 12 clusters in the data. Clusters 10, 11 and 12 were the only significant ones and contained 1125, 1050 and 825 ICA-estimates, respectively (out of a total of 4425 estimates). Those three clusters exhibited a high repeatability within the same subject and across subjects. In particular, ICA-estimates within cluster 10 were obtained from 15 different subjects while 14 and 11 subjects contributed to clusters 11 and 12, respectively. All remaining clusters were much smaller and had high cross-similarities with one or more of the three major clusters. Clusters 10, 11 and 12 were selected for further analy-

sis since they conveyed the most replicable and stable features of EEG-alpha rhythm.

The normalized scalp topographies corresponding to the representative centrotypes of clusters 10, 11 and 12 are shown in Fig. 8 (right panel). The single electrical dipoles most likely to be generating each of those topographies were located in caudal regions of the thalamus (cluster 10, $x = 9$, $y = -25$, $z = 9$), in the precuneus (cluster 11, $x = 2$, $y = -60$, $z = 28$), and in the middle occipital gyrus, within the limits of the cuneus (cluster 12, $x = 11$, $y = -97$, $z = 13$). The corresponding localization probability maps are shown in Fig. 9. Lillieford's test rejected the Gaussianity hypothesis for the centrotype ICA-estimates of clusters 10, 11 and 12 ($p < 0.01$).

Fig. 10 summarizes the results regarding directed flows between the alpha generators in the pre-cuneus (P), the cuneus (C) and the thalamus (T). There was a clear bidirectional link between the generation of EEG-alpha in the thalamus and the precuneus. EEG-alpha oscillations originated in the thalamus were mainly driven by EEG-alpha generated in the cuneus in 12 subjects ($p < 0.01$). By contrary, the thalamic source was mainly driven by EEG-alpha generation in the pre-cuneus only in 1 subject ($p < 0.01$). Similarly, the thalamic source had a main effect on the generation of EEG-alpha in the cuneus in 12 of the subjects ($p < 0.01$) whereas the hypothesis that the inflow to the cuneus was larger from the precuneus than from the thalamus did not reach significance ($p > 0.01$) in any of the subjects. Moreover, the participation of the precuneus in the EEG alpha generation did not exert a major effect on either the thalamus or the cuneus, which rules out the possibility that the bidirectional flow between thalamus and cuneus might be due to indirect flows through the precuneus. Only in 1 subject the flow of EEG-alpha activity from

pre-cuneus to thalamus was significantly larger ($p < 0.01$) than the flow from the cuneus to the thalamus. Overall, the precuneus seemed to behave like a sink of EEG-alpha activity generated in the thalamus and/or the cuneus, whereas the major mechanism regulating EEG-alpha generation was a strong bidirectional causal feedback between thalamus and cuneus. The origin of the EEG-alpha activity inflow to the precuneus is uncertain because the strong bidirectional link between thalamus and cuneus does not allow us to discard the possibility that the flow from thalamus to precuneus (from cuneus to precuneus) is actually caused by an indirect flow through the cuneus (thalamus). This issue could be clarified by incorporating additional information-flow measures like the PDC, which is a topic that we are currently investigating.

4 Discussion

In this study we have presented a robust methodology for estimating directional flows of activity between EEG sources. The major features of the proposed *MVAR-EfICA* approach are that (i) it removes spurious flows between scalp EEG signals due to volume conduction effects and (ii) it does not make any a-priori assumptions about the intracranial localization of the underlying EEG generators.

An advantage of the proposed methodology is that each column of the estimated leadfield matrix $\hat{\mathbf{\Phi}}$ determines the localization of a single generator. It has been previously reported that, by localizing each independent source of activity separately, the localization error can be significantly reduced (Tang et al., 2002). In this context, we have to admit that, although the conductivity values of scalp, skull and brain used in this study can be considered as a de-

facto standard in head modeling, they might be far from realistic (Gonçalves et al., 2003), suggesting the necessity of measuring them in vivo for each subject in order to decrease errors associated with the EEG source locations. Moreover, for the estimated leadfield matrix to be a valid approximation of the true leadfield matrix the assumptions of the model need to be fulfilled, mainly that the residuals in Eq. 3 are mutually independent and that the number of EEG sensors is greater than the number of relevant EEG sources.

The poor performance of *ThinICA-MVAR* is mainly explained by the fact that high-order ICA contrasts are very disturbed by the presence of time-lagged correlations between the sources, which violate the ICA assumption that the sources lack any kind of temporal structure. *MVAR-EfICA* overcomes this problem by applying ICA on the residuals of the MVAR model, which are (ideally) free of any temporal structure and still contain the same instantaneous spatial dependencies of the original sources.

Only few source space studies have attempted to provide a global pattern of directional connectivity across a population of subjects (Astolfi et al., 2007a; Supp et al., 2007). In the present study, ICASSO effectively integrated the information obtained from several subjects and provided a concise and simple description of the whole population. This comes at the cost of requiring that the number of active EEG sources and their intracranial localizations are similar across subjects. In the sight of the results, this seemed to be the case of the alpha-EEG datasets used here. However, this assumption might not hold for more complex EEG patterns and/or more heterogeneous populations, which is a topic for further research.

The proposed *MVAR-EfICA* analysis methodology requires that there are not

instantaneous flows of activity between brain generators and that the instantaneous innovation processes driving each generator are non-Gaussian. Instantaneous synaptic flows between neuronal populations located more than few mm apart are unlikely due to axonal propagation delays (Freeman, 2000). Although there is no fundamental reason to believe that the innovation process is non-Gaussian, several previous studies have also found meaningful non-Gaussian sources of brain activity (e.g. Makeig et al., 2002; Huang et al., 2007; Astolfi et al., 2007a), suggesting the existence of non-Gaussian generators in the brain.

The most crucial step in the *MVAR-EfICA* procedure is the MVAR modeling step. If the MVAR model does not provide an accurate representation of the sources or if the model coefficients are not accurately estimated *MVAR-EfICA* will probably fail to reveal the underlying connectivity pattern. This is also true for all connectivity studies that use MVAR modeling. Yet, we have to recognize that our technique might be specially disturbed by inaccuracies of the MVAR model because of the sensitivity of ICA to the presence of correlations in the residuals of the MVAR model. We are currently investigating the possibility of improving the decorrelation of the model residuals by replacing the MVAR model with the more general multivariate autoregressive moving-average model.

Brain oscillations in the range of alpha activity are one of the fundamental electrophysiological phenomena of the human EEG. This brain activity can be easily identified by its topographic distribution (maximum amplitude over parieto-occipital regions), frequency range ($8 - 13$ Hz), and reactivity (it suffers a dramatic amplitude attenuation with the opening of the eyes (IFSECN, 1974). The study of alpha oscillations has generated a vast amount of literature

related with physiological, maturational, clinical, and cognitive aspects (Schurmann and Başar, 2001; Pfurtscheller et al., 1996; Niedermeyer and Lopes da Silva, 1993).

Highly coherent alpha oscillations with significant phase shifts have been observed in both the visual cortex and the lateral geniculate nucleus in non-human mammals (Chatila et al., 1993; Lopes da Silva et al., 1973, 1980), supporting the involvement role of thalamo-cortical circuits in the generation of waking-alpha rhythm. Neocortical neurons located in the layer V of the occipital cortex seem to be intrinsic alpha generators, as revealed by results from *in vitro* preparations (Silva et al., 1991) and *in vivo* recordings (Lopes da Silva and Storm van Leeuwen, 1977). They may receive thalamic inputs in order to maintain activation of cortical columns at an optimal level depending on the brain activation state. The number and exact location of alpha generators remain, however, unclear.

From the analysis of the EEG alpha rhythm recorded from 20 volunteers under resting conditions, we found that the bidirectional feedback between thalamus and cuneus was crucial in the EEG alpha generation. The precuneus seemed to play a secondary (or independent) role and was not the source of any causal inflow neither to the thalamus nor to the cuneus. This finding is consistent with (Schreckberger et al., 2004) reporting a positive correlation between EEG alpha power and metabolism of the lateral thalamus as well as occipital cortex (cuneus) and adjacent parts of the parietal cortex (precuneus) in humans. Our results also revealed that thalamocortical synaptic transmission remained alike from thalamus to cortex and vice-versa, which is in agreement with neural simulations showing that bidirectional coupling between distant brain areas engenders strong oscillatory activity (David and

Friston, 2003). These findings, together with results from human studies employing 3D equivalent dipole modeling (Isaichev et al., 2001; Başar et al., 1997; Schurmann et al., 2000), support the notion that complex interactions between local and non-local EEG sources, instead of a single or multiple isolated neural generators, are responsible for the genesis of the human alpha rhythm (Nunez et al., 2003).

In conclusion, we presented a novel methodology for measuring directed (causal) flows of activity between brain areas using the EEG. The proposed approach is based on well-established techniques such as MVAR modeling, ICA, clustering and swLORETA. Simulated experiments showed improved robustness and accuracy with respect to more traditional approaches. We further evaluated the validity of our method using EEG recordings of alpha waves from a set of 20 control subjects. The proposed technique estimated brain locations and causal flows of brain activity in agreement with the most recent findings about the generation mechanisms of the alpha rhythm in humans.

5 Acknowledgments

This research was supported by grants from the European Union (FP6-2005-NEST-Path 043309), by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program 2006-2011), the Spanish Ministry of Education and Science (SAF2005-00398), Regional Ministry of Innovation, Science and Enterprise, Junta de Andalucia (CTS-229), and the CIBERNED (Network for Biomedical Research in Neurodegenerative Diseases, Spanish Ministry of Health).

References

Akaike, H., 1971. Autoregressive model fitting for control. Ann. Inst. Stat. Math. 23, 163–180.

Astolfi, L., Bakardjian, H., Cincotti, F., Mattia, D., Grazia Marciani, M., De Vico Fallani, F., Colosimo, A., Salinari, S., Miwakeichi, F., Yamaguchi, Y., Martinez, P., Cichocki, A., Tocci, A., Babiloni, F., 2007a. Estimate of causality between independent cortical spatial patterns during movement volition in spinal cord injured patients. Brain Topograph. 19, 107–123.

Astolfi, L., Cincotti, F., Mattia, D., Grazia Marciani, M., Baccala, L. A., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., Christopher Edgar, J., Miller, G. A., He, B., Babiloni, F., 2007b. Comparison of different cortical connectivity estimators for high-resolution EEG recordings. Hum. Brain Mapp. 28 (2), 143–157.

Baccalá, L. A., Sameshima, K., 2001. Partial directed coherence: A new concept in neural structure determination. Biol. Cybern. 84, 463–474.

Bakardjian, H., Cichocki, A., Cincotti, F., Mattia, D., Babiloni, F., Grazia Marciani, M., De Vico Fallani, F., Miwakeichi, F., Yamaguchi, Y., Martinez, P., Salinari, S., Tocci, A., Astolfi, L., 2006. Estimate of causiality between cortical spatial patterns during voluntary movements in normal subjects. International Journal of Bioelectromagnetism 8 (1), II/1–II/18.

Başar, E., Schürmann, M., Başar-Eroglu, C., Karakaş, S., 1997. Alpha oscillations in brain functioning: an integrative theory. Int. J. Psychophysiol. 26 (1–3), 5–29.

Bernasconi, C., König, 1999. On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. Biol. Cybern. 81, 199–210.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S., 2004. Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality. P. Natl. Acad. Sci. USA 101 (26), 9849–9854.

Chatila, M., Milleret, C., Rougeul, A., Buser, P., 1993. Alpha rhythm in the cat thalamus. C. R. Acad. Sci. III 316, 51–58.

Cheung, Y.-m., Xu, L., 2003. Dual multivariate auto-regressive modeling in state space for temporal signal separation. IEEE T. Syst. Man Cyb. 33, 386–398.

David, O., Friston, K. J., 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20 (3), 1743–1755.

Eichler, M., 2006. On the evaluation of information flow in multivariate systems by the directed transfer function. Biol. Cybern. 94, 469–482.

Franaszczuk, P. J., Bergey, G. K., 1998. Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. Brain Topogr. 11, 13–21.

Freeman, W. J., 2000. Characteristics of the synchronization of brain activity imposed by finite conduction velocities of axons. Int. J. Bifurcation Chaos 10 (10), 2307–2322.

Freiwald, W. A., 1999. Testing nonlinearity and directedness of interactions between neural groups in the macaque inferotemporal cortex. J. Neurosci. Meth. 94, 105–119.

Gómez-Herrero, G., 2008. MVAR-ICA toolbox for MATLAB.
URL `http://www.cs.tut.fi/~gomezher/projects/eeg/`

Gonçalves, S., Munck, J. d., Verbunt, J., Heetharr, R., da Silva, F., 2003. *In vivo* measurement of brain and skull resistivities using an EIT-based method and the combined analysis of SEF/SEP data. IEEE T. Biomed. Eng. 50,

1124–1128.

Granger, C. W. J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., 2004. Modulation of long-range neuronal synchrony reflects temporal limitations of visual attention in humans. Proc. Natl. Acad. Sci. 101, 13050–13055.

He, P., Wilson, G., Russell, C., 2004. Removal of ocular artifacts from electroencephalogram by adaptive filtering. Med. Biol. Eng. Comp. 42, 407–412.

Himberg, J., Hyvärinen, A., 2005. ICASSO toolbox.
URL `http://www.cis.hut.fi/projects/ica/icasso/`

Himberg, J., Hyvärinen, A., Esposito, F., 2004. Validating the independent components of neuroimaging time-series via clustering and visualization. NeuroImage 22, 1214–1222.

Hoechstetter, K., Bornfleth, H., Weckesser, D., Ille, N., Berg, P., Scherg, M., 2004. BESA source coherence: A new method to study cortical oscillatory coupling. Brain Topogr. 16 (4), 233–238.

Huang, R.-S., Jung, T.-P., Delorme, A., Makeig, S., 2007. Tonic and phasic electroencephalographic dynamics during continuous compensatory tracking. NeuroImage. In press, doi: 10.1016/j.neuroimage.2007.10.036.

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for Independent Component Analysis. IEEE T. on Neural Networks 10 (3), 626–634.

Hyvärinen, A., 2001. Complexity pursuit: Separating interesting components from time series. Neural Comput. 13 (4), 883–898.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons.

International Federation of Societies for Electroencephalography and Clinical

Neurophysiology (IFSECN), 1974. A glossary of terms commonly used by clinical electroencephalographers. Electroencephalogr. Clin. Neurophysiol. 376, 538–548.

Isaichev, S. A., Derevyankin, V. T., Koptelov, Y. M., Sokolov, E. N., 2001. Rhythmic alpha-activity generators in the human EEG. Neurosci. Behav. Physiol. 31 (1), 49–53.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., Sejnowski, T. J., 2000. Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37, 163–178.

Kamiński, M., Blinowska, K. J., 1991. A new method of the description of the information flow in the brain structures. Biol. Cybern. 65, 203–210.

Kamiński, M., Ding, M., Truccolo, W., Bressler, S., 2001. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assesment of significance. Biol. Cybern. 85, 145–157.

Koldovský, Z., Tichavský, P., Oja, E., 2006. Efficient variant of algorithm fastica for independent component analysis attaining the Cramer-Rao lower bound. IEEE T. Neural Networks 17 (5), 1265–1277.

Kús, R., Kamiński, M., Blinowska, K., 2004. Determination of EEG activity propagation: pair-wise versus multichannel estimate. IEEE T. Biomed. Eng. 51 (9), 1501–1510.

Lehmann, D., Faber, P. L., Gianotti, L. R. R., Kochi, K., Pasqual-Marqui, R. D., 2006. Coherence and phase locking in the scalp EEG and between LORETA model sources, and microstates as putative mechanisms of brain temporo-spatial functional organization. J. Physiology-Paris 99, 29–36.

Lopes da Silva, F. H., Lierop, T. H. M. T., Schrijer, C. F., Storm van Leeuwen, W., 1973. Organization of thalamic and cortical alpha rhythms: spectra and coherences. Electroencephalogr. Clin. Neurophysiol. 35, 627–639.

Lopes da Silva, F. H., Storm van Leeuwen, W., 1977. The cortical source of the alpha rhythm. Neurosci. Lett. 6, 237–241.

Lopes da Silva, F. H., Vos, J. E., Mooibroek, J., van Rotterdam, A., 1980. Relative contributions of the intracortical and thalamo-cortical processes in the generation of alpha rhythms, revealed by partial coherence analysis. Electroencephalogr. Clin. Neurophysiol. 50, 449–456.

Makeig, S., Westerfield, M., Jung, T., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T., 2002. Dynamic brain sources of visual evoked responses. Science 295, 690–694.

Malmivuo, J., Plonsey, R., 1995. Bioelectromagnetism: Principles and applications of bioelectric and biomagnetic fields. Oxford University Press, New York, available online: `http://butler.cc.tut.fi/~malmivuo/bem/book/index.htm`.

Meinecke, F. C., Ziehe, A., Kurths, J., Müller, K.-R., 2005. Measuring phase synchronization of superimposed signals. Phys. Rev. Lett. 94, 084102.

Niedermeyer, E., Lopes da Silva, F. H. (Eds.), 1993. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 3rd Edition. Williams & Wilkins, Baltimore.

Nolte, G., Meinecke, C., Ziehe, A., Müller, K.-R., 2006. Identifying interactions in mixed and noisy complex systems. Phys. Rev. E 73, 051913.

Nunez, P. L., Wingeier, B. M., Silberstein, R. B., 2003. Spatial-temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale measurements, and global binding of local networks. Hum. Brain Mapp. 13 (3), 125–164.

Palmero Soler, E., Dolan, K., Hadamschek, V., Tass, P. A., 2007. swLORETA: a novel approach to robust source localization and synchronization tomography. Phys. Med. Biol. 52, 1783–1800.

Pfurtscheller, G., Stancak, A., Neuper, C., 1996. Event-related synchronization (ERS) in the alpha band-an electrophysiological correlate of cortical idling: a review. Int. J. Psychophysiol. 24, 39–46.

Richards, J. E., 2004. Recovering dipole sources from scalp-recorded event-related-potentials using component analysis: principal component analysis and independent component analysis. Int. J. Psychophysiol. 54, 201–220.

Särelä, J., Vigário, R., 2003. Overlearning in marginal distribution-based ICA: Analysis and solutions. J. Mach. Learn. Res. 4, 1447–1469.

Schlögl, A., 2006. A comparison of multivariate autoregressive estimators. Signal Process. 86, 2426–2429.

Schneider, T., Neumaier, A., 2001. Algorithm 808: ARFIT - a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. ACM T. Math. Soft. 27 (1), 58–65.

Schreckberger, M., Lange-Asschenfeld, C., Lochmann, M., Mann, K., Siessmeier, T., Buchholz, H.-G., Bartenstein, P., Gründer, 2004. The thalamus as the generator and modulator of EEG alpha rhythm: a combined PET/EEG study with lorazepam challenge in humans. NeuroImage 22, 637–644, a stable positive correlation between EEG alpha power and metabolic activity of the lateral thalamic nuclei in humans was found even under pharmacological alpha suppression.

Schurmann, M., Başar, E., 2001. Functional aspects of alpha oscillations in the EEG. Int. J. Psychophysiol. 39, 151–158.

Schurmann, M., Demiralp, T., Başar, E., Başar-Eroglu, C., 2000. Electroencephalogram alpha (8-15 Hz) responses to visual stimuli in cat cortex, thalamus, and hippocampus: a distributed alpha network? Neurosci. Lett. 292, 175–178.

Schwarz, T., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.

Silva, L. R., Y., A., W., C. B., 1991. Intrinsic oscillations of neocortex generated by layer 5 pyramidal neurons. Science 251, 432–435.

Supp, G. G., Schlögl, A., Trujillo-Barreto, N., Müller, M. M., Gruber, T., 2007. Directed cortical information flow during human object recognition: Analyzing induced eeg gamma-band responses in brain's source space. PLoS One 2 (8), e684, doi:10.1371/journal.pone.0000684.

Tang, A. C., Pearlmutter, B. A., Malaszenko, N. A., Phung, D., Reeb, B. C., 2002. Independent components of magnetoencephalography: Localization. Neural Computation 14 (8), 1827–1858.

Uhlhass, P. J., Singer, W., 2006. Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. Neuron. 52, 155–168.

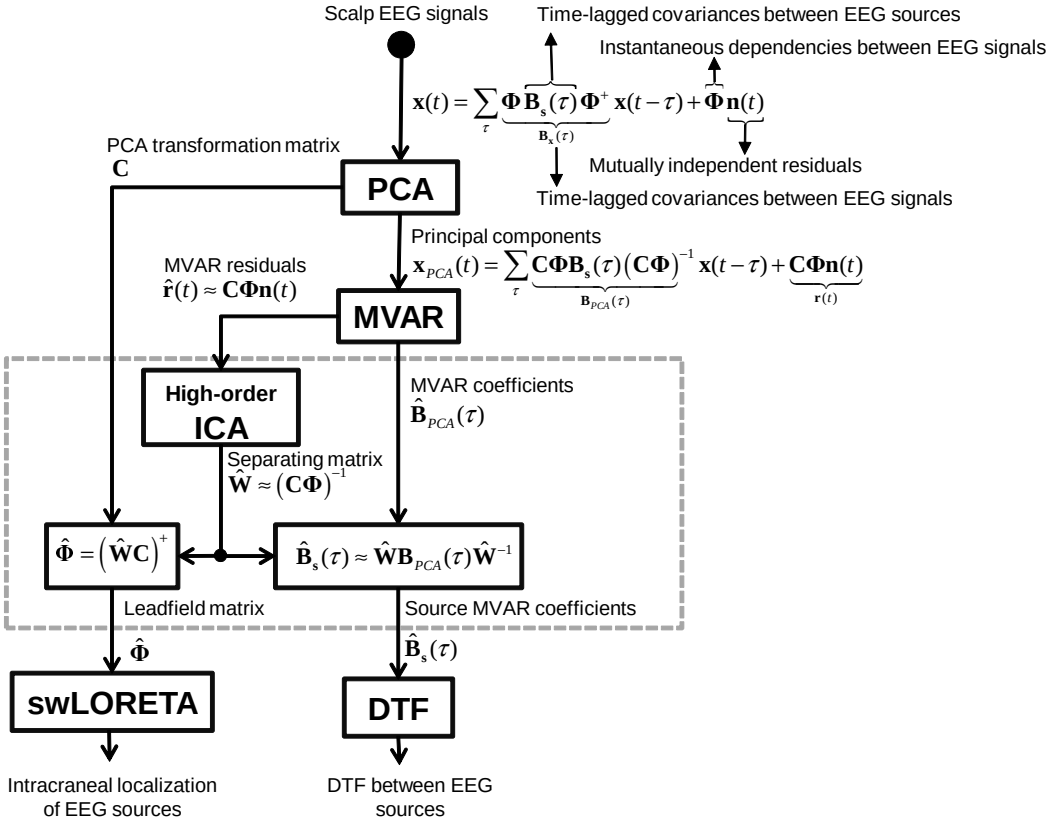

Fig. 1. Block diagram of the proposed *MVAR-EfICA* methodology. The PCA and ICA blocks model instantaneous cross-dependencies between scalp EEG signals, which are exclusively caused by volume conduction (matrix $\mathbf{\Phi}$ in the diagram). The PCA block takes care of second order correlations whereas the ICA block models dependencies of greater orders. By contrary, the MVAR block models time-lagged covariances between scalp EEG signals (matrices $\mathbf{B}_{\mathbf{x}}(\tau)$ in the diagram), which are partially caused by time-lagged flows of activity between EEG sources (matrices $\mathbf{B}_{\mathbf{s}}(\tau)$ in the diagram). By combining the PCA, MVAR and ICA models it is possible to obtain an estimate of the leadfield matrix and of the coefficients of the MVAR model driving the EEG sources. From the leadfield matrix, swLORETA obtains the probable intracraneal localization of the EEG sources. From the coefficients of the source MVAR model it is straightforward to compute the DTF. The dashed rectangle identifies the blocks that are modified when ICASSO is introduced.

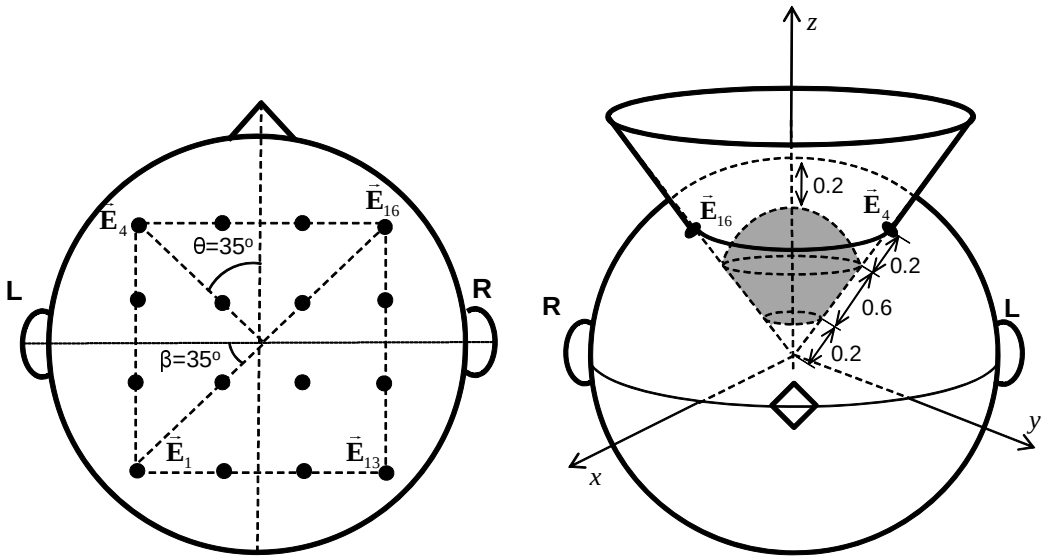

Fig. 2. (Left) Location of the electrodes in the simulated head with normalized radius $r = 1$. The black-filled circles illustrate the positions of the electrodes. The position of the corner electrodes $\vec{\mathbf{E}}_1, \vec{\mathbf{E}}_4, \vec{\mathbf{E}}_{13}, \vec{\mathbf{E}}_{16}$ were fixed. The other electrodes were placed uniformly in θ and β in such a way that the angular differences between two neighboring electrodes were $\Delta\theta = \Delta\beta = 70^o/3$. (Right) Possible locations of the noisy dipoles. The valid locations (in grey) were within the volume generated by the intersection of the head sphere with a cone whose vertex is located in the center of the head and whose lateral surface contains the four corner electrodes $\vec{\mathbf{E}}_1, \vec{\mathbf{E}}_4, \vec{\mathbf{E}}_{13}, \vec{\mathbf{E}}_{16}$. Additionally, the noisy dipoles were required to be located at a minimum distance of 0.2 from the scalp and at the same minimum distance from the center of the head.

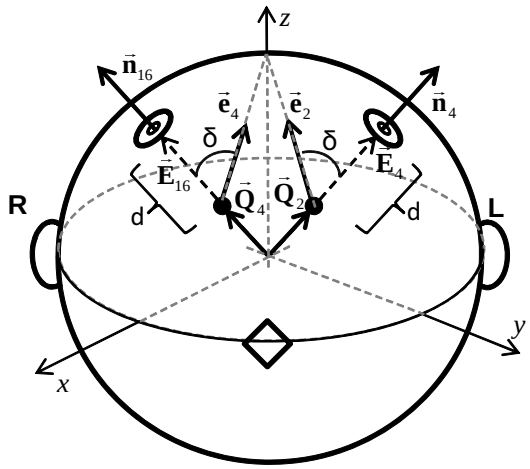

Fig. 3. Anterior frontal view of the simulated head with the positions of the two frontal source dipoles ($\vec{\mathbf{Q}}_2$, $\vec{\mathbf{Q}}_4$) marked with filled black dots and with the position of the two frontal corner electrodes ($\vec{\mathbf{E}}_4$, $\vec{\mathbf{E}}_{16}$) marked with empty circles. The source dipoles were located in the radii that connected the center of the head with the corner electrodes. The distance d between the dipoles and the scalp was a free parameter of the simulations and its value was normalized according to the head radius ($r = 1$). The dipoles orientation vectors $\vec{\mathbf{e}}_2$, $\vec{\mathbf{e}}_4$ formed an angle δ with their corresponding location vectors $\vec{\mathbf{Q}}_2$, $\vec{\mathbf{Q}}_4$. Their rotation angle around their radial component was random in each simulation surrogate. If $\delta = 0°$ the dipoles were radial whereas if $\delta = 90°$ they were tangential to the scalp surface. The positions and orientations of dipoles 1 and 3 were exactly symmetrical to those of dipoles 2 and 4 in respect of the midline coronal plane.

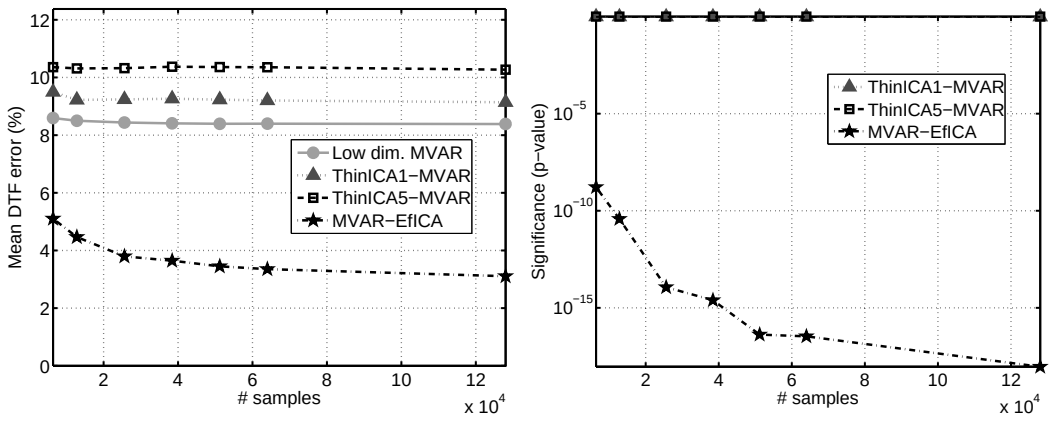

Fig. 4. (Left) Mean DTF estimation error for the *low-dimensional MVAR* approach, the two variants of *ThinICA-MVAR* and *MVAR-EfICA* versus the number of simulated data samples. (Right) Probability (p-value) that the *low-dimensional MVAR* method performed better (achieved a lower mean DTF error) than *MVAR-EfICA* and/or *ThinICA-MVAR* in a random simulation surrogate.

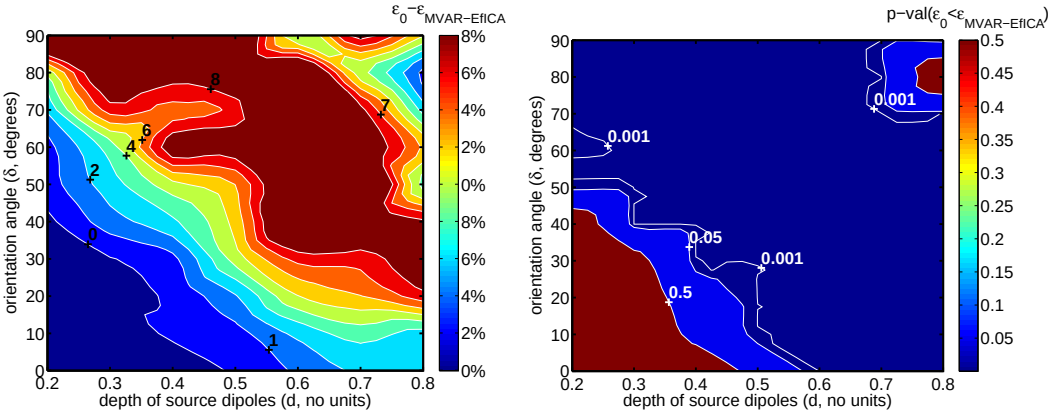

Fig. 5. Volume conduction effects.(Left) The color scale shows the value of the difference $\bar{\epsilon}_0 - \bar{\epsilon}_{MVAR-EfICA}$, where $\bar{\epsilon}_0$ denotes the mean DTF error obtained with the traditional *low-dimensional MVAR* approach and $\bar{\epsilon}_{MVAR-EfICA}$ denotes the error of the proposed *MVAR-EfICA* approach. Indeed, the more positive this value the greater the average improvement obtained with *MVAR-EfICA*.(Right) Probability (p-value) that *low-dimensional MVAR* is more accurate than *MVAR-EfICA*, i.e. the probability of the event $\bar{\epsilon}_0 < \bar{\epsilon}_{MVAR-EfICA}$. The smaller this value is the greater the significance of the improvement obtained with *MVAR-EfICA*

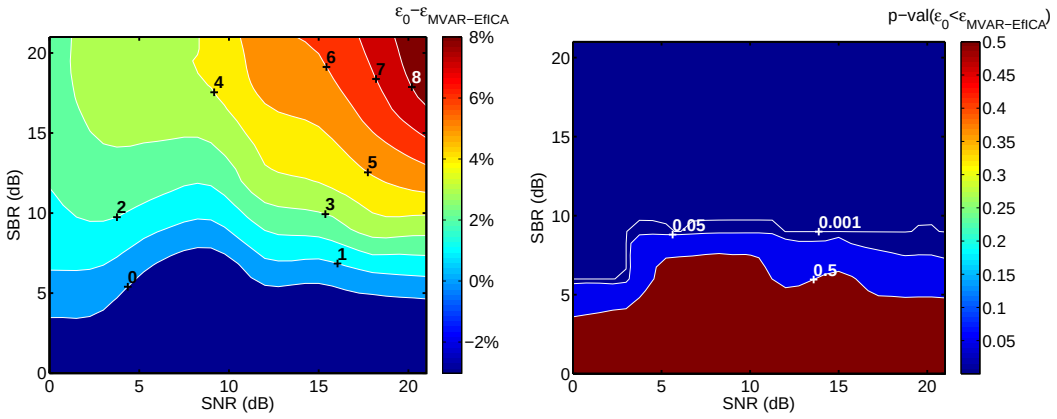

Fig. 6. Effects of Gaussian thermal noise at the EEG sensors (SNR) and biological noise ($SBNR$).The interpretation of this figure is analogous to that of Fig. 5.

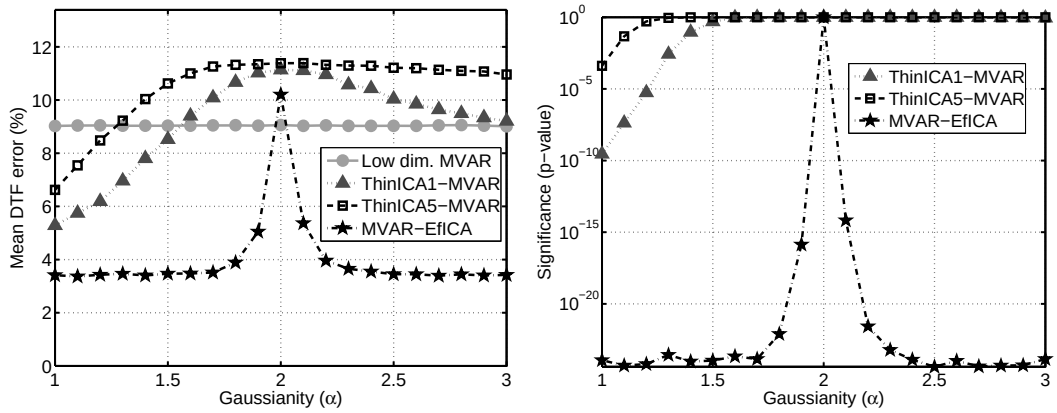

Fig. 7. (Left) Effects of the Gaussianity of the MVAR model residuals on the average DTF error. (Right) Probability (p-value) that the traditional *low-dimensional MVAR* approach ourperforms the alternative techniques.

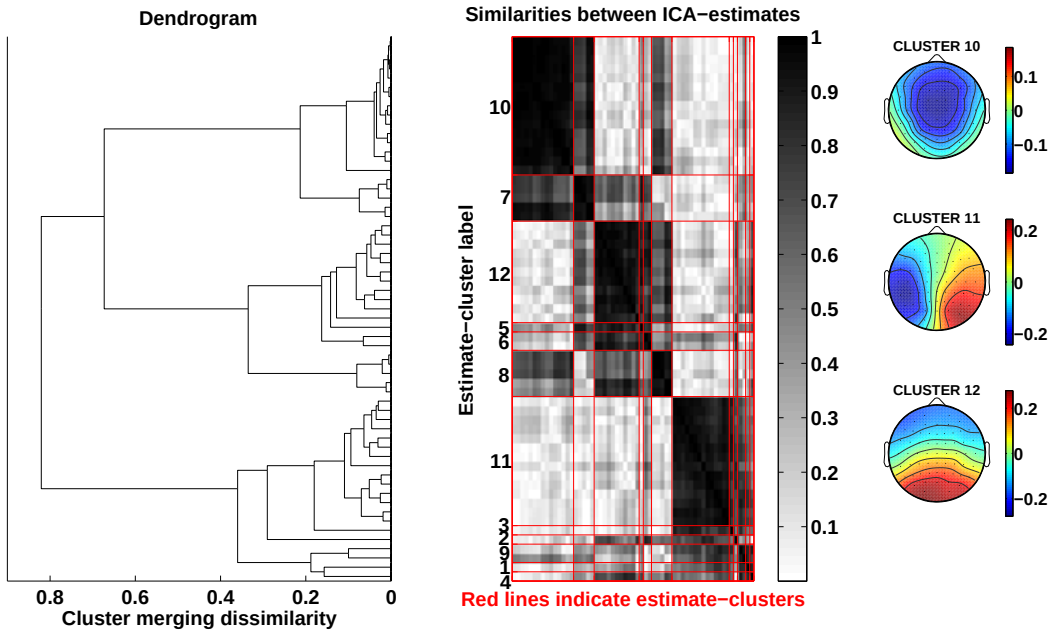

Fig. 8. (Left panel) Dendrogram illustrating the arrangement in 12 clusters (as suggested by the R-index) of the ICA-estimates obtained with ICASSO. The horizontal axis represents the dissimilarity values at which clusters are merged at each possible partition level. The vertical axis indexes ICA-estimates. (Middle panel) Similarity matrix. The color scale indicates the cross-correlation coefficient between the scalp topographies of individual ICA-estimates. Clusters of ICA-estimates are indicated with red lines and their correponding labels are depicted in the left vertical axis. (Right panel) Normalized distributions of scalp potentials corresponding to the centrotypes of clusters 10, 11 and 12, which are, by far, the largest and most compact.

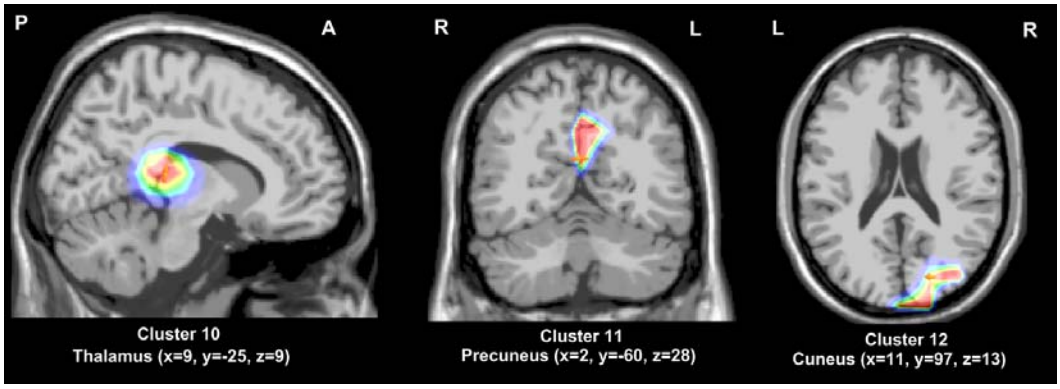

Fig. 9. Localization obtained with swLORETA of electric dipole sources for the scalp distribution of alpha oscillations associated with clusters 10, 11 and 12 (thalamus, precuneus and cuneus, respectively).

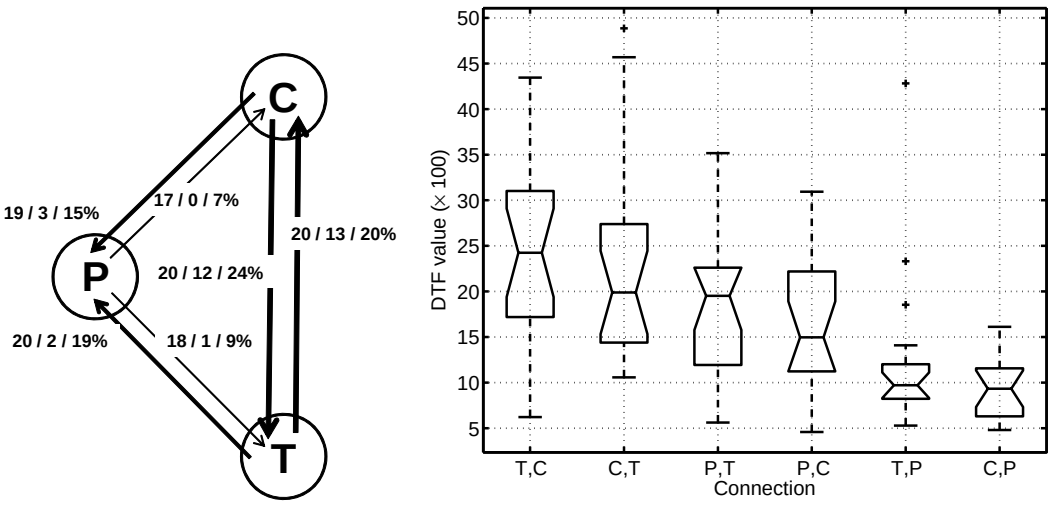

Fig. 10. (Left) Causal flows between EEG-alpha generators in the precuneus (P), cuneus (C) and thalamus (T). Each flow is characterized by three numbers. The first number corresponds to the number of subjects for which the flow was significant ($p = 0.01$). The second number is the number of subjects for which the flow was identified as the most important inflow to the destination EEG source ($p < 0.01$). The third number is the median DTF value across subjects. Based on the results, the directional flows were ranked according to their qualitative significance into three groups (identified by different line widths in the diagram): (i) the bidirectional flow between thalamus and cuneus, (ii) the inflows to the precuneus originated in cuneus and thalamus, (iii) the outflows from precuneus to cuneus and thalamus. (Right) Spread across subjects of DTF values corresponding to directional flows of activity between EEG-alpha generators in the precuneus (P), the cuneus (C) and the thalamus (T). The horizontal axis depicts the six possible directional flows, with X,Y meaning an inflow to generator X originated in generator Y. The vertical axis represents corresponding DTF values (in percentages). The lines within the boxes indicate the lower quartile, median and upper quartile values. The lines extending from each end of the boxes show the extent of the spread of DTF values across subjects. The notches in each box represent a robust estimate of the uncertainty about the medians of each box.