Unifying blind separation and clustering for resting-state EEG/MEG functional connectivity analysis

Jun-ichiro Hirayama[1]
Takeshi Ogawa[1]
Aapo Hyvärinen[2, 1]
[1]Cognitive Mechanisms Laboratories, Advanced Telecommunications Research Institute International (ATR), 2-2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan.
[2]Department of Computer Science and HIIT, University of Helsinki, Gustaf Hällströmin katu 2b, 00560 Helsinki, Finland.

Keywords: Unsupervised learning, Blind source separation, Functional connectivity, Electroencephalography, Magnetoencephalography

Abstract

Unsupervised analysis of the dynamics (non-stationarity) of functional brain connectivity during rest has recently received a lot of attention in both the neuroimaging and neuroengineering communities. Most studies have used functional magnetic resonance imaging (fMRI), but electroencephalography (EEG) and magnetoencephalography (MEG) also hold great promise for analyzing non-stationary functional connectivity with high temporal resolution. However, previous EEG/MEG analyses divided the problem into two consecutive stages: first, the separation of neural sources, and second, the connectivity analysis of the separated sources. Such non-optimal division into two stages may bias the result because of the different prior assumptions made about the data in the two stages. Here, we propose a unified method for separating EEG/MEG sources and learning their functional connectivity (coactivation) patterns. We combine blind source separation (BSS) with unsupervised clustering of the activity levels of the sources in a single probabilistic model. A BSS is performed on the Hilbert transforms of band-limited EEG/MEG signals, and coactivation patterns are learned by a mixture model of source envelopes. Simulation studies show that the unified approach often outperforms conventional two-stage methods, further indicating the benefit of using Hilbert transforms to deal with oscillatory sources. Experiments on resting-state EEG data, acquired in conjunction with a cued motor imagery/non-imagery task, also show that the states (clusters) obtained by the proposed method often correlate better with physiologically meaningful quantities than those obtained by a two-stage method.

1 Introduction

Unsupervised machine learning techniques play a fundamental role in the analysis of spontaneous (resting-state) neuroimaging signals by exploring the intrinsic statistical structures of such signals without relying on extrinsic covariates about tasks or stimulation protocols. The structures or features obtained can then be examined based on neurophysiological knowledge often using the features in a group comparison, or possibly by finding similar structures in other signals already associated with tasks or stimuli.

In recent years, there has been growing interest in exploring the patterns and dynamics of resting-state functional brain connectivity (Friston, 1994) based on unsupervised signal analyses. To find patterns in non-stationary functional connectivity, most studies have relied on such standard techniques as independent component analysis (ICA) (Brookes et al., 2011; Smith et al., 2012), principal component analysis (PCA) (Leonardi et al., 2013), and K-means clustering (Liu et al., 2013; Allen et al., 2014). New unsupervised analysis methods have also been actively developed (e.g., Haufe et al., 2010; Zhang and Hyvärinen, 2010; Hyvärinen et al., 2010b; Hirayama and Hyvärinen, 2012; Ramkumar et al., 2012, 2014; Dähne et al., 2014) to incorporate the specific nature of neuroimaging signals.

Such methods for analyzing functional brain connectivity, primarily based on signals' own statistics, are potentially very useful not only for neuroscientific investigations but also for neuroengineering applications. For example, a key challenge in an emerging new direction in brain-computer interface (BCI) research is to covertly acquire user's unobserved states during everyday life behaviors (Zander and Kothe, 2011; Lance et al., 2012). Since brain activity cannot be very well controlled in everyday life situations, no reliable class labels are available for discriminating the unobservable states of users. Such difficult data could be tackled if unsupervised analysis discovered connectivity patterns reflecting the user's cognitive states. Unsupervised connectivity analysis might also shed light on the neurophysiological basis of the BCI paradigms commonly used so far, such as those based on motor imagery, i.e., imagining body movements (Grosse-Wentrup, 2009, 2011).

Motivated by such potential applications, in this paper we focus on developing an unsupervised analysis method for finding connectivity-related signal features from electroencephalography (EEG); our method may also be readily applied to magnetoencephalography (MEG) because of its fundamental similarity. EEG/MEG's high temporal resolution is particularly useful for analyzing non-stationary connectivity, as compared to functional magnetic resonance imaging (fMRI), which is used in most connectivity studies.

The analysis of functional connectivity patterns in EEG/MEG, however, is not straightforward because the neural sources are mixed by volume conduction (and/or field spread) into sensor signals. Two-stage analysis has been conventionally performed by first separating the neural sources from the given sensor signals and analyzing the connectivity patterns based on those separated sources. In neuroimaging literature, electromagnetic inverse problems are often solved to separate (estimate) the activity of dipolar sources on the cortical grid, with an additional effort of physical forward modeling. On the other hand, in exploratory signal analysis related to BCI, blind source separation (BSS) methods (including ICA) are especially useful, since they greatly simplify the interpretation

of the results by decomposing the data into components, and the inverse problem can be solved to localize the obtained components afterwards (Hironaga and Ioannides, 2007; Doesburg and Ward, 2009). Note that these components are actually called "sources" in BSS literature, and a component can be an integration of multiple correlated dipolar sources.

The problem is that conventional two-stage analysis, i.e., first source separation and then connectivity analysis, is "neither a principled nor an optimal solution to the overall problem" (Makeig et al., 2012). Source separation methods rely on specific prior assumptions about the sources, which are not necessarily consistent with what the connectivity analysis assumes about them; such inconsistency between prior assumptions might bias the results. A more desirable unified treatment would be obtained by extending the conventional generative (forward) model used for source separation by including prior assumptions about the sources that are consistent with the connectivity analysis. Typically, connectivity analysis can be formulated as learning of a specific parametric model of the sources, and both layers of the model may be learned simultaneously, unified by the principle of statistical parameter estimation.

Here, we present a unified method for analyzing functional connectivity patterns in EEG/MEG sources with a jointly solved BSS, based on a novel two-layer extension of the conventional BSS/ICA generative model. In line with previous resting-state MEG studies (de Pasquale et al., 2010; Brookes et al., 2011; Ramkumar et al., 2014), we are particularly interested in finding coherent (frequently occurring) patterns of activity levels (coactivations) of oscillatory sources in a frequency band of interest. The connectivities are based on envelopes and ignore phase information, but our model is rather different from power-to-power coherences.

To properly model the source envelopes, our model uses a complex-valued formulation of BSS based on the Hilbert transform, which is a key departure from related extensions of ICA based on modeling real-valued data (Hyvärinen and Hoyer, 2000; Hyvärinen et al., 2001a; Valpola et al., 2004; Karklin and Lewicki, 2005; Kawanabe and Müller, 2005; Osindero et al., 2006; Köster and Hyvärinen, 2010; Haufe et al., 2010; Hirayama and Hyvärinen, 2012). Another important novelty here is using a finite mixture model of sources, in which they are assumed to exhibit different coactivation patterns corresponding to a finite number of unobserved "brain states." This corresponds to performing unsupervised clustering on the coactivations of the sources, inspired by the use of K-means clustering in previous resting-state fMRI studies (Liu et al., 2013; Allen et al., 2014). Due to the simplicity of the mixture model, our two-layer model is tractable, unlike previous two-layer models that are often difficult to learn. Our entire model can be readily estimated (optimized) by the maximum likelihood method without resorting to any approximations.

The rest of this paper is organized as follows. First, we present our proposed method based on a novel two-layer extension of the generative BSS/ICA model called the latent coactivity mixture model (LCMM) (Section 2). Then we provide simulation studies (Section 3) and real EEG data analysis (Section 4) to validate the unified approach. Finally, we discuss the results and open issues (Section 5). Preliminary results were presented at the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'14) (Hirayama et al., 2014).

2 Latent coactivity mixture model

2.1 Background: blind source separation

Before introducing our new model, we start by discussing the generative (forward) model conventionally used for the blind separation of EEG/MEG sources. Let $\boldsymbol{x}(t) = (x_1(t), x_2(t), \ldots, x_d(t))^\mathsf{T} \in \mathbb{R}^d$ be a multivariate EEG/MEG signal, sampled at discrete time points indexed by $t = 1, 2, \ldots, N$, which is assumed to have already been (band-pass) filtered so that each $x_j(t)$ is limited to a certain frequency band of interest. Sensor signal vector $\boldsymbol{x}(t)$ is then assumed to follow a linear generative model given by

$$\boldsymbol{x}(t) = \mathbf{A}\boldsymbol{s}(t), \tag{1}$$

where $\boldsymbol{s}(t) = (s_1(t), s_2(t), \ldots, s_d(t))^\mathsf{T} \in \mathbb{R}^d$ denotes the vector of the source signals and $\mathbf{A} \in \mathbb{R}^{d \times d}$ is called the mixing matrix, which is assumed to be non-singular so that demixing matrix $\mathbf{W} := \mathbf{A}^{-1}$ exists. Both $\boldsymbol{s}(t)$ and $\mathbf{A}$ are unknown and estimated from data in the BSS setting.

Note that in Eq. (1), the number of sources is assumed to equal (effective) dimensionality d of the sensor signal, as a fundamental setting in a standard ICA; this greatly simplifies the mathematical treatment. In practice, d can be selected to be smaller than the original number of sensor channels, typically by discarding the ineffective dimensions (with too small variances) using PCA. It should also be noted that each source (or component) $s_j(t)$ does not necessarily correspond to any single electrical dipole; instead, each $s_j(t)$ may describe the total effect of multiple correlated dipolar activities.

To solve the BSS problem, we need to make further assumptions about the statistical properties of the sources based on prior knowledge. Independent component analysis (ICA) typically assumes that d sources are non-Gaussian and mutually independent (Hyvärinen et al., 2001b), which theoretically guarantees the identifiability of both $\mathbf{A}$ and $\boldsymbol{s}(t)$'s, up to the scaling and permutation of the sources. However, the independence assumption might lead to a solution that is weakly functionally connected (i.e., weakly statistically dependent), even when the true sources are strongly dependent, which is not consistent with the goal of connectivity analysis. This motivated us to develop an appropriate model of functionally connected (dependent) sources.

2.2 Definition of latent coactivity mixture model

Our main interest here is modeling the envelopes (i.e., amplitudes) of narrow-band source signals (Onton and Makeig, 2009; Zhang and Hyvärinen, 2010; Brookes et al., 2011) for which the "real-valued" BSS model of Eq. (1) is not convenient. Envelopes can be modeled more easily with complex analytic signals $\widetilde{s}_j(t)$ (Schreier and Scharf, 2010), defined as

$$\widetilde{s}_j(t) = s_j(t) + i\mathcal{H}[s_j](t), \tag{2}$$

where $\mathcal{H}[f]$ denotes the Hilbert transform of signal $f(t)$ and i is an imaginary unit. The envelope of $s_j(t)$ is given by the modulus of $\widetilde{s}_j(t)$. This simple algebraic dependence of the envelope on $\widetilde{s}$ greatly simplifies the developments below.

We thus formulate a "complex-valued" BSS with a similar transformation of sensor signal vector $\boldsymbol{x}(t)$:

$$\widetilde{\boldsymbol{x}}(t) = \mathbf{A}\widetilde{\boldsymbol{s}}(t), \tag{3}$$

where complex sensor signal $\widetilde{\boldsymbol{x}}(t) = (\widetilde{x}_1(t), \widetilde{x}_2(t), \ldots, \widetilde{x}_d(t))^\mathsf{T} \in \mathbb{C}^d$ can be directly computed from the original one by $\widetilde{x}_j(t) := x_j(t) + i\mathcal{H}[x_j](t)$. Here, sources $\widetilde{s}_j(t)$ are further assumed to be centered (i.e., $\mathrm{E}[\widetilde{s}_j(t)] = 0$) without loss of generality by always subtracting the (sample) mean from $\widetilde{\boldsymbol{x}}$. Both mixing matrix $\mathbf{A}$ and complex-valued source signal $\widetilde{\boldsymbol{s}}(t) = (\widetilde{s}_1(t), \widetilde{s}_2(t), \ldots, \widetilde{s}_d(t))^\mathsf{T}$ are again unobserved and estimated from the data.

Note that mixing matrix $\mathbf{A}$ in Eq. (3) is identical to the original one in Eq. (1) because of the Hilbert transform's linearity. This ensures that the $\mathbf{A}$ columns can be interpreted directly as defining spatial topographies in the original sensor space. For simplicity, we constrain $\mathbf{A}$ to be real-valued, while a complex-valued $\mathbf{A}$ could also be straightforwardly used, which might be useful to deal with sources synchronized in different phases (Hyvärinen et al., 2010a).

We next define a coactivation (connectivity) structure between sources $\widetilde{s}_j(t)$. The fundamental assumption here is that a system generating the data can be in a finite number of different states, corresponding to different patterns of source amplitudes. Given the state at time point t, sources $\widetilde{s}_j(t)$ are generated based on a multivariate Student-t distribution (specified below) which implements the average source amplitudes specific to that state and generates random phases.

Thus, our model of connectivities is not based on explicitly measuring some form of correlations between the sources or their envelopes, as is typically done in electrophysiology. Instead, we characterize the interactions of the sources by dividing their joint activity into a number of typical patterns of envelopes, which intuitively express the idea that certain sources tend to be coactivated. Such coactivation does imply correlations of envelopes, i.e., power-to-power coherence, but provides a more detailed analysis of the coactivation than merely computing correlations.

The proposed latent coactivity mixture model (LCMM) is thus summarized as a two-layer generative model of complex sensor signal vector $\widetilde{\boldsymbol{x}}(t)$ as follows:

1. At each time point t, the system generating the data takes one of a finite number of different states (clusters) indexed by $k = 1, 2, \ldots, K$, according to multinomial probability distribution with cluster probabilities $\eta_1, \eta_2, \ldots, \eta_K$, where $\eta_k \geq 0$ and $\sum_{k=1}^{K} \eta_k = 1$.

2. Given that the system belongs to the k-th state at time t, source vector $\widetilde{\boldsymbol{s}}(t)$ is specifically generated by a complex multivariate Student-t distribution (Schreier and Scharf, 2010) with circular (see below) and mutually uncorrelated sources; they have state-conditional variances or expected powers (squared amplitudes), given by

$$\mathrm{E}[|\widetilde{s}_j(t)|^2]_k = \frac{\nu}{\nu - 2} b_{jk}, \quad j = 1, 2, \ldots, d, \tag{4}$$

where $\mathrm{E}[\cdot]_k$ denotes the conditional expectation given the k-th state, ν is an integer called the degrees of freedom, and nonnegative vector $\boldsymbol{b}_k = (b_{1k}, b_{2k}, \ldots, b_{dk})^\mathsf{T}$,

called the coactivation pattern, specifies the expected levels of the source envelopes.[1]

3. Complex sensor signal $\widetilde{\boldsymbol{x}}(t)$ is given as a linear instantaneous mixture of $\widetilde{\boldsymbol{s}}(t)$ with unknown mixing matrix $\mathbf{A}$, as in Eq. (3), common to each state k.

The N Hilbert-transformed sensor signal vectors, $\widetilde{\boldsymbol{x}}(1), \widetilde{\boldsymbol{x}}(2), \ldots, \widetilde{\boldsymbol{x}}(N)$, are simply assumed to be independently and identically distributed (i.i.d.), as is commonly done in many ICA methods. Such a simplification is mainly done for purposes of mathematical and computational tractability, but it could be relaxed by further modeling the autocorrelation structures of the sources (in future work).

The circularity of the sources means that the phase of each source is distributed uniformly and independently of its amplitude. Our method focuses on modeling (analyzing) an amplitude-to-amplitude type of connectivity (coactivation), ignoring phase-to-phase or phase-to-amplitude types of connectivity. In the circular case, the probability density function (pdf) of complex multivariate Student-t distribution (Schreier and Scharf, 2010) with scatter matrix $\boldsymbol{\Sigma}$ and $\nu\,(>0)$ degrees of freedom is given by

$$\widetilde{\mathcal{T}}(\widetilde{\boldsymbol{s}};\boldsymbol{\Sigma},\nu) = \frac{2^d\Gamma(d+\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^d|\boldsymbol{\Sigma}|}\left(1+\frac{2}{\nu}\widetilde{\boldsymbol{s}}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{s}}\right)^{-d-\frac{\nu}{2}}, \tag{5}$$

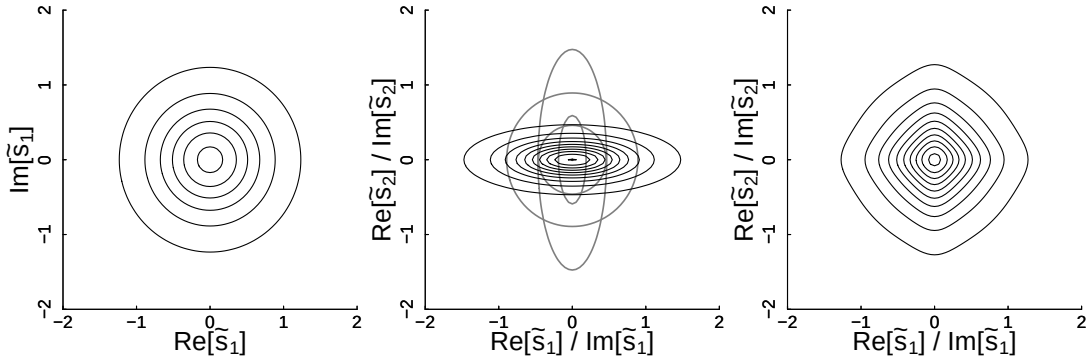
where $\cdot^{\mathsf{H}}$ denotes the Hermitian transpose and $\Gamma(\cdot)$ denotes the Gamma function. Source vector $\widetilde{\boldsymbol{s}}$ in LCMM has a state-conditional pdf given by $\widetilde{\mathcal{T}}(\widetilde{\boldsymbol{s}};\mathrm{diag}(\boldsymbol{b}_k),\nu)$. Figs. 1 (a) and (b) illustrate the conditional pdf for a bivariate case.

The particular choice of the Student-t model is mainly motivated by its robustness to outliers in the estimation of (co)variances (i.e., coactivation patterns), as has been thoroughly studied in the literature (see e.g., Ollila and Koivunen, 2003; Mahot et al., 2013).[2] Scatter matrix $\boldsymbol{\Sigma}$ is proportional to the covariance matrix if it exists, i.e., $\mathrm{E}[\widetilde{\boldsymbol{s}}\widetilde{\boldsymbol{s}}^{\mathsf{H}}] = \{\nu/(\nu-2)\}\boldsymbol{\Sigma}$ for $\nu > 2$, and maximum likelihood estimate (MLE) of $\boldsymbol{\Sigma}$ is often used as a robust alternative of the sample covariance matrix even with $\nu \leq 2$. Diagonal scatter matrix $\mathrm{diag}(\boldsymbol{b}_k)$ in LCMM implies Eq. (4), and the estimate of $\boldsymbol{b}_k$ serves as a robust estimator of the state-conditional variances; note that even if the variance does not exist when $\nu \leq 2$, $\boldsymbol{b}_k$ can have a finite MLE, giving a more general scale parameter estimate. Typically, since EEG/MEG signals contain a large amount of noises or artifacts from outside of the brain, robustness is a desirable property in practice. On the other hand, our model includes no explicit noise term in the generative model of Eq. (1), which is mainly for computational simplicity as in standard ICA methods.

Graphical representations of the dependency structure in LCMM are given in Fig. 2. The sources in LCMM are not independent of each other (Fig. 2(a); see also Fig. 1

[1] To be precise, the exact relation between $\boldsymbol{b}_k$ and the source variances in Eq. (4) is no longer valid if $\nu \leq 2$, since the variance is infinite or undefined. However, even if $\nu \leq 2$, the estimated $\boldsymbol{b}_k$ can still be interpreted as modeling the variance levels, while small ν implies canceling the effects from the outliers from the viewpoint of robust estimation (as explained in the text).

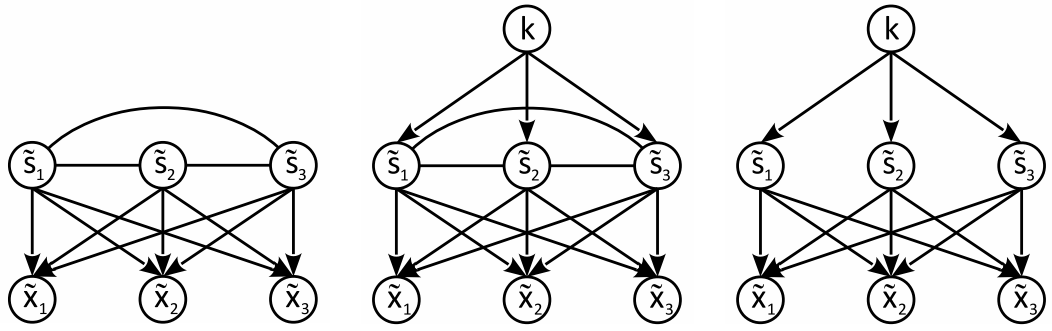
[2] Note that other heavy-tailed distributions in the complex elliptical family (Schreier and Scharf, 2010) commonly have the robustness property and thus could also be used as the source model in LCMM. However, the examination of their differences is beyond the scope of this paper.



(a) $\widetilde{s}_1$: Real and imaginary parts (b) $\widetilde{s}_1$ and $\widetilde{s}_2$: given a state (c) Mixture density

Figure 1: Illustration of complex multivariate Student-t distribution for two circular complex sources. (a,b): Bivariate Student-t pdf $\widetilde{\mathcal{T}}((\widetilde{s}_1, \widetilde{s}_2), \mathrm{diag}(\boldsymbol{b}_1), \nu)$ with $\boldsymbol{b}_1 = (1, .1)$ and $\nu = 2$; (a) illustrates pairwise marginal density on real and imaginary parts of single source $\widetilde{s}_1$, where spherical equiprobability contours imply circularity; the four types of pairwise marginals between $\widetilde{s}_1$ (real/imaginary) and $\widetilde{s}_2$ (real/imaginary) have the same form illustrated in (b) (black solid lines), where two other examples are also shown (gray solid lines). (c): Pairwise marginal densities for a mixture of the three Student-t pdfs given in (b): $\boldsymbol{b}_1 = (1, .1)$, $\boldsymbol{b}_2 = (.1, 1)$, $\boldsymbol{b}_3 = (.8, .8)$, $\boldsymbol{\eta} = (.1, .1, .8)$, and $\nu = 2$.

(c)), i.e., $p(\widetilde{\boldsymbol{s}}) \neq \prod_{j=1}^{d} p(\widetilde{s}_j)$, in contrast to standard complex-valued BSS/ICA models. Unlike the related generalizations of ICA modeling energy correlations, the sources are even conditionally dependent (Fig. 2(b)), given the higher-order latent variables (here: state k). This is due to our choice of the Student-t model instead of a Gaussian model. In the limit of $\nu \to \infty$ the pdf is reduced to a complex Gaussian and the sources become conditionally independent (Fig. 2(c)), but they remain dependent over the whole data set (with state k marginalized out).



(a) State k marginalized (b) With finite ν (c) With $\nu \to \infty$ (Gaussian)

Figure 2: Graphical representations of dependency structure in LCMM

2.3 Parameter estimation

According to LCMM's generative model, we obtain the pdf of complex sensor signal $\widetilde{\boldsymbol{x}}$ by the transformation of random variables from $\widetilde{\boldsymbol{s}}$ to $\widetilde{\boldsymbol{x}}$:

$$p(\widetilde{\boldsymbol{x}}; \mathbf{W}, \mathbf{B}, \boldsymbol{\eta}, \nu) = |\det \mathbf{W}|^2 \sum_{k=1}^{K} \eta_k \widetilde{\mathcal{T}}(\mathbf{W}\widetilde{\boldsymbol{x}}; \mathrm{diag}(\boldsymbol{b}_k), \nu), \tag{6}$$

where we explicitly indicate the model parameters in the left-hand side (after the semicolon), and $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K)$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_K)^\mathsf{T}$ collect the coactivation patterns and the state probabilities, respectively. Note that in Eq. (6), the determinant is squared because the same transformation is required for both real and imaginary parts. The pdf can also be expressed as

$$p(\widetilde{\boldsymbol{x}}; \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}, \nu) = \sum_{k=1}^{K} \eta_k \widetilde{\mathcal{T}}(\widetilde{\boldsymbol{x}}; \mathbf{A}^\mathsf{T}\mathrm{diag}(\boldsymbol{b}_k)\mathbf{A}, \nu). \tag{7}$$

The model is thus a constrained form of the mixture of multivariate Student-t distributions, in which the state-dependent scatter matrices are tied with common parameter $\mathbf{A}$. It can be readily seen from Eq. (7) that the probability density does not change if $\mathbf{A}$ (or $\mathbf{W}$) and $\boldsymbol{b}_k$ are simultaneously replaced by $\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ (or $\mathbf{D}^{\frac{1}{2}}\mathbf{W}$) and $\mathbf{D}\boldsymbol{b}_k$, respectively, where $\mathbf{D}$ is any non-singular diagonal matrix. This is the well-known scaling ambiguity inherent to BSS/ICA; we fix the scale by setting every column of $\mathbf{A}$ to have unit Euclidean norm.

The parameters of interest in LCMM can be easily estimated by the maximum likelihood method, i.e., maximizing $\sum_{t=1}^{N} \ln p(\widetilde{\boldsymbol{x}}(t); \mathbf{W}, \mathbf{B}, \boldsymbol{\eta}, \nu)$ with respect to the model parameters when the latent variables are marginalized out. Importantly, this does not require any approximation in contrast to other hierarchical BSS models. For simplicity, we fix degrees of freedom ν to a constant (we set $\nu = 2$ in Sections 3 and 4 below; this choice leads to an infinite variance and strong robustness) and learn the other parameters $\{\mathbf{W}, \mathbf{B}, \boldsymbol{\eta}\}$ since ν typically has only a small effect on the final solution (at least if set relatively small for ensuring robustness). The detailed form of the objective function and its derivatives are given in Appendix A.

We propose to use a quasi-Newton method[3] to efficiently optimize the likelihood using reparameterization (Salakhutdinov et al., 2003) given by

$$\eta_k = \frac{\exp(\lambda_k)}{\sum_{k'=1}^{K} \exp(\lambda_{k'})}, \tag{8}$$

so that the η_ks automatically satisfy constraints $\eta_k \geq 0$ and $\sum_{k=1}^{K} \eta_k = 1$. On the other hand, we don't constrain the scaling of $\mathbf{W}$ or $\mathbf{B}$ during the optimization, but rescale them after obtaining the final solution so that every column of $\mathbf{A}$ has a unit norm. Any standard optimization software can be used for solving unconstrained optimization on new parameter set $\{\mathbf{W}, \mathbf{B}, \boldsymbol{\lambda}\}$. We found that this quasi-Newton method is more efficient than the well-known expectation-maximization method (simulations not shown).

[3] We used a Matlab implementation of the limited-memory BFGS by Mark Schmidt, available at http://www.di.ens.fr/~mschmidt/Software/minFunc.html

After estimating the model parameters, the (real-valued) sources are separated by $\boldsymbol{s}(t) = \mathbf{W}\boldsymbol{x}(t)$ or by taking the real part of $\widetilde{\boldsymbol{s}}(t) = \mathbf{W}\widetilde{\boldsymbol{x}}(t)$ for any $\boldsymbol{x}(t)$ or $\widetilde{\boldsymbol{x}}(t)$. This is possible because the demixing matrix is the same for both the original and the Hilbert-transformed data. The states are inferred by computing their posterior probability:

$$p(k \mid \widetilde{\boldsymbol{x}}) = \frac{\eta_k \widetilde{\mathcal{T}}(\mathbf{W}\widetilde{\boldsymbol{x}}; \operatorname{diag}(\boldsymbol{b}_k), \nu)}{\sum_{k'=1}^{K} \eta_{k'} \widetilde{\mathcal{T}}(\mathbf{W}\widetilde{\boldsymbol{x}}; \operatorname{diag}(\boldsymbol{b}_k), \nu)}. \tag{9}$$

The maximum a posteriori (MAP) estimate of the state is given by taking the state that maximizes this posterior, which gives the final result of the model-based clustering of the source coactivations.

2.4 Choosing the number of states by BIC

Another important issue in learning mixture models is the choice of the number K of states or clusters. We use the Bayesian information criterion (BIC) to select the best K minimizing

$$\mathrm{BIC}(K) := -2 \ln \widehat{L} + M \ln N, \tag{10}$$

where $\widehat{L}$ denotes the maximum of the likelihood obtained numerically as explained above and the number of free parameters M in LCMM is given by $M = K - 1 + d^2 + Kd - d$. The use of BIC for the model order selection in mixture models has been extensively studied in statistics. There are theoretical results regarding statistical consistency (Keribin, 2000), and BIC often exhibits state-of-the-art performance, as shown empirically (Steele and Raftery, 2010).

2.5 Relation to previous two-layer extensions of BSS/ICA

Many previous attempts have been made to extend BSS/ICA based on Eq. (1), particularly to deal with the residual dependency structures between power s_j^2 or magnitudes $|s_j|$ of the sources often observed in ICA results (Hyvärinen and Hoyer, 2000; Hyvärinen et al., 2001a; Valpola et al., 2004; Karklin and Lewicki, 2005; Kawanabe and Müller, 2005; Zhang and Hyvärinen, 2010; Hirayama and Hyvärinen, 2012); these works were not necessarily concerned with EEG/MEG.

Initial developments made fixed prior assumptions about the dependencies of the sources without estimating any parameters in the source model (Hyvärinen and Hoyer, 2000). However, we are concerned with models in which the parameters in the second layer (i.e., connectivity patterns) are estimated as well (Karklin and Lewicki, 2005; Osindero et al., 2006; Köster and Hyvärinen, 2010). Typically, these models are based on a (generalized) linear model of squared sources $\boldsymbol{s}^2 := (s_1^2, s_2^2, \ldots, s_d^2)^{\mathsf{T}}$, as already proposed by (Hyvärinen et al., 2001a):

$$\mathrm{E}[\boldsymbol{s}^2(t) \mid \boldsymbol{u}(t)] = \phi(\mathbf{B}\boldsymbol{u}(t)), \tag{11}$$

where $\mathbf{B}$ and $\boldsymbol{u}(t) = (u_1(t), u_2(t), \ldots, u_K(t))^{\mathsf{T}}$ are the second-layer mixing matrix and source vector, and ϕ is a strictly monotonic function that is applied element-wise.

Given variance $\mathrm{E}[s_j^2 \mid \boldsymbol{u}]$, each source $s_j(t)$ is then usually assumed to follow a certain probability distribution: either Gaussian (Hyvärinen et al., 2001a; Valpola et al., 2004) or non-Gaussian (Karklin and Lewicki, 2005). The real-valued linear mixing (Eq. (1)) finally produces observed signals $\boldsymbol{x}(t)$.

The following are the main problem with these previous models for EEG/MEG connectivity analysis: 1) they do not properly model the envelopes of the oscillatory sources, and 2) the likelihood is often intractable without resorting to approximations, because the continuous latent variable $\boldsymbol{u}(t)$ needs to be integrated out. A recent study (Cadieu and Olshausen, 2012) on the statistical modeling of natural movies actually addressed the first issue without resolving the second one.

In fact, a close connection between LCMM and Eq. (11) is implied by the well-known fact that Student-t distribution belongs to the Gaussian scale-mixture family (see e.g., Bishop, 2006, for a real case). We can equivalently re-formulate our model by assuming that $\widetilde{s}_j(t)$ is a complex (circular) Gaussian conditionally on state k and introducing a scaling variable $u_k(t) \geq 0$ that follows an inverse Gamma distribution. The conditional variance can then be written:

$$\mathrm{E}[|\widetilde{\boldsymbol{s}}|^2(t) \mid u_k(t)]_k = u_k(t)\boldsymbol{b}_k, \tag{12}$$

where $|\widetilde{\boldsymbol{s}}|^2 := (|\widetilde{s}_1|^2, |\widetilde{s}_2|^2, \ldots, |\widetilde{s}_d|^2)^\mathsf{T}$ denotes the squared envelopes. Now consider a simplified complex-valued counterpart of the previously used Eq. (11) given by

$$\mathrm{E}[|\widetilde{\boldsymbol{s}}|^2(t) \mid \boldsymbol{u}(t)] = \mathbf{B}\boldsymbol{u}(t). \tag{13}$$

The LCMM in Eq. (12) has essentially the same form, if we can constrain it so that only a single variable u_k takes a non-zero value at a time.

Hence, although closely related, LCMM has notable differences from previous energy-correlation models. First, it models the (squared) envelopes $|\widetilde{\boldsymbol{s}}|^2$ instead of the (squared) magnitudes $\boldsymbol{s}^2$, thus properly dealing with oscillatory sources (together with Hilbert transform). Second, the model is tractable and fast to learn because it has only one discrete latent variable instead of multiple continuous ones.

2.6 Real-valued variant of LCMM

To separately evaluate the effect of using complex-valued formulation instead of a real-valued kind, we also examine a real-valued counterpart of LCMM in our simulation study below. The model can also be seen as a simplification of a previous two-layer BSS/ICA such that only a single variable u_k in Eq. (11) takes a non-zero value at a time, where nonlinearity ϕ is set to an identity function, implying that $\mathrm{E}[\boldsymbol{s}^2(t) \mid u_k(t)]_k = u_k(t)\boldsymbol{b}_k$. More specifically, the real-valued LCMM is given as a constrained form of a mixture of multivariate Student-t distributions and its pdf is given by

$$p(\boldsymbol{x}; \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}, \kappa) = \sum_{k=1}^{K} \eta_k \mathcal{T}(\boldsymbol{x}; \mathbf{A}^\mathsf{T}\mathrm{diag}(\boldsymbol{b}_k)\mathbf{A}, \kappa), \tag{14}$$

where the Student-t pdf for real vector $\boldsymbol{s} \in \mathbb{R}^d$ is generally given by

$$\mathcal{T}(\boldsymbol{s}; \boldsymbol{\Sigma}, \kappa) = \frac{\Gamma(\frac{d+\kappa}{2})}{\Gamma(\frac{\kappa}{2})(\kappa\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \left(1 + \frac{1}{\kappa}\boldsymbol{s}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{s}\right)^{-\frac{d+\kappa}{2}}. \tag{15}$$

Again, the pdf can be expressed using $\mathbf{W} = \mathbf{A}^{-1}$ as

$$p(\boldsymbol{x}; \mathbf{W}, \mathbf{B}, \boldsymbol{\eta}, \kappa) = |\det \mathbf{W}| \sum_{k=1}^{K} \eta_k \mathcal{T}(\mathbf{W}\boldsymbol{x}; \mathrm{diag}(\boldsymbol{b}_k), \kappa), \tag{16}$$

and the same quasi-Newton optimization is used for estimating $\{\mathbf{W}, \mathbf{B}, \boldsymbol{\eta}\}$ (specifically with $\kappa = 1$ in Sections 3 and 4 below). Finally, the state posterior is also given by

$$p(k \mid \boldsymbol{x}) = \frac{\eta_k \mathcal{T}(\mathbf{W}\boldsymbol{x}; \mathrm{diag}(\boldsymbol{b}_k), \kappa)}{\sum_{k'=1}^{K} \eta_{k'} \mathcal{T}(\mathbf{W}\boldsymbol{x}; \mathrm{diag}(\boldsymbol{b}_k), \kappa)}. \tag{17}$$

3 Simulations on artificial data

We next quantitatively compare the proposed method with existing approaches for the analysis of EEG/MEG data. We start with simulated EEG/MEG data so that the ground truth is known and can be systematically controlled; we provide a real EEG analysis in Section 4. The goal of the simulation study below is to validate the two key ideas of the proposed method: the complex-valued formulation and the unified estimation principle of the two stages of analysis.

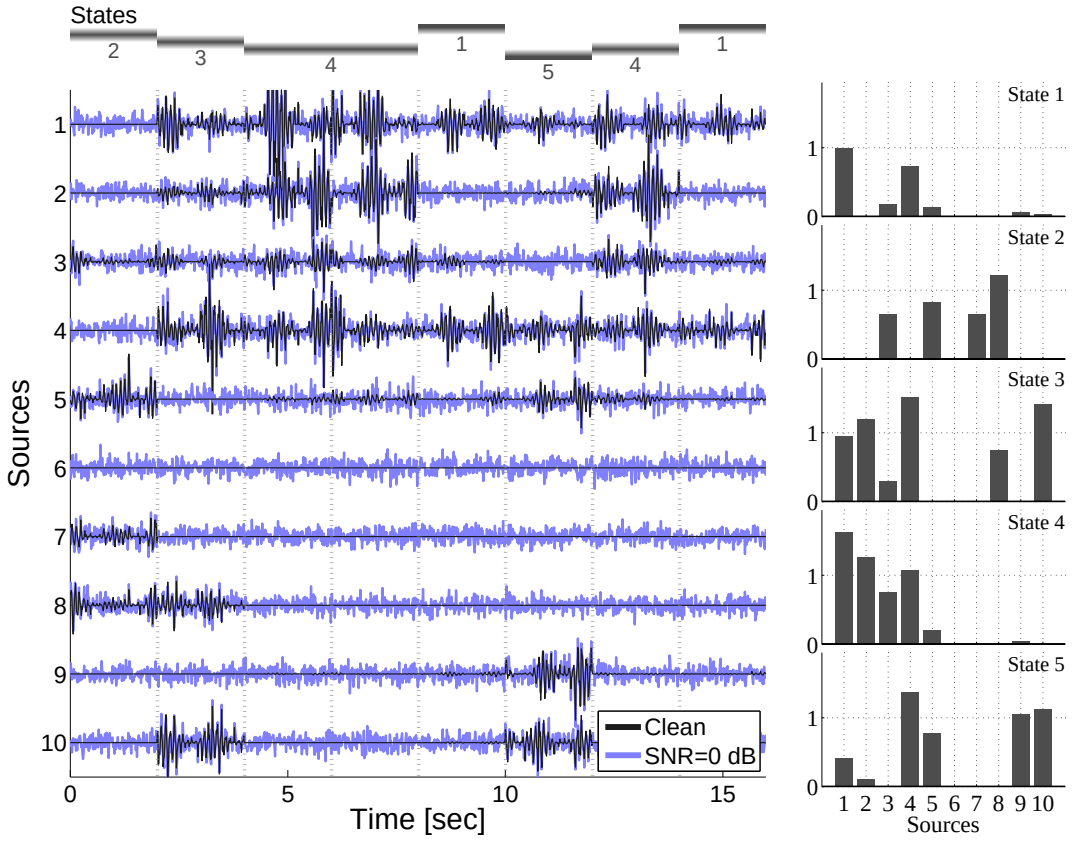
3.1 Methods

The simulated EEG/MEG signals were created as follows. First, we applied a band-pass filter (9.5-10.5 Hz) to ten Gaussian temporally white signals sampled virtually at 75 Hz to simulate the alpha-range activities. Then these oscillatory signals were jointly amplitude-modulated block-wise in every 2-second window (150 samples) to show the state-dependent coactivation patterns. For this purpose, we first created vectors $\boldsymbol{b}_k$ whose entries were independently sampled from a standard Gaussian distribution but set to zero if negative; this was repeated until at least two entries satisfied $b_{jk} \geq 0.05$. Then the j-th oscillatory signal was multiplied by $\sqrt{b_{jk}}$ with state k randomly chosen for each block with uniform probability ($k = 1, 2, \ldots, 5$). The sources with non-zero b_{jk}'s were actually coactivated, while very small activity levels were avoided with this procedure.

So that the amplitudes of these coactivated sources have non-zero (positive) correlations within each state, they were further modulated globally by a noisy sinusoidal signal, generated by sampling from a Gamma distribution $\mathrm{Gamma}(2\xi(t), 2)$ [4] with $\xi(t) = 0.9 \sin(2\pi f t/75 + \phi) + 1$ where $f = 1$ Hz and phase ϕ was randomly selected. Then Gaussian white noise was added, where the noise variance was set to have a given value of a signal-to-noise ratio (SNR), defined as the ratio of the variance. Fig. 3 illustrates an example of ten sources before and after adding the noise. Finally, sources $s_j(t)$ were linearly mixed into the same number (i.e., 10) of sensor signals $x_j(t)$ with square mixing matrix $\mathbf{A}$ generated randomly from the standard Gaussian distribution.

The clustering and source separation performances on these simulated data were compared among the following methods: 1) **LCMM**, 2) **LCMM (real)**, 3) **ICA+MixT**,

[4] The pdf is given by $p(x) = b^a x^{a-1} \exp(-bx)/\Gamma(a)$ if $x \sim \mathrm{Gamma}(a, b)$.



(a) Simulated sources (b) Coactivation patterns

Figure 3: Simulated data: Example of simulated sources and states. (a) Ten simulated sources $s_j(t)$ before (black) and after (gray) adding Gaussian noise with signal-to-noise ratio (SNR) of 0 dB. Clean sources (i.e., before adding noise) exhibit one of five different coactivation patterns $\boldsymbol{b}_k$ in each block (separated by vertical dotted lines) with their amplitudes co-modulated within each block (see text for more details). Top horizontal bars and numbers indicate five states $k = 1, 2, \ldots, 5$, corresponding to five coactivation patterns. (b) Five coactivation patterns. k-th panel shows values of b_{jk} for ten sources $j = 1, 2, \ldots, 10$.

and 4) **ICA+Kmeans**. The first one, **LCMM**, is the proposed approach, based on the joint maximum likelihood estimation of the two layers of the complex-valued LCMM. The second one, **LCMM (real)**, denotes the real-valued counterpart of **LCMM** (see Section 2.6), which can be seen as laying between our proposed method and the previous two-layer BSS/ICA models. The latter two, **ICA+MixT** and **ICA+Kmeans**, perform two-stage analysis. Both first used the complex-valued FastICA (Bingham and Hyvärinen, 2000) (with real-valued $\mathbf{W}$) for separating complex sources $\widetilde{\boldsymbol{s}}(t)$; then **ICA+MixT** directly learned the mixture of the Student-t model (as in LCMM) on the separated sources, while **ICA+Kmeans** performed standard k-means clustering on log-amplitudes $\ln|\widetilde{s}_j(t)|$, where the mean log-amplitude over the channels was subtracted at every t to compensate for the global modulation.

We used the adjusted mutual information (AMI) (Vinh et al., 2010) and the Amari index (Amari et al., 1996) as specific performance measures for clustering and source separation, respectively. AMI[5] corrects normalized mutual information (NMI) between true cluster k and estimated cluster $\widehat{k}$ for chance agreements: $\mathrm{AMI} = (\mathrm{NMI} - \overline{\mathrm{NMI}})/(1-\overline{\mathrm{NMI}})$ where $\mathrm{NMI} = I(k,\widehat{k})/\max\{H(k),H(\widehat{k})\}$ $(0 \le \mathrm{NMI} \le 1)$ and $\overline{\mathrm{NMI}}$ denotes the expectation of NMI under random permutations of the cluster labels where the numbers of clusters and cluster members are unchanged (I and H denote the sample mutual information and marginal entropy). AMI is thus expected to be zero under this random permutation and is upper-bounded by one; the upper bound is achieved only when the two clusterings are perfectly matched. The Amari index, which is a standard performance measure for linear BSS problems, is defined by

$$\text{Amari index} = \sum_{j=1}^{d}\left(\sum_{j'=1}^{d}\frac{|\chi_{jj'}|}{\max_k|\chi_{jk}|}-1\right)+\sum_{j=1}^{d}\left(\sum_{j'=1}^{d}\frac{|\chi_{jj'}|}{\max_k|\chi_{kj'}|}-1\right), \qquad (18)$$

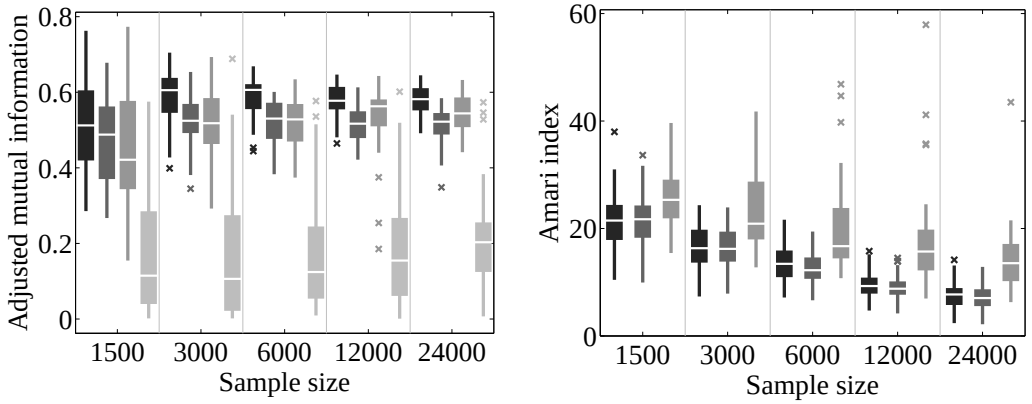
where $\chi_{jj'}$ denotes the (j,j')-th element of matrix $\mathbf{A}^{-1}\mathbf{A}^{\mathrm{true}}$ with estimated and true mixing matrices $\mathbf{A}$ and $\mathbf{A}^{\mathrm{true}}$, respectively. This index is nonnegative and equals zero if and only if the true mixing matrix is recovered up to the permutation and scaling of the columns.

3.2 Results

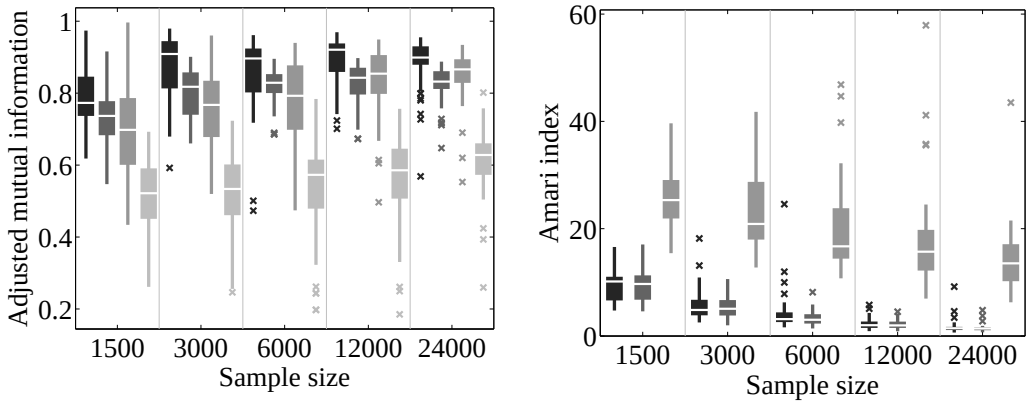
Figure 4 quantitatively compares the performances of clustering (panels on the left) and source separation (panels on the right) by the above four methods with different numbers of clusters estimated by the model, $K = 2, 5, 8$, while the number of true states is always 5. Each boxplot displays the result of 50 runs in each of the different sample sizes N. The SNR of the sources was specifically set to 20 dB in this figure. In the panels on the right, "ICA" corresponds to both **ICA+MixT** and **ICA+Kmeans**.

As is clearly seen in the left three panels, **LCMM** achieved the highest AMI (medians) in every condition, and **LCMM (real)** consistently showed a lower AMI. In contrast, on the right-hand-side panels, these two methods showed very similar Amari indices. These results imply that the complex-valued formulation of our LCMM is particularly beneficial for obtaining better clustering without degenerating the source sep-

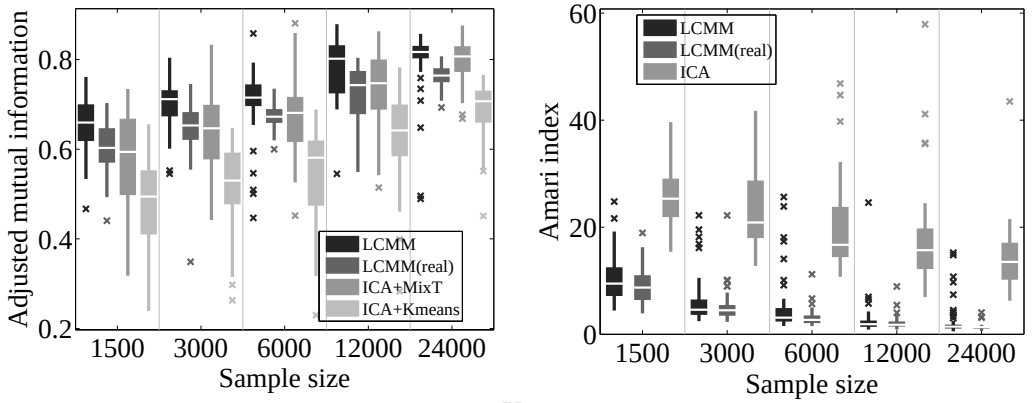
[5] We used the matlab code available at https://sites.google.com/site/vinhnguyenx/softwares.



(a) $K = 2$



(b) $K = 5$



(c) $K = 8$

Figure 4: Simulated data (with SNR=20dB): Performances in clustering (left) and source separation (right) with different settings of number of clusters estimated, (a) $K = 2$, (b) $K = 5$, and (c) $K = 8$, evaluated respectively by adjusted mutual information (AMI) (Vinh et al., 2010) and Amari index (Amari et al., 1996). Actual number of clusters was always five. AMI is scaled between 0 (completely random) and 1 (perfect clustering); Amari index becomes zero if estimated $\mathbf{A}$ recovers the true mixing matrix up to the permutation and scaling of the columns. Thus, on the left, high values are good; on the right, low values are good. Each panel displays boxplots at different sample sizes N. Each boxplot indicates median, interquartile range, and entire data range of 50 runs, excluding outliers indicated by 'x'. See text for legends of methods.

aration. The two-stage methods, **ICA+MixT** and **ICA+Kmeans**, also exhibited lower (i.e., worse) AMI values than those of **LCMM** (but not necessarily of **LCMM (real)**), while they exhibited higher (i.e., worse) Amari indices. The two-stage methods are less accurate in both clustering and source separation than the proposed (complex-valued) LCMM method.

Although we obtained the best performance with $K = 5$ (true number of clusters), the relative performance among the four methods was qualitatively similar for different Ks. That is, the proposed method outperforms other methods even when the number of clusters K is misspecified.

To further examine how the result changes with different noise levels, we also conducted simulations with different SNRs for generating the source signals. The number of clusters K in the model was simply set at the true one ($K = 5$). Fig. 5 shows the result in the same format as that of Fig. 4. The relative performance of the four methods was qualitatively similar to Fig. 4 in every SNR setting. This showed that **LCMM** improved clustering without degenerating the source separation over the other methods even when the SNR is relatively low.

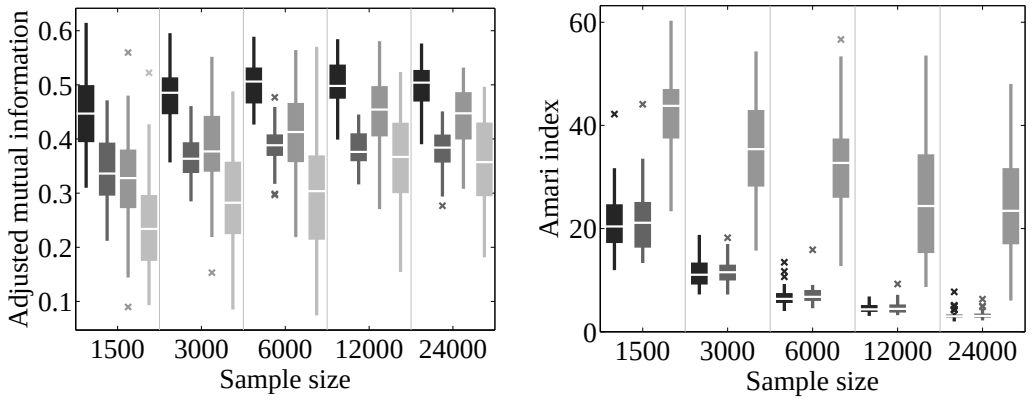
Finally, we demonstrated the use of BIC for selecting the number of clusters. Here, we computed the BIC for $K = 2, 3, \ldots, 10$ and chose the K that minimizes the value. Note that in the simulation setting here, LCMM does not completely match the true data-generating model due to the additional Gaussian noise in the sources. The number of clusters selected thus often exceeded five, as shown in Fig. 6, while higher SNRs (e.g., 20 or 30 dB) resulted in values closer to five. In practice, since EEG/MEG usually has a low SNR, these results indicate that the number of clusters will likely be overestimated by BIC. However, spatial topographies $\boldsymbol{a}_j$ also learned by LCMM can be used to identify and discard such irrelevant clusters that only contain noise or artifacts instead of physiologically meaningful patterns.

4 Experiments on resting-state EEG data

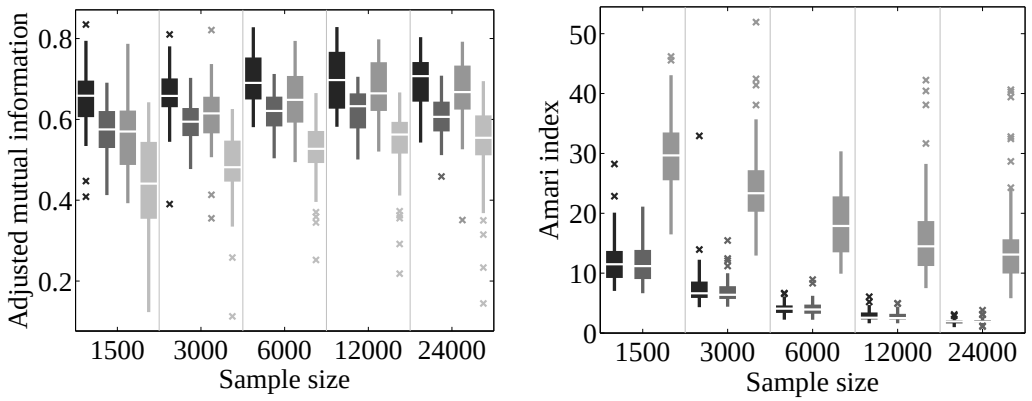
Next, we demonstrate the advantages of using LCMM compared to two-stage methods in a real EEG data analysis. The target data are resting-state EEGs acquired before and after a BCI-related task, which we expect to contain task-relevant brain states possibly due to mental rehearsal or retrieval. We examined the patterns (states) found in the resting-state EEGs based on the labeled EEG data during task as well as their spatial topographies on sensor channels.

4.1 EEG data

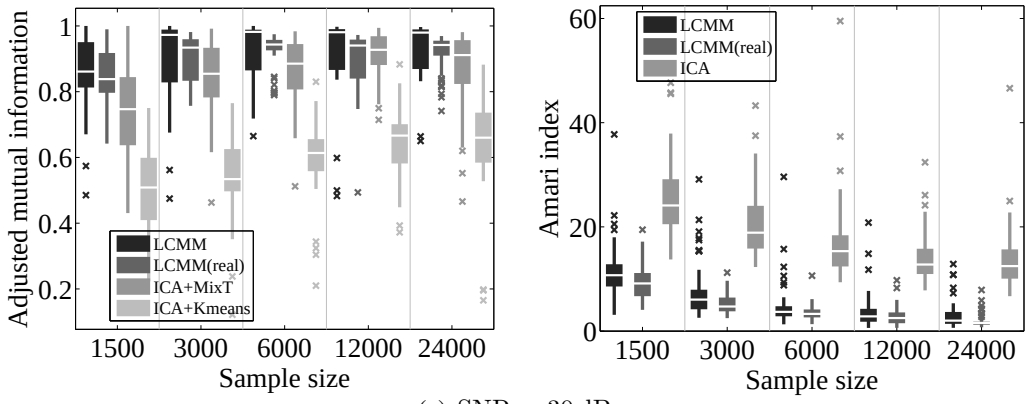
Five healthy subjects (three males, two females, 28+/-11 years old) participated in our EEG experiment. We placed a headcap with EEG electrodes on their heads with electric-conductive gel SIGNAGEL (Parker Laboratories Inc., Fairfield, NJ, USA) to reduce the impedance of the electrodes. We positioned 64-channel active electrodes based on the international 10-20 system, and connected them to an ActiveTwo amplifier (BioSemi, Amsterdam, The Netherlands). An experimental protocol of this study was approved by the ethical committee at ATR.



(a) $\mathrm{SNR} = 0\,\mathrm{dB}$



(b) $\mathrm{SNR} = 10\,\mathrm{dB}$



(c) $\mathrm{SNR} = 30\,\mathrm{dB}$

Figure 5: Simulated data (with $K = 5$ (true)): clustering (left) and source separation (right) performance with different noise levels. See caption of Fig. 4 for details.

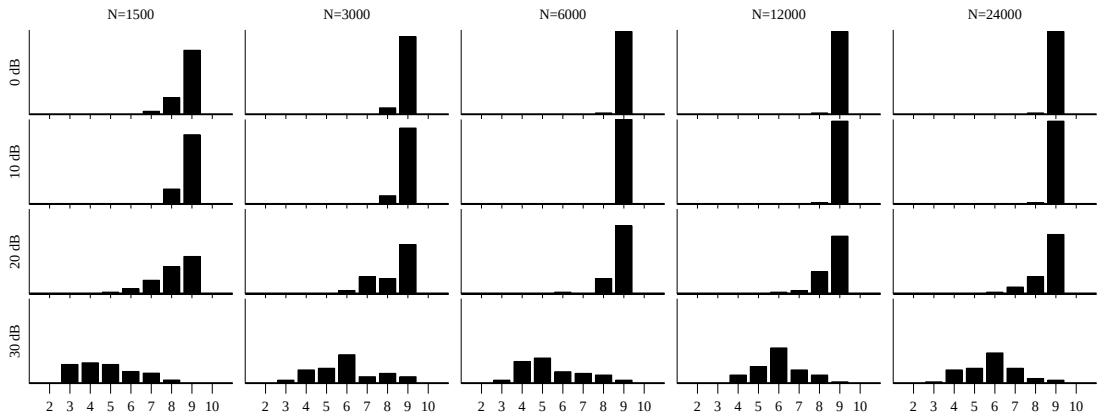


Figure 6: Simulated data: normalized counts of number of clusters K selected by minimizing BIC over $K = 1, 2, \ldots, 10$. Rows and columns correspond respectively to different noise levels (signal-to-noise ratios) of simulated sources and to different sample sizes N. True number of clusters was five.

The brain activity was measured in each subject during resting states with eyes open and while the subjects performed a cued motor imagery/non-imagery task. The experiment consisted of two resting-state (RS) sessions and six task sessions between the two RS sessions. In each RS session (5 minutes), the subject was instructed to relax without thinking of anything in particular and without sleeping and to focus on a fixation point at the screen's center. In each task session, the subjects performed a number of task trials in each of which they randomly took one of the following three actions for three seconds after a visually-cued onset: 1) *left*: covert imagination of a left-hand movement, 2) *right*: covert imagination of a right-hand movement, and 3) *idle*: no imagination of hand movements.

The EEG data were acquired at a sampling frequency of 256 Hz, band-pass filtered off-line to 1-50 Hz (fourth-order Butterworth, zero-phase), and re-referenced to the common average. Part of the RS data and some trials in the task data were rejected due to gross contamination. Typical ocular, cardiac, and muscular artifacts were also identified and removed by FastICA (Hyvärinen, 1999) with visual inspection and frequency analysis, which was done separately for each subject and also for the RS and task data. The data were further band-pass filtered in certain frequency bands of interest and then Hilbert-transformed. We focused on two frequency bands of interest, 8-12 Hz (alpha) and 13-30 Hz (beta), in line with previous neurophysiological studies on motor imagery (e.g., Pfurtscheller and Neuper, 1997) as well as those on resting-state brain networks (e.g., Brookes et al., 2011).

4.2 Method

All the LCMM parameters were estimated from the RS data alone where the two (pre- and post-) RS sessions were combined. We emphasize that the task data and labels were not used for the parameter estimation but only to validate the learned model in a post-hoc manner. Before the parameter estimation, the RS data were spatially prewhitened and dimensionality-reduced by PCA, so that 99% of the sample variance

was kept. Note that the effective dimensionality of the data had already been reduced above by removing the artifactual dimensions using ICA. As a result, the number of sources d was selected as $12, 12, 11, 7$, and 31 for the five subjects in the alpha band and $17, 12, 17, 14$, and 31 in the beta band. For comparison, we also applied the two-stage method **ICA+MixT**, as explained in Section 3.1, to the same data to examine the effect of unifying the two stages of the analysis by **LCMM**. In both methods, we ran the algorithm ten times, each from different initial parameters to converge, for every preselected number of states $K = 10, 20, \ldots, 100$. The best K in **LCMM** was chosen so that the median of the ten BIC values achieved the minimum, and the same number K was also used in **ICA+MixT**.

We then evaluated how well each coactivation pattern $\boldsymbol{b}_k$ found in the RS data discriminates the two physiologically different brain states corresponding to the motor imagery (i.e., *left* and *right*) and non-imagery (i.e., *idle*) labeled in the task EEG data. We used a standard performance criterion for binary discrimination, the area under the receiver operating characteristic (ROC) curve or AUC and evaluated them as follows. We used the k-th signal model $p(\widetilde{\boldsymbol{x}} \mid k)$ to detect whether a task trial was in the k-th state. This state was supposed to be detected if log-likelihood $\ln p(\widetilde{\boldsymbol{x}} \mid k)$, averaged over the imagery/non-imagery period after the cued onset (where the initial 0.5 seconds of this 3-second period were discarded to avoid the transient effect), was above or below a certain threshold. The log-likelihood of the k-th state, up to the irrelevant scaling and additive constants, can be evaluated at each time point by $\ln(1 + \sum_{j=1}^{d} b_{jk}^{-1} |\widetilde{s}_j|^2)$, as derived by setting $\nu = 2$ in Eq. (5) and removing the time-invariant constant terms from its logarithm. The occurrence of the k-th state may be associated arbitrarily with motor imagery or with non-imagery for each case of which the ROC curve was drawn by plotting the true positive rate against the false positive rate at many different threshold values. Thus we had two AUC values (computed by a trapezoidal rule) for each state from which the greater one was simply chosen; the AUC becomes close to one if the state occurrence discriminates well between motor imagery and non-imagery, and it is close to 0.5 if the occurrence does not discriminate it at all. Note that this AUC does not distinguish between *left* and *right* because that seemed too difficult in these data based on our preliminary analysis (not shown here).

To compare these states that exhibited high AUC values, we further examined the time courses of their log-likelihood and the sensor-level topographies corresponding to them. The topographies were drawn by evaluating how the occurrence of each state changed the power in the frequency band of interest at each site of the electrodes. We first obtained a robust estimate of the state-conditional variance (power) for each sensor channel by $\sigma_{ck}^2 \propto \sum_j b_{jk} a_{cj}^2$ ($c = 1, 2, \ldots, 64$), where a_{cj} denotes the estimated mixing coefficient from the j-th source to the c-th sensor channel and total variance $\sum_c \sigma_{ck}^2$ was simply normalized to one because it was undefined due to the choice of $\nu = 2$. The topographies were then plotted by spatially interpolating the percentage deviations $100 \times (\sigma_{ck}^2 - \overline{\sigma}_c^2)/\overline{\sigma}_c^2$ from the "grand variance" $\overline{\sigma}_c^2$ over 64 electrodes (see Fig. 7 for the layout and channel names). Grand variance $\overline{\sigma}_c^2$ was set to the expectation of state-conditional variances σ_{ck}^2 during the task period, i.e., $\overline{\sigma}_c^2 = \sum_k \eta_k \sigma_{ck}^2$, where η_k was re-estimated by the posterior state probabilities during the imagery/non-imagery periods of the task trials.

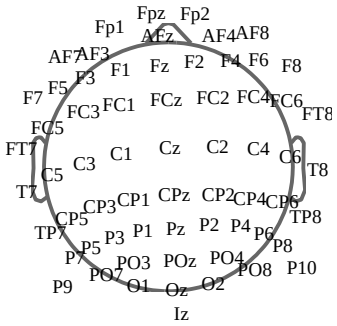


Figure 7: Real EEG analysis: Channel names and spatial layout of 64 electrodes. Layout corresponds to spatial topographies on scalp shown in Figs. 9 and 10.

4.3 Results

The discriminability of each state between the motor imagery and idling trials was examined in terms of the sample distribution of their AUC values. Fig. 8 shows the normalized histograms of the AUC collected from the ten different runs of the parameter estimation, where each panel shows the two normalized histograms by **LCMM** ("unified") and **ICA+MixT** ("2-stage") for a particular subject and frequency band. The number of samples (AUC values) summarized in each panel was thus $10 \times K$, where number of states K was selected by BIC as $K = 70, 70, 90, 20$ and 90 for the five subjects in the alpha band, and $K = 30, 10, 20, 10$ and 70 in the beta band. The figure shows that in most cases, the unified method exhibited greater variation in AUC than the two-stage method. For example, the range of AUC clearly expanded to both sides of the distribution in Subjects 1 (beta), 3 (beta), and 5 (alpha), while mostly to the right (greater) side in Subjects 3 (alpha), 4 (alpha, beta) and 5 (beta). The greater variation in the AUC distribution, even if it was both-sided, implies that it has a greater chance to contain brain states that exhibit higher AUC values. The two vertical lines in each panel further indicate the upper quartiles of AUC for the unified (solid line) and two-stage methods (dashed line), showing consistent increases of the upper quartiles by **LCMM** from those by **ICA+MixT**. In other words, the lower bound of the top 25% of the AUC values consistently shifted to the right as a consequence of unifying the two stages of analysis.

To get insights into the difference of these high-AUC states between the unified and two-stage methods, we further analyzed the average dynamics of the states that exhibited the best (highest) AUC in each subject and frequency band. Figs. 9 and 10 display the trial-averaged time courses of the log-likelihood of those states obtained for the alpha and the beta bands, respectively, around the 3-second period of motor imagery/non-imagery. For the alpha band (Fig. 9), the dynamics clearly differ between the motor imagery (*left* and *right*) and the non-imagery (*idle*) conditions in **LCMM** (left column), especially in Subjects 1, 2, and 5, with smaller differences among them in **ICA+MixT** (right column). For the beta band (Fig. 10), the dynamics again exhibited similar differences between the imagery and non-imagery conditions in every subject. The results by **LCMM** (left column) and **ICA+MixT** (right column) are very similar in Subjects 2, 3, and 4, but in Subjects 1 and 5 the dynamics again clearly differ between

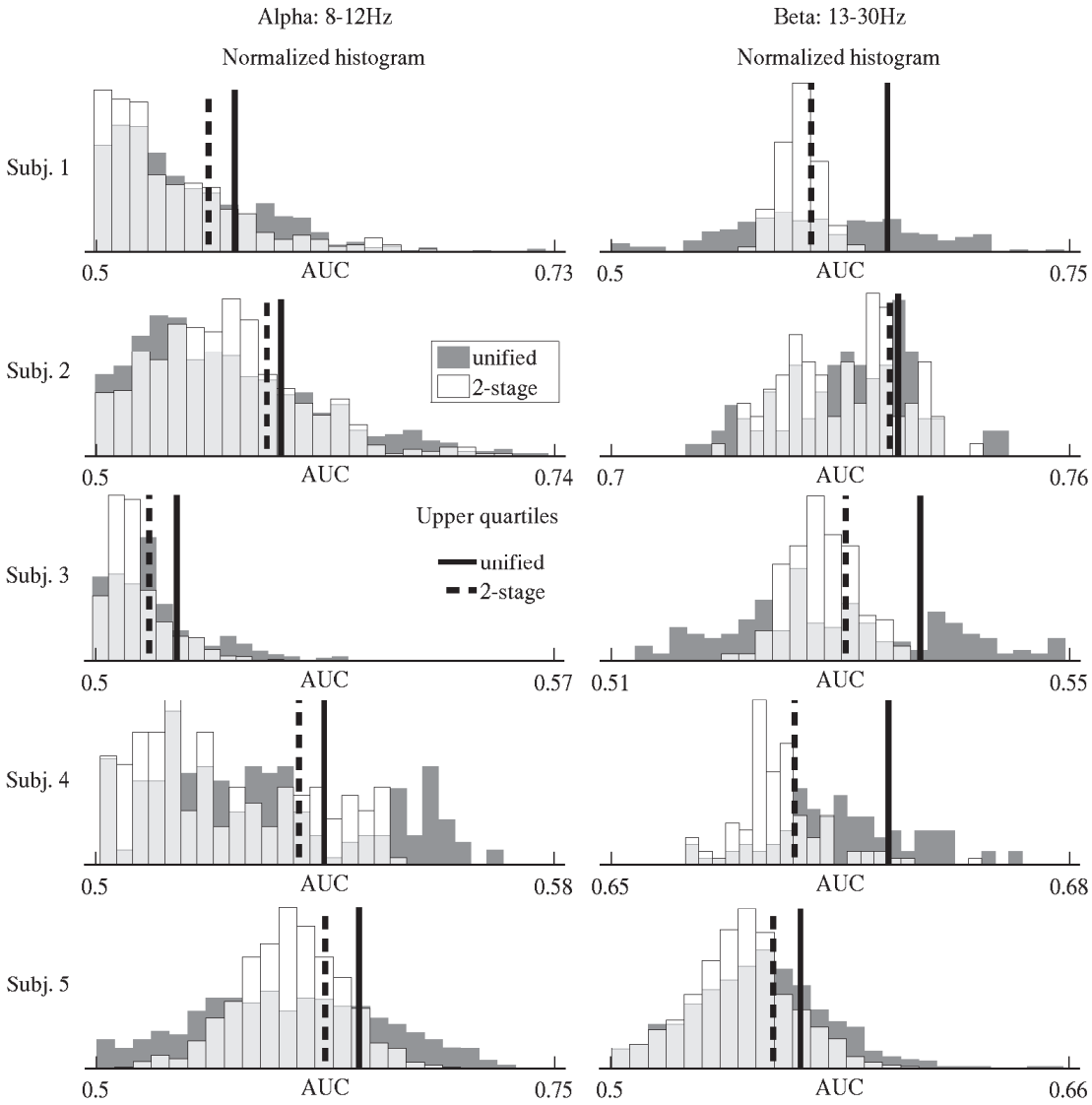


Figure 8: Real EEG analysis: Distributional difference of AUC values evaluated for each obtained state between unified and two-stage methods. Each panel shows two histograms corresponding to **LCMM** (unified) and **ICA+MixT** (2-stage) for particular subjects and frequency bands, as indicated on left and right sides of figure, respectively. In each panel, two normalized histograms are superimposed with their overlap shown by transparency. Vertical axis is scaled in each panel, and horizontal axis denotes AUC. Two vertical lines indicate upper quartiles for two methods: **LCMM** (solid line) and **ICA+MixT** (dashed line).

the imagery and non-imagery conditions in **LCMM** with fewer differences among them in **ICA+MixT**.

The corresponding spatial topographies of the power deviations, given in Figs. 9 and 10, provide further neurophysiological insights in conjunction with those of the averaged dynamics. For example, Fig. 9 (left column) suggests that the best-AUC states of Subjects 1 and 2 are both associated with the event-related decrease of frontal alpha power in the *idle* condition but not in the two motor imagery conditions. This is readily seen in the decrease of the log-likelihood during the *idle* condition (green line) from the baseline level and with those topographies that exhibit positive values on the central areas. The figure also suggests that the best-AUC state of Subject 5 is associated with the event-related decrease of the alpha powers during the *left* and *right* conditions at the bilateral Rolandic (central) areas, which are the major regions of interest related to motor imagery (Pfurtscheller and Neuper, 1997). This is seen in the log-likelihood increase during motor imagery in conjunction with the negative deviations in the alpha power around those areas, as indicated by the topographies. The topographies obtained for the beta band (Fig. 9) seem more difficult to interpret. Further neurophysiological interpretation is beyond the scope of this paper.

These topographic changes in power, seen at the sensor level, are actually caused by different coactivation patterns at the level of the underlying sources. Fig. 11 shows an example of the coactivation patterns ($\boldsymbol{b}_k$) and the topographies ($\boldsymbol{a}_j$ and their element-wise squares) corresponding to each source obtained by LCMM. The top row shows the coactivation pattern of the twelve sources, obtained for Subject 2 in the alpha band, which corresponds to the best-AUC state shown in Fig. 9 (left column, Subject 2). Here in this state, sources 3 and 10 have the largest powers, followed by 5, 6, and 7, and the rest have relatively small powers. With the coactivation patterns for other states, perhaps these high-AUC states can be characterized by the relative deactivation of sources 1 and 2 or the relative activation of sources 5, 6, and 7, for example. Even though we omit the details, the topographies given at the bottom will be useful for further interpretation.

5 Discussion

We presented a novel unsupervised method for non-stationary functional connectivity analysis of EEG/MEG sources. Our simulation studies confirmed that the proposed unified method often outperformed the conventional two-stage method, in terms of both source separation and clustering performances. Real EEG data analysis also showed that the unified method finds coactivity patterns that discriminate well between motor imagery and idling states with a higher probability than the two-stage method. In addition, the proposed method performs clustering in the source space to give further neurophysiological insights beyond sensor-space clustering (e.g., Britz et al., 2010; Shi and Lu, 2008), as demonstrated in Section 4.3.

Our real EEG data analysis suggested that the log-likelihood values of the states, obtained by our unsupervised analysis, may provide better higher-order signal features relevant to BCI (e.g., for the onset detection of motor imagery) than those by conventional methods. However, as seen in Section 4.3, LCMM often did not consistently improve the discriminability (AUC) of every state but rather strengthened their con-

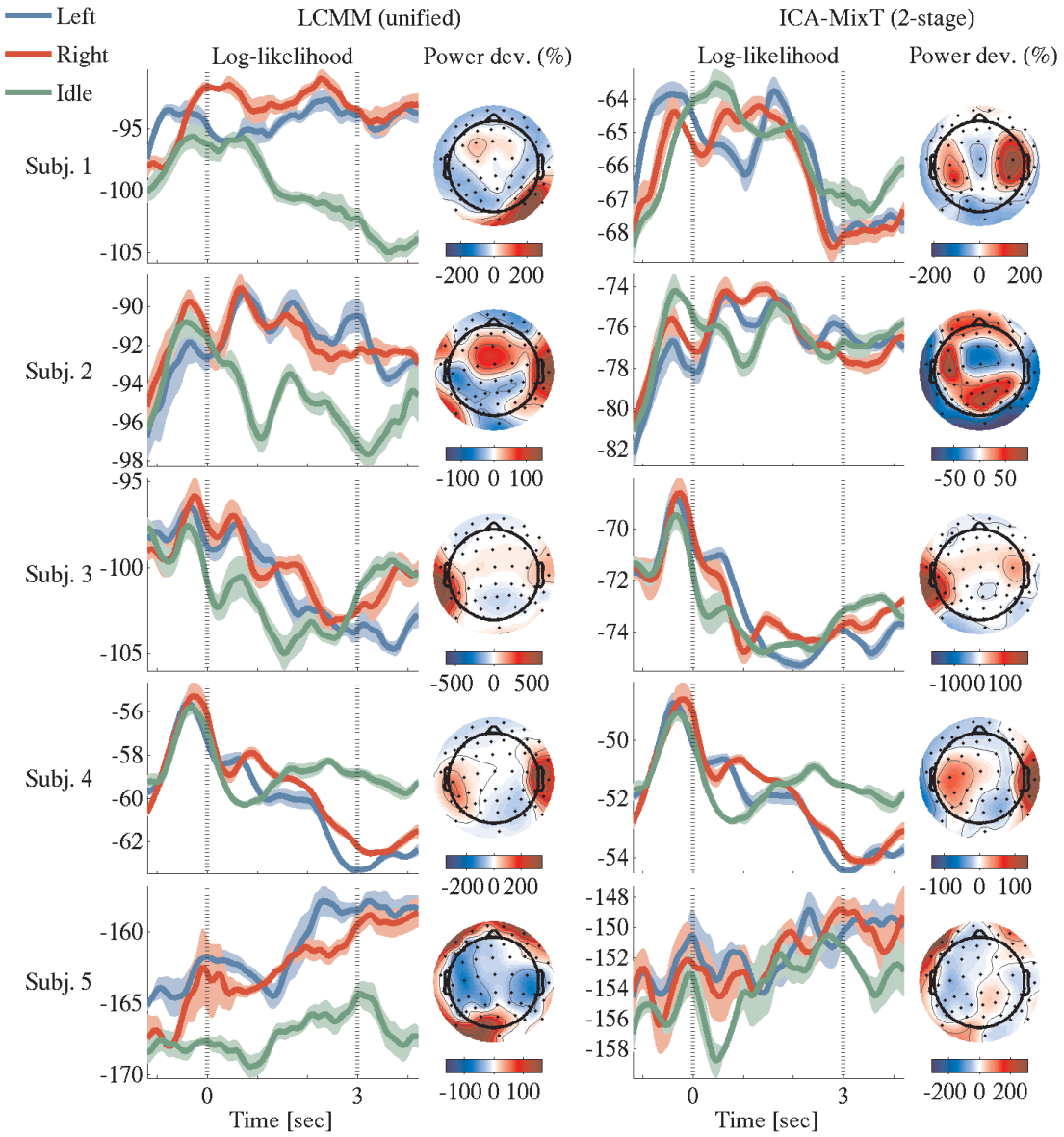


Figure 9: Real EEG analysis: Temporal dynamics of log-likelihood of states that exhibited best (highest) AUC in Fig. 8 for alpha band (8-12 Hz). Each panel displays time courses of log-likelihood, trial-averaged in each of three task conditions, *left* (blue), *right* (red), and *idle* (green), for a particular subject and by either **LCMM** (unified) or **ICA+MixT** (2-stage), as indicated on left and top of figure. Solid lines indicate moving-averages using time windows of 0.5 seconds, where shaded intervals indicate standard deviation in each moving window. Corresponding scalp topographies shows percentage deviations in frequency-band power at each electrode (Fig. 7 for channel names) interpolated spatially, which represents how the occurrence of state changes the power from the grand mean at the sensor level.

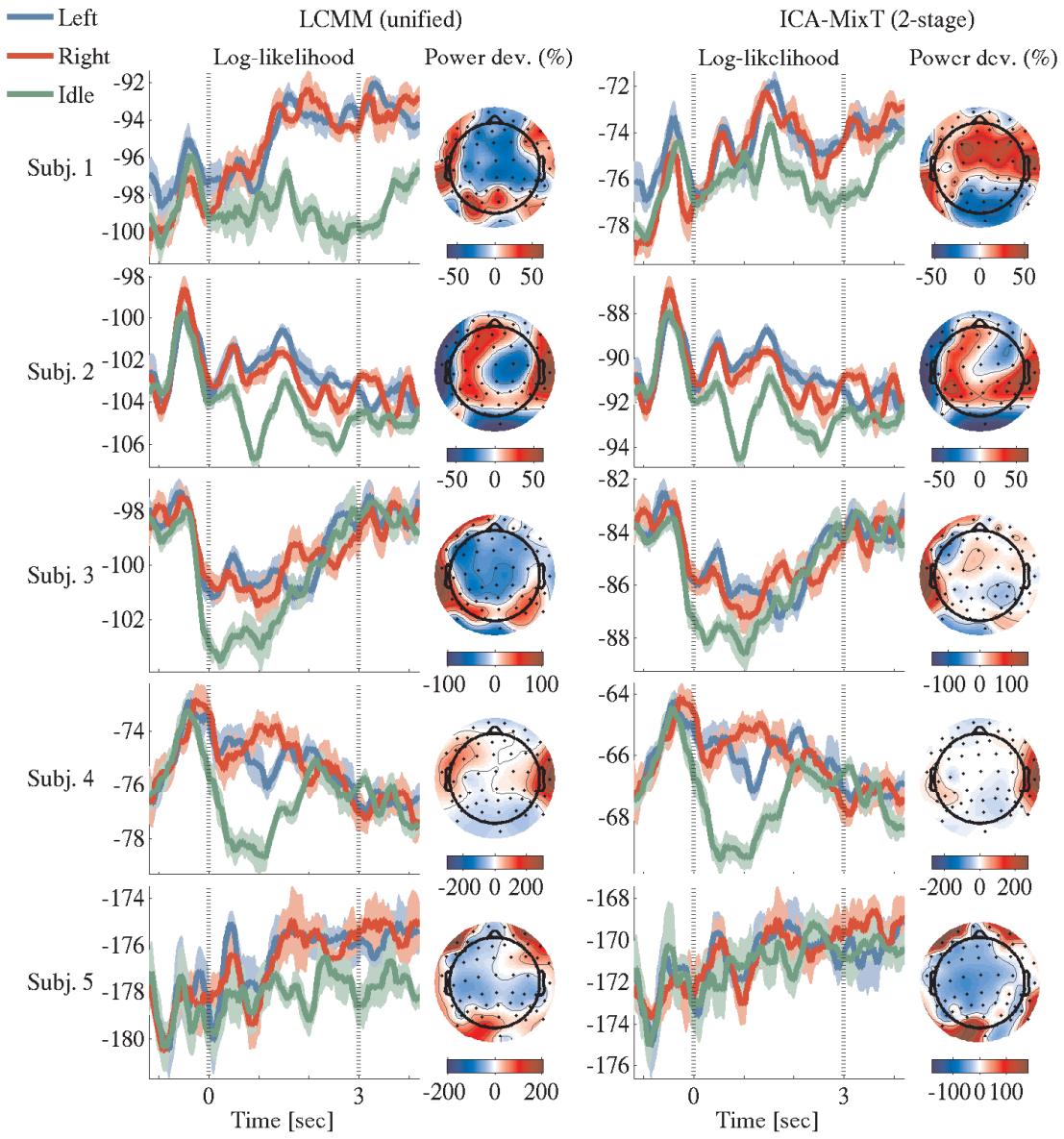


Figure 10: Real EEG analysis: Temporal dynamics of log-likelihood of states that exhibited best (highest) AUC in Fig. 8 for beta band (13-30 Hz). See Fig. 9 for more details. Moving-average was taken by time windows of 0.25 seconds.

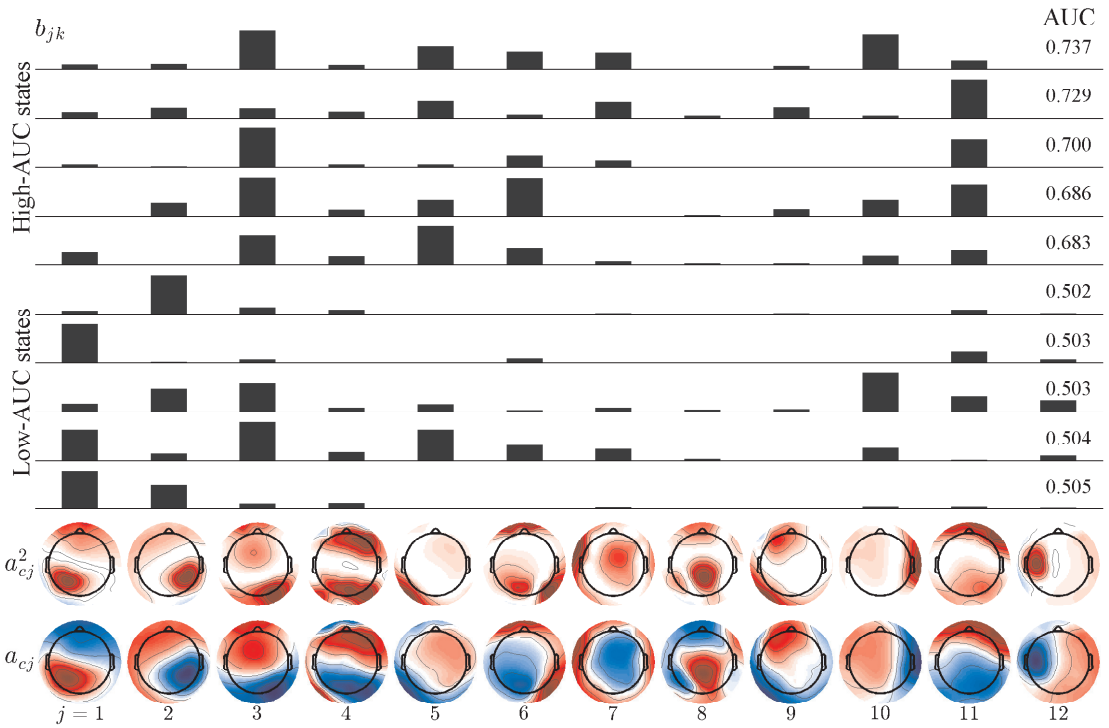


Figure 11: Real EEG analysis: Examples of coactivation patterns $\boldsymbol{b}_k$ of sources and spatial topographies corresponding to each source obtained by LCMM (Subject 2, alpha band). The ten barplots display coactivations associated with ten different states k. Five from top are those of highest AUC values, and the rest are lowest AUC values, as indicated at right of each row. Vertical length of the j-th bar (from left) in the k-th row represents relative value of b_{jk} with vertical axis scaled separately in each row. Scalp topographies at bottom show values of mixing coefficients a_{cj} (lower) and its squared a_{cj}^2 (upper), where blue and red colors correspond to negative and positive signs, respectively. Note that the signs of coefficients a_{cj} may be arbitrarily flipped for each j due to indeterminacy inherent to LCMM (or ICA). Numbers j at bottom indicate twelve sources.

trasts: the states obtained by LCMM may contain both more discriminative (higher AUC) and more indiscriminative (lower AUC) ones than those by the two-stage methods. Hence, we must carefully select the discriminative features (states) while avoiding the indiscriminative ones to successfully apply our method to BCI. Such a feature selection could be done, e.g., based on neurophysiological interpretations or using additional task-based experimental calibration. This remains open for future investigation.

Some recent studies have combined BSS with effective (directional) connectivity analysis to analyze the causality between neural activities, for example, autoregressive (AR) models (Gómez-Herrero et al., 2008; Haufe et al., 2010, see also Fukushima et al. (2015) to solve the inverse problem rather than BSS), generalized autoregressive conditional heteroscedasticity (GARCH) models (Zhang and Hyvärinen, 2010), or a structural equation model (SEM) (Hirayama and Hyvärinen, 2012). Our idea of unifying BSS and EEG/MEG connectivity analysis in a hierarchical statistical model is thus not completely novel, but in the present study, we focused on a different type of statistical connectivity: functional connectivity based on envelope correlations. More importantly, we focused on analyzing non-stationary functional connectivity in terms of underlying patterns/states and their dynamics. This is a conceptually crucial difference from previous studies that were concerned with static connectivity, i.e., those unchanging over time.

The LCMM proposed here was shown to be a reasonable statistical model to perform our specific analysis on resting-state EEG/MEG signals, but obviously it has some limitations from the perspective of generative signal modeling. First, each LCMM state can describe only the positive correlations in the source envelopes, as is readily seen from the interpretation of Eq. (12), where latent factor $u_k(t)$ and coefficients $\boldsymbol{b}_k$ are both positive. The states themselves are, on the other hand, correlated negatively since they do not occur simultaneously due to the assumption of the finite mixture model. Hence, our method cannot analyze any brain states characterized by negative envelope correlations or states that are positively correlated with each other. Second, the i.i.d. assumption is too simplistic since brain activity is obviously not independent over time. More general models corresponding to Eq. (13), possibly with nonlinearly ϕ in Eq. (11) as well as some temporal dynamics model on latent variables, will make the model more realistic. However, the model complexity must be carefully controlled to achieve good predictability and to maintain the model's tractability.

Another practical limitation of our method is that it can currently handle only a single frequency band of interest. Although it is very typical to limit an EEG/MEG analysis to a certain frequency band, cross-frequency interactions are sometimes of particular interest. Some recent studies have actually combined a complex-valued ICA with time-frequency decomposition for spontaneous EEG/MEG analysis with more flexibility on the spectral nature (e.g., Hyvärinen et al., 2010a; Ramkumar et al., 2012). A similar technique can probably be used to extend our method to analyze the data in multiple frequency bands.

Acknowledgments We gratefully thank Dr. Motoaki Kawanabe and Dr. Okito Yamashita for their helpful comments and discussions and Dr. Takayuki Suyama who provided the opportunity for this research. This work was supported by a contract with the

Ministry of Internal Affairs and Communications entitled, 'Novel and innovative R&D making use of brain structures' and by JSPS KAKENHI Grant Number 25730155.

A Maximum likelihood estimation of LCMM

In our LCMM, the maximum likelihood estimates of parameters of interest $\{\mathbf{W}, \mathbf{B}, \boldsymbol{\eta}\}$ are given by the minimizer of the negative (average) log-likelihood $\overline{L} := (1/N)\sum_{t=1}^{N} L(t)$, where t-th term, $L(t) := -\ln p(\widetilde{x}(t); \mathbf{W}, \mathbf{B}, \boldsymbol{\eta}) + \text{const.}$, is given by

$$L(t) = -\ln \sum_{k=1}^{K} \eta_k \exp\left[-G\left(\sum_{j=1}^{d} \frac{|\boldsymbol{w}_j^{\mathsf{T}} \widetilde{\boldsymbol{x}}(t)|^2}{b_{jk}}\right) - \sum_{j=1}^{d} \ln b_{jk}\right] - 2\ln|\det \mathbf{W}|, \quad (19)$$

where $G(u) = (d + \frac{\nu}{2}) \ln(1 + 2u/\nu)$, according to Eqs. (5) and (6). We minimize $\overline{L}$ by a quasi-Newton method with respect to $\mathbf{W}$, $\mathbf{B}$, and $\boldsymbol{\lambda}$ using the reparametrization of Eq. (8). This requires the first derivatives of $L(t)$, given by

$$\frac{\partial L(t)}{\partial b_{ik}} = q_k(t)\left\{\frac{1}{b_{ik}} - g\left(\sum_{j=1}^{d} \frac{|\widetilde{s}_j(t)|^2}{b_{jk}}\right)\frac{|\widetilde{s}_i(t)|^2}{b_{ik}^2}\right\}, \quad (20)$$

$$\frac{\partial L(t)}{\partial w_{ij}} = \left\{\sum_{k=1}^{K} \frac{q_k(t)}{b_{ik}} g\left(\sum_{j=1}^{d} \frac{|\widetilde{s}_j(t)|^2}{b_{jk}}\right)\right\}(\widetilde{s}_i(t)\widetilde{x}_j^*(t) + \widetilde{s}_i^*(t)\widetilde{x}_j(t)) - 2\mathbf{W}_{ij}^{-\mathsf{T}}, \quad (21)$$

$$\frac{\partial L(t)}{\partial \lambda_k} = q_k(t) - \eta_k, \quad (22)$$

where $g(u) := G'(u) = (\nu + 2d)/(\nu + 2u)$, asterisk $\cdot^*$ denotes a complex conjugate, $\mathbf{W}_{ij}^{-\mathsf{T}}$ denotes the (i,j)-element of the transposed inverse matrix of $\mathbf{W}$. $q_k(t)$ denotes the posterior probability of the k-th state, given by

$$q_k(t) \propto \eta_k \exp\left(-G\left(\sum_{j=1}^{d} \frac{|\widetilde{s}_j(t)|^2}{b_{jk}}\right) - \sum_{j=1}^{d} \ln b_{jk}\right), \quad (23)$$

where $\propto$ means that the left-hand side is proportional to the right-hand side up to a constant factor independent of k. Notice that Eq. (23) is equivalent to Eq. (9). To obtain Eq. (21) above, we used the relation given by

$$\frac{\partial|\widetilde{s}_i|^2}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}}\left\{\mathrm{Re}[\boldsymbol{w}_i^{\mathsf{T}}\widetilde{\boldsymbol{x}}]^2 + \mathrm{Im}[\boldsymbol{w}_i^{\mathsf{T}}\widetilde{\boldsymbol{x}}]^2\right\} \quad (24)$$

$$= 2(\boldsymbol{w}_i^{\mathsf{T}}\,\mathrm{Re}[\widetilde{\boldsymbol{x}}])\,\mathrm{Re}[\widetilde{x}_j] + 2(\boldsymbol{w}_i^{\mathsf{T}}\,\mathrm{Im}[\widetilde{\boldsymbol{x}}])\,\mathrm{Im}[\widetilde{x}_j] \quad (25)$$

$$= 2\,\mathrm{Re}[\widetilde{s}_i]\,\mathrm{Re}[\widetilde{x}_j] + 2\,\mathrm{Im}[\widetilde{s}_i]\,\mathrm{Im}[\widetilde{x}_j] \quad (26)$$

$$= \widetilde{s}_i\widetilde{x}_j^* + \widetilde{s}_i^*\widetilde{x}_j, \quad (27)$$

where $\mathrm{Re}[z]$ and $\mathrm{Im}[z]$ denote the real and imaginary parts of complex number z, respectively.

References

Allen, E. A. et al. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex*, 24(3):663–676.

Amari, S., Cichocki, A., and Yang, H. H. (1996). A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems 8 (NIPS1995)*, pages 757–763.

Bingham, E. and Hyvärinen, A. (2000). A fast fixed-point algorithm for independent component analysis of complex valued signals. *International Journal of Neural Systems*, 10(1):1–8.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Britz, J., Ville, D. V. D., and Michel, C. M. (2010). BOLD correlates of EEG topography reveal rapid resting-state network dynamics. *NeuroImage*, 52:1162–1170.

Brookes, M. J. et al. (2011). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proceedings of the National Academy of Sciences of the United States of America*, 108(40):16783–16788.

Cadieu, C. F. and Olshausen, B. A. (2012). Learning intermediate-level representations of form and motion from natural movies. *Neural Computation*, 24(4):827–866.

Dähne, S. et al. (2014). Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, 96:334–348.

de Pasquale, F. et al. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):6040–6045.

Doesburg, S. M. and Ward, L. M. (2009). Synchronization between sources: Emerging methods for understanding large-scale functional networks in the human brain. In Perez-Velazquez, J. L. and Wennberg, R., editors, *Coordinated Activity in the Brain*, pages 25–42. Springer, New York.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2:56–78.

Fukushima, M., Yamashita, O., Knösche, T., and Sato, M. (2015). MEG source reconstruction based on identification of directed source interactions on whole-brain anatomical networks. *NeuroImage*, 105:408–427.

Gómez-Herrero, G., Atienza, M., Egiazarian, K., and Cantero, J. (2008). Measuring directional coupling between EEG sources. *NeuroImage*, 43:497–508.

Grosse-Wentrup, M. (2009). Understanding brain connectivity patterns during motor imagery for brain-computer interfacing. In *Advances in Neural Information Processing Systems 21 (NIPS 2008)*, pages 561–568.

Grosse-Wentrup, M. (2011). What are the causes of performance variation in brain-computer interfacing? *International Journal of Bioelectromagnetism*, 13(3):115–116.

Haufe, S., Tomioka, R., Nolte, G., Müller, K.-R., and Kawanabe, M. (2010). Modeling sparse connectivity between underlying brain sources for EEG/MEG. *IEEE Transactions on Biomedical Engineering*, 57(8):1954–1963.

Hirayama, J. and Hyvärinen, A. (2012). Structural equations and divisive normalization for energy-dependent component analysis. In *Advances in Neural Information Processing Systems 24 (NIPS2011)*, pages 1872–1880.

Hirayama, J., Ogawa, T., and Hyvärinen, A. (2014). Simultaneous blind separation and clustering of coactivated EEG/MEG sources for analyzing spontaneous brain activity. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2014)*, pages 4932–4935.

Hironaga, N. and Ioannides, A. A. (2007). Localization of individual area neuronal activity. *NeuroImage*, 34(4):1519–1534.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634.

Hyvärinen, A. and Hoyer, P. O. (2000). Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12(7):1705–1720.

Hyvärinen, A., Hoyer, P. O., and Inki, M. (2001a). Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001b). *Independent Component Analysis*. John Wiley & Sons.

Hyvärinen, A., Ramkumar, P., Parkkonen, L., and Hari, R. (2010a). Independent component analysis of short-time Fourier transforms for spontaneous EEG/MEG analysis. *NeuroImage*, 49:257–271.

Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. (2010b). Estimation of a structural vector autoregression model using non-Gaussianity. *Journal of Machine Learning Research*, 11:1709–1731.

Karklin, Y. and Lewicki, M. S. (2005). A hierarchical Bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, 17:397–423.

Kawanabe, M. and Müller, K.-R. (2005). Estimating functions for blind separation when sources have variance dependencies. *Journal of Machine Learning Research*, 6:453–482.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62(1):49–66.

Köster, U. and Hyvärinen, A. (2010). A two-layer model of natural stimuli estimated with score matching. *Neural Computation*, 22:2308–2333.

Lance, B. J., Kerick, S. E., Ries, A. J., Oie, K. S., and McDowell, K. (2012). Brain-computer interface technologies in the coming decades. *Proceedings of the IEEE*, 100(Centennial-Issue):1585–1599.

Leonardi, N. et al. (2013). Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *NeuroImage*, 83:937–950.

Liu, X., Chang, C., and Duyn, J. H. (2013). Decomposition of spontaneous brain activity into distinct fMRI co-activation patterns. *Frontiers in Systems Neuroscience*, 7:101.

Mahot, M. et al. (2013). Asymptotic properties of robust complex covariance matrix estimates. *IEEE Transactions on Signal Processing*, 61(13):3348–3356.

Makeig, S. et al. (2012). Evolving signal processing for brain-computer interfaces. *Proceedings of the IEEE*, 100(Centennial-Issue):1567–1584.

Ollila, E. and Koivunen, V. (2003). Robust antenna array processing using M-estimators of pseudo-covariance. In *Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC2003)*, pages 7–10.

Onton, J. and Makeig, S. (2009). High-frequency broadband modulations of electroencephalographic spectra. *Frontiers in Neuroscience*, 159:99–120.

Osindero, S., Welling, M., and Hinton, G. E. (2006). Topographic product models applied to natural scene statistics. *Neural Computation*, 18:381–414.

Pfurtscheller, G. and Neuper, C. (1997). Motor imagery activates primary sensorimotor area in humans. *Neuroscience Letters*, 239:65–68.

Ramkumar, P., Parkkonen, L., Hari, R., and Hyvärinen, A. (2012). Characterization of neuromagnetic brain rhythms over time scales of minutes using spatial independent component analysis. *Human Brain Mapping.*, 33(7):1648–1662.

Ramkumar, P., Parkkonen, L., and Hyvärinen, A. (2014). Group-level spatial independent component analysis of Fourier envelopes of resting-state MEG data. *NeuroImage*, 86:480–491.

Salakhutdinov, R., Roweis, S. T., and Ghahramani, Z. (2003). Optimization with EM and expectation-conjugate-gradient. In *Proceedings of the 20th International Conference on Machine Learning (ICML2003)*, pages 672–679.

Schreier, P. J. and Scharf, L. L. (2010). *Statistical Signal Processing of Complex-Valued Data*. Cambridge University Press.

Shi, L. C. and Lu, B. L. (2008). Dynamic clustering for vigilance analysis based on EEG. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2008)*, pages 54–57.

Smith, S. M. et al. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8):3131–3136.

Steele, R. J. and Raftery, A. E. (2010). Performance of Bayesian model selection criteria for Gaussian mixture models. In *Frontiers of Statistical Decision Making and Bayesian Analysis*, chapter 4.1, pages 113–130. Springer.

Valpola, H., Harva, M., and Karhunen, J. (2004). Hierarchical models of variance sources. *Signal Processing*, 84(2):267–282.

Vinh, N. X., Epps, J., and Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11:2837–2854.

Zander, T. O. and Kothe, C. (2011). Towards passive braincomputer interfaces: applying braincomputer interface technology to humanmachine systems in general. *Journal of Neural Engineering*, 8(2).

Zhang, K. and Hyvärinen, A. (2010). Source separation and higher-order causal analysis of MEG and EEG. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI 2010)*, pages 709–716.