NeuroImage 118 (2015) 598–612

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Solving the EEG inverse problem based on space–time–frequency structured sparsity constraints

Sebastián Castaño-Candamil [a,c,*], Johannes Höhne [b], Juan-David Martínez-Vargas [c], Xing-Wei An [d], German Castellanos-Domínguez [c], Stefan Haufe [e,f,**]

[a] BrainLinks-BrainTools, Albert-Ludwig Universität Freiburg, Germany
[b] Neurotechnology Group, Technische Universität Berlin, Germany
[c] Signal Processing and Recognition Group, Universidad Nacional de Colombia, Colombia
[d] Department of Biomedical Engineering, Tianjin University, PR China
[e] Laboratory for Intelligent Imaging and Neural Computing, Columbia University in the City of New York, USA
[f] Machine Learning Group, Technische Universität Berlin, Germany

ARTICLE INFO

Article history:
Received 28 December 2014
Accepted 7 May 2015
Available online 3 June 2015

Keywords:
EEG
MEG
Inverse problem
Spatio-temporal priors
Structured sparsity
Non-stationarity

ABSTRACT

We introduce STOUT (spatio-temporal unifying tomography), a novel method for the source analysis of electroencephalograpic (EEG) recordings, which is based on a physiologically-motivated source representation. Our method assumes that only a small number of brain sources are active throughout a measurement, where each of the sources exhibits focal (smooth but localized) characteristics in space, time and frequency. This structure is enforced through an expansion of the source current density into appropriate spatio-temporal basis functions in combination with sparsity constraints. This approach combines the main strengths of two existing methods, namely Sparse Basis Field Expansions (Haufe et al., 2011) and Time–Frequency Mixed-Norm Estimates (Gramfort et al., 2013). By adjusting the ratio between two regularization terms, STOUT is capable of trading temporal for spatial reconstruction accuracy and vice versa, depending on the requirements of specific analyses and the provided data. Due to allowing for non-stationary source activations, STOUT is particularly suited for the localization of event-related potentials (ERP) and other evoked brain activity. We demonstrate its performance on simulated ERP data for varying signal-to-noise ratios and numbers of active sources. Our analysis of the generators of visual and auditory evoked N200 potentials reveals that the most active sources originate in the temporal and occipital lobes, in line with the literature on sensory processing.

Introduction

In recent years, the advances in Neuroscience have led to a better understanding of cognitive processes in the human brain. One general goal is to identify brain areas related to certain cognitive processes or pathologies through measurements. Methods allowing for such kind of analyses are called neuroimaging techniques.

At the same time, it is also of importance to identify and characterize temporal brain activation patterns related to the cognitive phenomena under study. Magneto- and electroencephalography (MEG and EEG) have been widely used to study brain dynamics by identifying and analyzing temporal activation patterns, e.g., neural rhythms, event-related potentials (ERP), epileptic spikes, among others (e. g., Michel et al., 2004; Galka et al., 2004; Blankertz et al., 2011). M/EEG recordings also contain spatial information, because they are usually measured over the entire scalp using up to a few hundred sensors. Consequently, EEG and MEG have also been used as neuroimaging techniques (e. g., Zwoliński et al., 2010; Baillet et al., 2001; Toga and Mazziotta, 2002). While most of the considerations made here equally apply to MEG, we restrict the discussion to EEG in the following.

The pyramidal neurons believed to account for most of the EEG signal populate the entire cortical gray matter, and outnumber the available sensors by several orders of magnitude. Methods for estimating the generators of EEG activity therefore need to consider at least a few thousand potentially contributing brain sites as potential sources, which may be distributed evenly across the brain, or restricted to the cortical gray matter. Estimating the source distribution of brain electrical activity based on EEG measurements therefore amounts to solving

* Correspondence to: S. Castaño-Candamil, BrainLinks-BrainTools, Albert-Ludwig Universität Freiburg, Germany.
** Correspondence to: S. Haufe, Laboratory for Intelligent Imaging and Neural Computing, Columbia University in the City of New York, USA.
E-mail addresses: sebastian.castano@blbt.uni-freiburg.de (S. Castaño-Candamil), sh3397@columbia.edu (S. Haufe).

http://dx.doi.org/10.1016/j.neuroimage.2015.05.052
1053-8119/

an ill-posed and mathematically underdetermined inverse problem, where a unique solution can only be obtained by making additional assumptions (Baillet et al., 2001; Galka et al., 2004; Grech et al., 2008; Babadi et al., 2014). This can for example be done by way means of introducing prior beliefs on the structure of possible source configurations in a Bayesian inference framework (Nummenmaa et al., 2007).

With respect to neurophysiological plausibility, it has been argued that solutions with a simple spatial structure may be favored. There are various algorithmic approaches to enforce simplicity. The minimum norm estimate (MNE, Hämäläinen and Ilmoniemi, 1994), for example, minimizes the overall power of the sources, whereas the Low Resolution Tomography estimate (LORETA, Pascual-Marqui et al., 1994) explicitly enforces spatial smoothness of the sources based on the argument that neighboring voxels should be similarly active. Technically, both approaches can be implemented using ℓ_2-norm penalties. On the other hand, it has also been argued that, in event-related experimental designs, only a small fraction of the brain should be consistently activated. Consequently, methods assuming sparsity in the spatial domain have been proposed (Gorodnitsky et al., 1995; Matsuura and Okabe, 1995; Grech et al., 2008; Bolstad et al., 2009; Wipf and Nagarajan, 2009; Ou et al., 2008). Sparse methods are often based on the minimization of ℓ_1-norm regularization terms or, in a more general sense, on the minimization of the volume spanned by the active coefficients of the sources.

While being physiologically motivated, all these solutions practically suffer from undesired properties, which include spatial blurring—and the resulting inability to spatially separate multiple sources—, the presence of so-called ghost sources for minimum ℓ_2-norm solutions, as well as instability and spatial scattering for minimum ℓ_1-norm solutions (Haufe et al., 2008b, 2011; Grech et al., 2008; Tibshirani, 1994). To overcome these issues, several authors have proposed to combine spatial smoothness and sparsity to obtain *focal* source activations, be it through a combination of penalty terms (see Haufe et al., 2008a,b; Vega-Hernández et al., 2008), or through representing brain activity as the sum of a small number of spatial basis functions describing smooth localized patches of potentially active brain regions (Friston et al., 2008; Haufe et al., 2008a, 2011).

Besides enforcing a preferred spatial structure, prior information may also be included in the form of temporal constraints describing dynamics of neural activity. Specifically, it has been shown that time-frequency representations provide insightful information about the dynamics of neural processes (Miwakeichi et al., 2004; Durka et al., 2005; Trujillo-Barreto et al., 2008; Gramfort et al., 2013). Generally, brain activity may be non-stationary (e. g., event-related), which is not taken into account by classical methods. In contrast, Gramfort et al. (2013) address the non-stationarity issue by representing brain activity through a sparse set of time-frequency basis functions (atoms).

The vast majority of inverse methods for neuroimaging employ constraints either in the spatial or temporal domain, but not simultaneously in both domains. Thus, some methods are able to accurately describe non-stationary brain activations (e. g., Gramfort et al., 2013, TF-MxNE), but their solutions may be too focal; that is, solutions are not composed of smooth activation patches, but of non-contiguous spikes of activation. The opposite holds for other methods (e. g., Haufe et al., 2011, S-FLEX) that enforce spatial focality while being unable to describe non-stationary brain activations. Here, we propose to fill this gap by enforcing neurophysiologically motivated structure both in time and space, and thereby to unify the advantages of S-FLEX and TF-MxNE. Precisely, we propose a spatio-temporal decomposition of source activations, which depends on three components: (1) a predefined dictionary of spatial basis fields, (2) a predefined dictionary of temporal basis functions, and (3) a matrix of spatio-temporal coefficients that needs to be estimated. By adopting spatial and temporal "dictionaries" from Gramfort et al. (2013) and Haufe et al. (2011), our method — termed spatio-temporal unifying tomography (STOUT) — is able to reconstruct the time courses of potentially non-stationary source activations with focal spatial topographies. Moreover, by enforcing sparse structure through a weighted combination of spatial and temporal penalty terms, our method is able to "trade" spatial focality for a simpler time-frequency representation, and vice versa. This tradeoff is quantified by a single hyperparameter that allows to access to an entire spectrum of solutions ranging between S-FLEX and TF-MxNE.

The present manuscript is organized as follows. In the Methods section, we give an introduction to the EEG inverse problem and present existing solutions as well as our novel source imaging method STOUT. In the Experiments and Results sections, we assess the reconstruction of simulated ERP activity using STOUT as compared to state-of-the-art source imaging approaches. We also apply STOUT to real EEG data, where the task is to localize the generators of auditory and visual evoked potentials recorded during an oddball experiment. Then, we discuss the properties of our method in the Discussion section, and conclude our contributions in the Conclusion section.

Methods

EEG forward and inverse problem

The electromagnetic field measured by EEG may be represented by the following linear model (Baillet et al., 2001; Grech et al., 2008):

$$\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{J} + \epsilon. \tag{1}$$

Here, $\boldsymbol{Y} \in \mathbb{R}^{N_c \times N_t}$ is the EEG data measured at a set of N_c sensors at N_t time points, $\boldsymbol{J} \in \mathbb{R}^{3N_d \times N_t}$ (termed the *current density*) is the corresponding brain source activity matrix holding the 3D current vectors of N_d dipolar electrical brain sources at the N_t time points, and $\boldsymbol{L} \in \mathbb{R}^{N_c \times 3N_d}$ (the *lead field*) is a gain matrix representing the relationship between the current sources $\boldsymbol{J}$ and the measured EEG data $\boldsymbol{Y}$, composed as $\boldsymbol{L} = [\boldsymbol{L}_x, \boldsymbol{L}_y, \boldsymbol{L}_z]$, where the matrices $\boldsymbol{L}_{x/y/z}$ are the lead fields of the current sources in each direction x, y and z, respectively. We also assume that $\boldsymbol{Y}$ is affected by Gaussian distributed noise $\epsilon \in \mathbb{R}^{N_c \times N_t}$ with covariance $cov(\epsilon) = \boldsymbol{Q}_\epsilon \in \mathbb{R}^{N_c \times N_c}$, where $\boldsymbol{Q}_\epsilon$ is the noise covariance matrix. In practice, $\boldsymbol{Q}_\epsilon$ can be estimated from data using baseline measurements (Nagarajan et al., 2007), be derived from the lead field (assuming i.i.d. source activations), or simply set to the identity matrix. The latter approach is applied in the present work. Under this model, the maximum a-posteriori (MAP) estimate of $\boldsymbol{J}$ can be found as the minimizer of the following cost function, which is composed of a quadratic error term and a regularization term (Grech et al., 2008):

$$\underset{\boldsymbol{J}}{\operatorname{argmin}} \left\{ ||\boldsymbol{Y} - \boldsymbol{L}\boldsymbol{J}||^2_{\boldsymbol{Q}_\epsilon} + \lambda\Theta(\boldsymbol{J}) \right\}. \tag{2}$$

Here, $\|\boldsymbol{P}\|_{\boldsymbol{Q}_\epsilon} = \sqrt{\operatorname{tr}\left\{\boldsymbol{P}^T \boldsymbol{Q}_\epsilon^{-1} \boldsymbol{P}\right\}}$ denotes the Mahalanobis distance, $\lambda \in \mathbb{R}^+$ is a regularization constant, and $\Theta(\boldsymbol{J}) \in \mathbb{R}^+$ is a function which formalizes the constraints that are imposed upon the source activity.

Existing inverse solutions

The penalty function $\Theta(\boldsymbol{J})$ is commonly used to promote solutions with a certain spatial or temporal structure. Solution with purely smooth as well as purely sparse source activations have been argued to be neurophysiologically plausible (Hämäläinen and Ilmoniemi, 1994; Pascual-Marqui et al., 1994; Gorodnitsky et al., 1995; Matsuura and Okabe, 1995). An example of a spatially smooth method is the Low Resolution Tomography (LORETA) estimate (Pascual-Marqui