◆ Human Brain Mapping 30:1898–1910 (2009) ◆

EEG Source Imaging With Spatio-Temporal Tomographic Nonnegative Independent Component Analysis

Pedro A. Valdés-Sosa,[1]* Mayrim Vega-Hernández,[1] José Miguel Sánchez-Bornot,[1] Eduardo Martínez-Montes,[1] and María Antonieta Bobes[2]

[1]*Cuban Neuroscience Center, Neurostatistics Department, Cubanacán, Playa, Havana, Cuba*
[2]*Cuban Neuroscience Center, Cognitive Neurosciences Department, Cubanacán, Playa, Havana, Cuba*

Abstract: This article describes a spatio-temporal EEG/MEG source imaging (ESI) that extracts a parsimonious set of "atoms" or components, each the outer product of both a spatial and a temporal signature. The sources estimated are localized as smooth, minimally overlapping patches of cortical activation that are obtained by constraining spatial signatures to be nonnegative (NN), orthogonal, sparse, and smooth-in effect integrating ESI with NN-ICA. This constitutes a generalization of work by this group on the use of multiple penalties for ESI. A multiplicative update algorithm is derived being stable, fast and converging within seconds near the optimal solution. This procedure, spatio-temporal tomographic NN ICA (STTONNICA), is equally able to recover superficial or deep sources without additional weighting constraints as tested with simulations. STTONNICA analysis of ERPs to familiar and unfamiliar faces yields an occipital-fusiform atom activated by all faces and a more frontal atom that only is active with familiar faces. The temporal signatures are at present unconstrained but can be required to be smooth, complex, or following a multivariate autoregressive model. *Hum Brain Mapp 30:1898–1910, 2009.* © 2009 Wiley-Liss, Inc.

Key words: EEG; MEG; inverse solution; orthogonal; semi-negative matrix factorization; spatio-temporal source imaging; ICA

INTRODUCTION

An electrophysiological source image (ESI) is an estimate of those current source distributions in the brain that generate observed electrical (EEG) and magnetic (MEG) signals. The estimation procedure for ESI is known as the EEG/MEG Inverse Problem which is, unfortunately, severely ill posed and underdetermined [Michel et al., 2004]. As a consequence, any given observed data can be equally well explained by an infinite number of different ESI. However, a unique ESI may be obtained by selecting a given source configuration that not only explains the data but also satisfies additional constraints or prior information.

Over the past few years, several different choices of priors have been proposed for explaining the EEG/MEG at a given time instant (instantaneous ESI) in a history that starts with dipolar sources and has been extended with different types of distributed inverse solutions-to a great degree paralleling developments in statistical regression analysis and atomic decomposition methods in signal processing. This proliferation of ESI models immediately begs the question: Which is the best prior to use? Summarizing a recent trend our group has emphasized on the need of using ESI models satisfying multiple priors

*Correspondence to: Pedro A. Valdés-Sosa, Ave. 25, esq. 158, #15202, Cubanacán, Playa, Havana, Cuba.
E-mail: peter@cneuro.edu.cu

Received for publication 11 December 2008; Revised 16 February 2009; Accepted 25 February 2009

DOI: 10.1002/hbm.20784

Published online 17 April 2009 in Wiley InterScience (www.interscience.wiley.com).

simultaneously. The relative weight of each type of prior will then be determined by a data driven procedure. This strategy of different types of prior constraints to fit data will be termed multiple-prior electrophysiological source imaging and has been recently summarized in [Vega-Hernández et al., 2008]. There is an additional vast literature on Bayesian models for instantaneous source localization which we cannot summarize here completely. For excellent reviews see [Nummemnaa et al., 2007a,b; Wipf and Nagarajan, 2009]. It is not our purpose to review this literature exhaustively but rather to extend this family of generative models.

Instantaneous ESI only considers spatial constraints on sources. By contrast, [Scherg and Voncramon, 1986] pointed out early on, that major electrophysiological components could be modeled by a fixed set of neural generators with time varying amplitudes. This assumption is the basis of spatio-temporal ESI. There have been some attempts to apply temporal penalizations to the sources of distributed solutions [Baillet and Garnero, 1997; Galka et al., 2004; Yamashita et al., 2004]. However, a major problem with most of these methods is the sheer volume of data to be analyzed and visualized.

To overcome this data overload, some authors have tried first to find a meaningful decomposition of the space/time data into a small number of components or atoms (each with a spatial and temporal signature) and afterwards carry out ESI of the voltage atoms extracted. We shall call this multistep approach tomography of components. Methods used for tomography of component have ranged from those applied with Principal Component Analysis (PCA) [Koles et al., 1995; Marzetti et al., 2008], Independent Component Analysis (ICA) [Onton et al., 2006; Tang et al., 2002], Multilinear decompositions [Miwakeichi et al., 2004], to others based on projection on temporal bases [Ou et al., 2009; Zumer et al., 2008]. In fact it might be rewarding to apply source localization to other types of EEG decompositions. For example, in [Parra et al., 2005] bilinear decompositions are described that extract spatio-temporal atoms without restrictive spatial or anatomical assumptions.

We consider that these methods are statistically suboptimal and advocate that extraction of relevant atoms and ESI should be carried out simultaneously, a procedure that we shall call component tomography. To our knowledge this type of decomposition (that explicitly acknowledges the existence of a lead field) has only been carried out for PCA in source space (sPCA) as described by [Marzetti et al., 2008]. Unfortunately sPCA for sources, the same as PCA for the EEG suffers from rotational indeterminacy. This has led us to look for a tomographic ICA which, in our opinion, would benefit from the all advantages of this type of decomposition. ICA is a rapidly evolving subject which initially was identified with very specific methods such as gradient descent on a mutual information objective function that minimizes higher order dependency between sources. However, in their seminal paper [Roweis and Ghahramani, 1999] spelled out [Eqs. (23)–(30)] the general generative model underlying ICA which can even be extended to cope with temporal constraints such as in [Parra and Spence, 2000]. This approach has been followed by other authors [Friston, 2008]. In this context the EEG ICA problem can be stated as the minimization of the following objective function:

$$||\mathbf{V} - \mathbf{MG}||_2^2 + f(\mathbf{M})$$

The EEG data matrix $\mathbf{V}$ (channels by time points) is decomposed into the outer product of k atoms. Here the k spatial atoms are arranged by columns in the matrix $\mathbf{M}$, and the k temporal atoms are arranged by rows in the matrix $\mathbf{G} \cdot f(\mathbf{M})$ is a nonlinear penalty (corresponding to a prior) that enforces some type of component independence. Carrying out source localization of the k columns of $\mathbf{M}$ corresponds to what we call tomography of ICA components. Rather than do this we propose to minimize the following function:

$$||\mathbf{V} - \mathbf{KMG}||_2^2 + f(\mathbf{M})$$

where $\mathbf{K}$ is the lead field thus obtaining a tomographic ICA. Different types of ICA and tomographic ICA can be specified by selecting the nonlinear penalty function $f(\mathbf{M})$. There has been a recent flurry of interest in nonnegative ICA, initially described by Plumbley and Oja [Plumbley and Oja, 2004]. In this case the orthogonality usually associated with PCA is augmented with non negativity.

This article describes a new type of spatio-temporal ESI that extracts a unique and parsimonious set of source "atoms" such that the sources estimated are localized as smooth, minimally overlapping patches of cortical activation. Technically this is achieved by constraining spatial signatures to be nonnegative (NN), orthogonal, sparse and smooth. The method builds on recent results on nonnegative and semi nonnegative matrix factorizations [Ding et al., 2006b; Yoo and Choi, 2008] to date only applied to directly observed data. Here, we generalize these techniques to find components not in data space but rather in source space-in effect integrating ESI with NN-ICA. We shall therefore refer to this method as spatio-temporal tomographic NN ICA or STTONNICA. We present examples that show that tomographic ICA gives better results than tomography on ICA.

The article is organized as follows: In the next section, a summary of instantaneous ESI methods and their associated penalties lays the groundwork for the STTONNICA procedure. The next section specifies the generative model for the new method. Additionally, a multiplicative update algorithm is derived which is stable, fast, and converges within seconds near the optimal solution. The next two sections describe results with simulated and experimental data. In particular it is shown that STTONNICA analysis of ERPs to familiar and unfamiliar faces yields an

TABLE I. Names and references of methods using penalties based on minimizing different norms of the vector of parameters (Nonsmooth) or of the smoothed vector of parameters (Smooth)

	Nonsmooth			Smooth	
Norm	L0	L1	L2	L1	L2
Regression	Entropy Penalty, Subset selection (Chen, et al., 2001)	Lasso (Tibshirani, 1996)	Ridge (Hoerl and Kennard, 2000)	Lasso Fusion (Land and Fiedman, 1996)	Weighted ridge (Hämäläinen, 1993)
Signal Processing	Best Orthogonal Basis (Chen, et al., 2001)	Basis Pursuit (Chen, et al., 2001)	Methods of frames (Chen, et al., 2001)	—	—
ESI	Dipole (Scherg and Voncramon, 1986) MUSIC (Mosher and Leahy, 1998)	Minimum current estimates (Uutela et al., 1999)	Minimum norm (Hämäläinen and Ilmoniemi, 1994)	Vareta (Bosch-Bayard et al., 2001) Loreta-L1 (Silvaa et al., 2004) Focuss (Gorodnitsky et al., 1995)	Loreta (Pascual-Marqui et al., 1994) sLoreta (Pascual-Marqui, 2002)

occipital-fusiform atom activated by all faces and a more frontal atom that is only active with familiar faces as previously surmised. The discussion outlines possible extensions of the new technique.

PRIORS FOR INSTANTANEOUS ESI

We first restrict our attention to ESI for images at a single time instant. Without loss of generality we shall also confine our attention to EEG ESI with sources restricted within, and oriented perpendicularly to the cortical surface. The ESI forward problem is:

$$\mathbf{v}_t = \mathbf{K}\mathbf{g}_t + \varepsilon_t \tag{1}$$

where $\mathbf{v}_t$ is the vector of N_e voltages measured at time t, $\mathbf{K}$ is the lead field, ε_t is the sensor noise, and $\mathbf{g}_t$ the vector of N_s cortical sources usually with $N_s \gg N_e$. We must find estimates of $\mathbf{g}_t$. Instantaneous ESI impose spatial constraints on the currents $\mathbf{g}_t$. The corresponding ESI estimate can be stated as:

$$\hat{\mathbf{g}}_t = \arg\min_g \|\mathbf{v} - \mathbf{K}\mathbf{g}_t\|_2^2 + \lambda P(\mathbf{g}_t) \tag{2}$$

where $P(\mathbf{g}_t)$ is a penalization function and λ a regularization parameter. In most works $P(\mathbf{g}_t) = \|\mathbf{H}\mathbf{g}\|_p^2$ where the symbol $\|\cdot\|_p$ denotes the lp-norm. The functional (2) can also be interpreted as a penalized least squares regression or as a method for atomic decompositions in wavelet dictionaries. A variety of recent ESTI methods and their counterparts in statistics or signal analysis are shown in Table I.

When the penalty term is a quadratic function of $\mathbf{g}_t$ the estimators are explicit and linear. This type of problem is known in the linear regression field as ridge regression with different operators $\mathbf{H}$ [Hoerl and Kennard, 2000] that offer smooth solutions. In the past few years there have been advances in penalized least squares regression that highlight the use of non-convex (non quadratic) penalty functions, which can produce sparser or more concentrated solutions [Fan and Li, 2001; Tibshirani et al., 2005], especially useful for severely underdetermined problems like ESI as has been shown by several authors [Beucker and Schlitt, 1996; Matsuura and Okabe, 1995, 1997; Uutela et al., 1999]. It should be noted that some confusion arises due to the existence of several alternative algorithms that minimize the same objective function (2). For example some of the previously cited ESI methods are equivalent to FOCUSS) [Gorodnitsky et al., 1995; Gorodnitsky and Rao, 1997] and the Controlled Support MEG (CSMEG) [Nagarajan et al., 2006].

A more recent trend has been to use multiple priors simultaneously [Valdes-Sosa et al., 2006]. Specifically Eq. (2) generalizes to

$$\hat{\mathbf{g}}_t = \arg\min_{g_t} \|\mathbf{v}_t - \mathbf{K}\mathbf{g}_t\|_2^2 + \sum_{i=1}^{I} \lambda_i P_i(\mathbf{g}_t) \tag{3}$$

where $\sum \lambda_m P_m(\mathbf{g})$, with $m = 1, \ldots, M$, includes all prior information through the M penalty functions P_m. Each penalty term imposes a different type of constraint and has a corresponding regularization parameter λ_m which determines the relative importance given to each aspect of prior specification. These parameters are selected by a data driven procedure. In practice the Bayesian Information Criterion (BIC) is used for this purpose since it provides a measure that trades off data fitting and model complexity.

Equation (3) expresses the multiple penalized least squares model described in [Vega-Hernández et al., 2008]. Table II sets out some recent attempts to use this type of penalization in the ESI, statistical and signal processing

TABLE II. Names, equations and references of methods using multiple penalties

Name	Equation	Reference
Fused lasso	$\hat{\mathbf{g}}_{\text{FusedL}} = \arg\min \|\mathbf{v}-\mathbf{Kg}\|_2^2 + \lambda_{\text{sparse}_1}\|\mathbf{g}\|_1 + \lambda_{\text{sparse}_2}\|\mathbf{Dg}\|_1$	(Tibshirani et al., 2005)
Fused ridge	$\hat{\mathbf{g}}_{\text{RFused}} = \arg\min_{\mathbf{g}} \|\mathbf{v}-\mathbf{Kg}\|_2^2 + \lambda_{\text{smooth}_1}\|\mathbf{g}\|_2^2 + \lambda_{\text{smooth}_2}\|\mathbf{Lg}\|_2^2$	(Vega-Hernández et al., 2008)
Elastic net	$\hat{\mathbf{g}}_{\text{ENET}} = \arg\min_{\mathbf{g}} \|\mathbf{v}-\mathbf{Kg}\|_2^2 + \lambda_{\text{sparse}}\|\mathbf{g}\|_1 + \lambda_{\text{smooth}}\|\mathbf{g}\|_2^2$	(Zou and Hastie, 2005)
Elastic Net-L	$\hat{\mathbf{g}}_{\text{ENET-L}} = \arg\min_{\mathbf{g}} \|\mathbf{v}-\mathbf{Kg}\|_2^2 + \lambda_{\text{sparse}}\|\mathbf{Lg}\|_1 + \lambda_{\text{smooth}}\|\mathbf{Lg}\|_2^2$	(Vega-Hernández et al., 2008)
Smooth lasso	$\hat{\mathbf{g}}_{\text{S-LASSO}} = \arg\min_{\mathbf{g}} \|\mathbf{v}-\mathbf{Kg}\|_2^2 + \lambda_{\text{sparse}}\|\mathbf{g}\|_1 + \lambda_{\text{smooth}}\|\mathbf{Dg}\|_2^2$	(Hebiri, 2008)
Focal vector field reconstruction	$\hat{\mathbf{g}}_{\text{FVR}} = \arg\min_{\mathbf{g}} \|\mathbf{v}-\mathbf{Kg}\|_2^2 + \lambda_{\text{sparse}_1}\|\mathbf{g}\|_1 + \lambda_{\text{sparse}_2}\|\mathbf{Lg}\|_1$	(Haufe et al., 2008)
Controlled support MEG	$\hat{\mathbf{g}}_{\text{CSMEG}} = \arg\min_{\mathbf{g}} \|\mathbf{v}-\mathbf{Kg}\|_2^2 + \lambda_{\text{sparse}}\|\mathbf{g}\|_0 + (1-\lambda_{\text{sparse}})\|g\|_1$	(Nagarajan et al., 2006)

D is gradient (first differences), **L** is Laplacian operator.

literature. Vega et al. [Vega-Hernández et al., 2008] described a family of models (Elastic Net-L) that combine LORETA (Smooth L2 norm) and VARETA (smooth L1 norm) with a continuous transition of one type of ESI to the other:

$$\hat{\mathbf{g}}_{t\text{ENET-L}} = \arg\min_{\mathbf{g}_t} \|\mathbf{v}-\mathbf{Kg}_t\|_2^2 + \lambda_{\text{sparse}}\|\mathbf{Lg}_t\|_1 + \lambda_{\text{smooth}}\|\mathbf{Lg}_t\|_2^2 \quad (4)$$

It is precisely this blend of smoothness and sparseness that will be applied to the identification of functionally significant components in the next section.

SPATIO-TEMPORAL COMPONENT MODELS

The spatio-temporal model with k components is:

$$\mathbf{v}_t = \sum_{i=1}^{k} \mathbf{K}\mathbf{m}_i\mathbf{g}_{i,t}^T + \varepsilon_t \quad (5)$$

where the superscript T denotes vector transpose.

Here, a relatively small number, k, of components or atoms are used to fit the observed data. For the i-th atom its distribution on the cortex (spatial signature) is given by $\mathbf{m}_i$ and its temporal signature is given by $\mathbf{g}_{i,t}$. As before, N_e is the number of electrodes, N_t the number of time instants, N_s the number of cortical sources. $\mathbf{K}_{(N_e\times N_s)}$, is the lead field. Arranging the data, spatial signatures, and temporal signature in the matrices $\mathbf{V}_{(N_e\times N_t)} = [\mathbf{v}_{1(N_e\times 1)}\cdots\mathbf{v}_{N_t(N_e\times 1)}]$, $\mathbf{M}_{(N_s\times k)} = [\mathbf{m}_{1(N_s\times 1)}\cdots\mathbf{m}_{k(N_s\times 1)}]$, and $\mathbf{G}_{(k\times N_t)} = [\mathbf{g}_{1(N_t\times 1)}\cdots\mathbf{g}_{k(N_t\times 1)}]^T$ summarizes the generative model as.

$$\mathbf{V}_{(N_e\times N_t)} = \mathbf{K}_{(N_e\times N_s)}\mathbf{M}_{(\mathbf{N_s}\times\mathbf{k})}\mathbf{G}_{(k\times N_t)} + \mathbf{E}_{(N_e\times N_t)} \quad (6)$$

The multiple penalty objective function (3) for instantaneous ESI is adapted to Spatio-Temporal modeling as:

$$\|\mathbf{V}-\mathbf{KMG}\|_2^2 + \sum_{i=1}^{I}\lambda_i P_i(\mathbf{M}) \quad (7)$$

Although in Eq. (7) other penalty functions on **G** may be also included, a refinement postponed for future work. For instance, in the specific case of the Elastic Net-L model (hereinafter ENET), Eq. (5), takes the following form

$$\|\mathbf{V}-\mathbf{KSMG}\|_2^2 + \lambda_{\text{sparse}}\|\mathbf{M}\|_1 \quad (8)$$

where $\mathbf{S}_{(N_s\times N_s)}$ is a Smoothing Kernel defined as $\mathbf{S} = (I + \lambda_{\text{smooth}}\mathbf{L}^T\mathbf{L})^{-1}$.

STTONNICA

Unfortunately sparseness and smoothness do not guarantee uniqueness of the model since $\mathbf{MG} = \mathbf{MAA}^{-1}\mathbf{G}$. Uniqueness does follow by requiring that spatial signatures be nonnegative $\mathbf{M} \geq 0$, and that they be column-orthogonal ($\mathbf{M}^T\mathbf{M} = \mathbf{I}$, with $\mathbf{I}$ the identity matrix). The important paper [Ding et al., 2006a] demonstrates this uniqueness. These authors also point out that this form of matrix factorization is equivalent to k-means clustering since in each row only one element can be nonnegative thus specifying a cluster label. **Our specification of priors thus stipulates that we are searching for smoothly varying non-overlapping patches of cortical activation, with most of the cortex inactive.**

It should be mentioned that the particular form objective function (6) is based on the non-smooth nonnegative factorization algorithm. [Pascual-Montano et al., 2006].

Estimation of **M** and **G** proceeds as follows:

1. Initialize **M** with random positive numbers. Note that this initialization is NOT an orthogonal matrix-the algorithm quickly reaches this property.
2. Estimate **G** by ordinary least squares: $\hat{\mathbf{G}} = (\mathbf{X}^T\mathbf{X})^+\mathbf{X}^T\mathbf{V}$, where $\mathbf{X} = \mathbf{KSM}$ and $\mathbf{A}^+$ is the Moore-Penrose pseudo inverse of **A**.

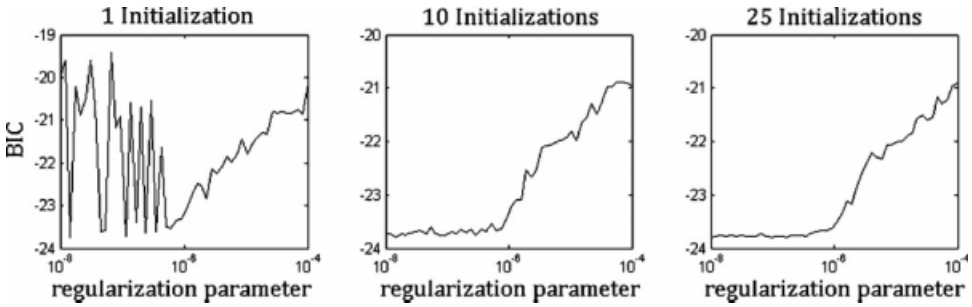

Figure 1.

BIC curves obtained as the minimum across a number of initializations for each value of the regularization parameter. Note that the noisy behavior of BCI for a small number of initializations is considerably reduced when this number is increased. The BCI curve for 100 initializations (not shown) closely resembles the one for 25 initializations. So, we decided to use this number for all subsequent runs of STTONNICA.

3. Estimate $\mathbf{M}$ through the multiplicative update rule: $\mathbf{M} = \mathbf{M}.* \sqrt{(\nabla_N./(\nabla_P + \tau))}$s, where .* and ./ respectively denote the dot-product and dot-division (element by element), τ is a small constant to avoid division by zero, and ∇_N, ∇_P are specified below.
4. Alternate between steps 2 and 3 until $\mathbf{M}$ converges.

The expressions needed for step 3 are:

$$\nabla_P = [(\mathbf{KS})^T\mathbf{V}\mathbf{G}^T]^- + [(\mathbf{KS})^T\mathbf{KSMG}\mathbf{G}^T]^+ + [\mathbf{M}\mathbf{G}^T\mathbf{KSM}]^+ + \lambda_{\text{sparse}}\mathbf{1}_{N_s}\mathbf{1}_k^T + \mu\mathbf{M}(\mathbf{M}^T\mathbf{M} - \mathbf{I})$$

$$\nabla_N = [(\mathbf{KS})^T\mathbf{V}\mathbf{G}^T]^+ + [(\mathbf{KS})^T\mathbf{KSMG}\mathbf{G}^T]^- + [\mathbf{M}\mathbf{G}^T\mathbf{KSM}]^- + [\mathbf{MG}\mathbf{G}^T\mathbf{M}^T(\mathbf{KS})^T\mathbf{KSM}]^+ + \lambda_{\text{sparse}}\mathbf{M}\mathbf{1}_{N_s}\mathbf{1}_k^T\mathbf{M}$$

where $[\mathbf{A}]^-$ and $[\mathbf{A}]^+$ represent the negative and positive values of the matrix A. The derivation of this multiplicative update rule is given in Appendix A.

This algorithm is carried out for a range of values of λ_{smooth} and λ_{sparse}. To compare models with different number of components and values of the regularization parameters the Bayesian Information Criterion (BIC) is evaluated as BIC $= \|\mathbf{V} - \mathbf{KS}\hat{\mathbf{M}}\hat{\mathbf{G}}\|_2^2 + df\cdot\log(n)/n$, where $\hat{\mathbf{M}}$ and $\hat{\mathbf{G}}$ are the estimators of the spatial and temporal signatures. Also *df* denotes the degrees of freedom of the model and *n* the number of data points.

STTONNICA, as all nonnegative factorization models is sensitive to initial conditions. In practice, it is common to apply the algorithm several times with different random initializations. AS shown in Figure 1, the BIC curve improves with the number of initializations. So, in the following, we set the algorithm to use 25 initializations for a more robust and reliable estimation.

RESULTS WITH SIMULATED DATA

A simulated data set was prepared with a small subset of voxels in the cortical surface of the MNI brain atlas [Trujillo-Barreto et al., 2004]. A total of 736 voxels were selected on an axial slice and 32 electrodes were located around the head at the same height (see Fig. 4A). White noise was added to the electrodes with a resulting power signal to noise ratio (SNR) of 14 dB.[1] The lead field for this configuration was computed using a three-sphere piecewise homogeneous and isotropic model for the head. The primary current density (PCD) can then be considered to be a function defined on a line interval (with the endpoints joined), in which neighboring elements are also neighbors in the brain space, as they are placed on the cortical surface. With this lead field iterative estimation of the STTONNICA model for a given set of regularization parameters took 1.2 s on an Intel Core 2 Duo, 2.4 GHz, 8 GB RAM desktop.

As an illustration of the performance of the method, two different scenarios were simulated for the spatial distribution of the PCD. In the first one, two Gaussian sources of width = 10 voxels were constructed, whose maxima were located at points with different depth relative to the sensors. Figure 2 (upper left) shows a linear representation of these sources, which are shown in left panels of Figure 4B,C as superimposed on the axial slice of the head. The corresponding time courses were simulated as perturbed rhythmic activity with different frequencies and amplitude (Fig. 2, upper right).

In the second scenario, we simulated three different Gaussian sources with the same depth, but different width (the first with width = 10 voxels and the other two with width = 30 voxels) as shown in Figure 3 top left and left panels in Figure 4D–F. The time course of the first source was simulated as a typical ERP, whereas the other two were the same oscillations used in the first simulated scenario (Fig. 3, upper-right panel). In both cases the data were generated from the average referenced lead field with the addition of a noise component corresponding to the SNR defined earlier.

[1] $\text{SNR} = 10\log_{10}\frac{\text{power signal}}{\text{power noise}}$.

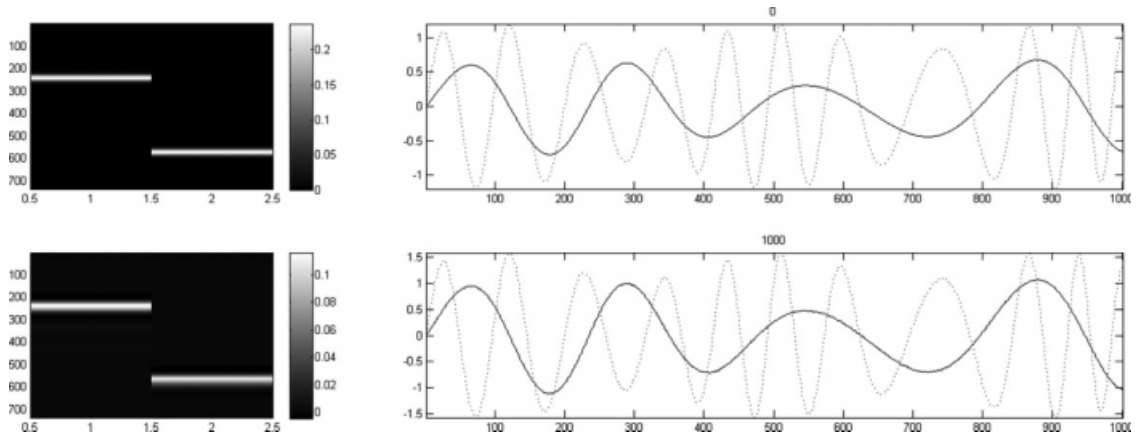

Figure 2.

STTONNICA recovery of two extended sources at different depths. The top row shows the spatial signature (left) and temporal signatures (right) of two simulated EEG atoms. Voltages were generated by using a lead field with a simplified EEG direct problem as shown in Figure 3A with the addition of noises in the sensors. In this simulation, one atom represents a deeper source and the other a more superficial one. The results of source imaging with STTONNICA are shown on the bottom with good recovery of the simulated spatial (left) and temporal (right) signatures.

For both types of simulations the STTONNICA ESI procedure was applied as described earlier. Because the sources lie on a line, a one-dimensional Laplacian $\mathbf{L} = \text{toeplitz}([2 - 1, \text{zeros}(N_s - 2)])$ was defined as a roughness penalty, from which the smoothing matrix was then obtained by inversion. The regularization parameters were found by minimizing the BIC criterion, a process that was feasible due to the speed of the computations. In any case, the reconstruction in both scenarios was excellent, even despite of the low signal to noise ratio, as can be seen comparing the top and bottom parts of Figures 2 and 3, as well as the right and left panels in Figures 4B–F.

To compare the performance of STTONNICA as a component tomography with that of a more common tomography of components, we used the data of the second simulated scenario described earlier. First, we decompose the data in three independent components with an ICA algorithm as described in [Onton et al., 2006]. Second, the

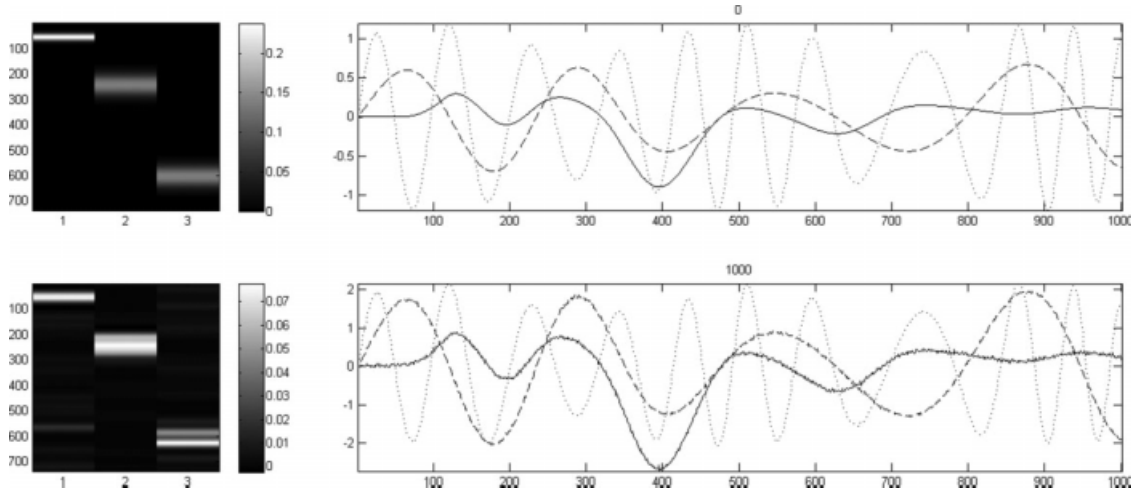

Figure 3.

STTONNICA recovery of three extended sources with varying widths. The top row shows the spatial signature (left) and temporal signatures (right) of three simulated EEG atoms generated as in Figure 1, in this case designed to test quality of recovery with sources with different spatial extensions. The results of source imaging with STTONNICA are shown on the bottom with good recovery of the simulated spatial (left) and temporal (right) signatures.

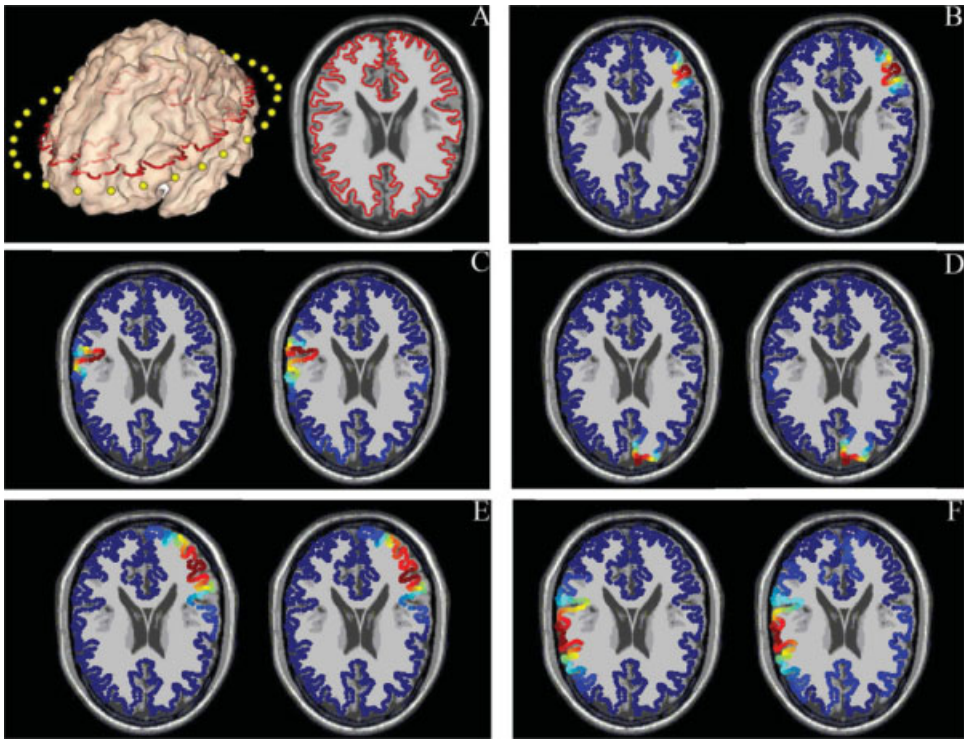

Figure 4.

Simulation tests of STTONNICA recovery of EEG atoms. Simulations were carried out with a simplified direct problem. Both cortical sources and electrodes were restricted to a single axial slice of the brain as shown in A. On the left is a three-dimensional rendering of the cortical surface. The yellow circles indicate electrode positions and the red line mark the voxels to which active sources were restricted. These voxels are marked also on the axial slice on the right. For the pairs of images denoted as B–F, on the left shows the simulated cortical source and the right side the spatial signature recovered by STTONNICA Figure 3B,C show the results for the two source simulation that corresponds to Figure 1, consisting of one superficial and one deep source. Figure 3D–F shows the corresponding results for sources at totally different locations. The quality of the reconstruction of the sources is evident.

topographical components were subjected to source localization using LORETA and ENET. Figure 5 shows the spatial maps obtained together with the true source and the already reported STTONICA solution. As can be seen tomographic ICA produces less phantom sources and has less crosstalk than tomography of ICA components.

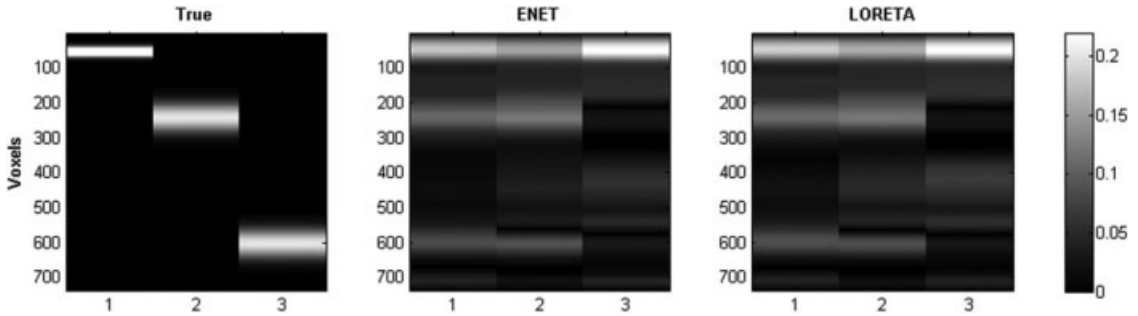

Figure 5.

Component tomography versus tomography of components. The left panel shows the three "true" simulated sources (second scenario). The middle and right panel shows the absolute value of ENET and LORETA solutions respectively, obtained from the three ICA components. All solutions were normalized for the easiness of comparison. These figures can be compared to the STTONICA solution shown in Figure 3.

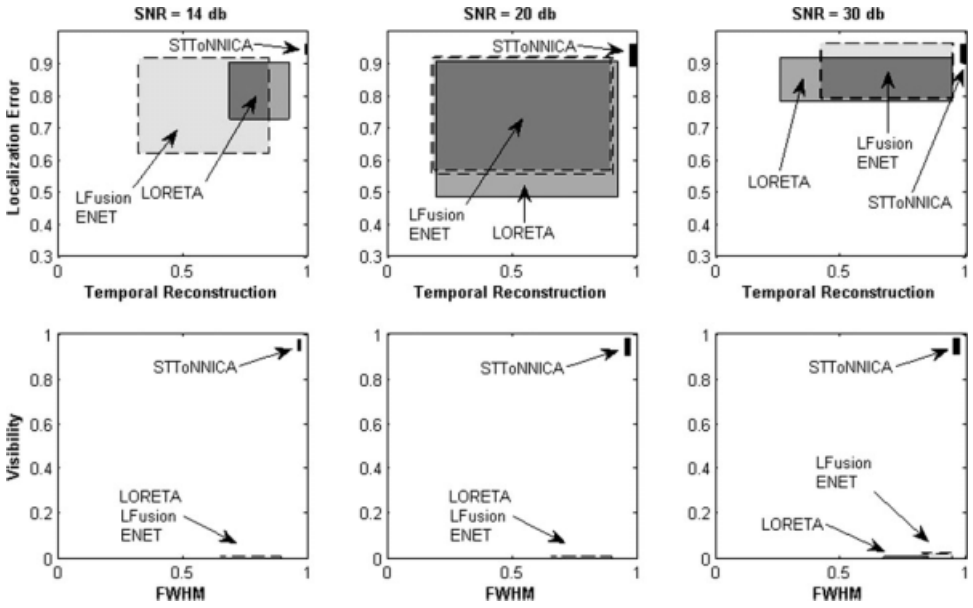

Figure 6.

Spatio-temporal components vs. instantaneous source localization. Comparison between STTONNICA, LORETA, ENET, and LFUSION according to measures of the quality of the reconstruction. Upper row: Localization error versus temporal reconstruction. Lower row: Visibility versus full width at half maximum. Each method is shown as a rectangle covering the percentiles 25 and 75 in both axes. STTONNICA appears in black and the other methods in gray tones.

A comparison can also be made between STTONNICA and other instantaneous ESI methods such as LORETA, ENET, and LFUSION. For this purpose, we prepared 20 different two-sources configurations, based on the first scenario described earlier. The first source was a narrow Gaussian with five different random localization of the maximum, while the second source was a wider Gaussian whose maximum varied in four different random positions. For each simulated data, three different values of SNR were used (14 db, 20 db, and 30 db) making a total of 60 simulations. LORETA, ENET, LFUSION, and STTONICA were applied to each of the simulated data.

Different quality measures were computed for assessing the goodness of reconstruction. These were the localization error, full width at half maximum (FWHM), visibility and temporal reconstruction. The first three are detailed in [Vega-Hernández et al., 2008], while the fourth refers to the correlation between estimated and true temporal waveforms. STTONNICA was required to estimate two components. However, the other three instantaneous methods were applied separately to each time point of the spatio-temporal data matrix. A representative spatial map was obtained for the time point in which the sum of the two simulated temporal courses is maximum in magnitude. Thus, this spatial maps shows simultaneously both sources, but we compare them separately with the corresponding true spatial map.

Figure 6 shows plots summarizing this comparison. The bi-dimensional plots correspond to the normalized quality measures (0 for the worst recovery, 1 for ideal 100% correspondence between estimated and true sources). Each method is represented by a rectangle covering from percentile 25–75 in both dimensions [Vega-Hernández et al., 2008].

SOURCE ANALYSIS OF ERPS FROM A FACE FAMILIARITY EXPERIMENT

As an illustration of the method using real data, ERPs were obtained from an oddball paradigm in which frequent unfamiliar faces (UNFAMILIAR) were used as standard stimuli (83%) and the infrequent target stimuli were face of acquaintances (FAMILIAR) (17%), which were selected among the family and close friends of each subject (i.e., this was different for each subject). Face stimuli were presented for 1,000 ms, with an inter-stimulus interval (ISI) of 1,000 ms. Data acquisition was carried out with 120 monopolar derivations, using electrodes mounted in an elastic cap homogeneously distributed over the scalp (subsample of the Electrocap 128 montage) referred to linked earlobes. Signals were amplified by a factor of 10,000 and filtered between 0.5–30 Hz (3 dB down, and a notch filter with peak at 60 Hz was used). The EEG was

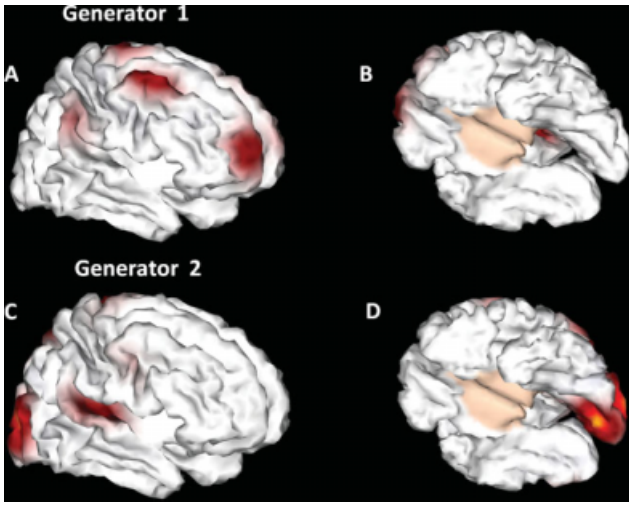

Figure 7.

Spatial Signatures of STTONICA atoms of ERPS to Familiar versus unfamiliar faces. Intensity coded map of the magnitude of the spatial signature of the two STONNICA atoms obtained from ERPS to familiar versus unfamiliar faces. On the left are views of the right hemisphere. On the left are views of the base of the brain with the frontal pole towards the left. Note the more frontal localization of Generator 1 and the more occipital one of Generator 2 with activation of the right fusiform area.

digitally recorded at a sampling rate of 200 Hz, with 700 ms epochs, and a 300 ms pre-stimulus baseline. Averaged evoked responses for each subject were digitally low-pass filtered (5.5 Hz cut-off) and DC corrected. A detailed description of the paradigm and recording parameters can be found in [Bobes et al., 2007]. As described in that paper both unfamiliar and acquaintance faces elicited a N170 component associated to face presentation. Additionally faces of acquaintance elicited a P3 component with two subcomponents, an early P3 with latency at about 350 ms and larger over frontal sites and another with latency at about 500 ms, larger over central-parietal sites. For analysis with STONNICA the grand average of 10 subjects was obtained. Before obtaining the grand average, individual recordings were visually inspected and electrodes with excessive noise were eliminated and substituted by an interpolation of the 11 closest neighbors.

To finesse the issue of registration of different brains a representative head and brain was used as a template. The head used was the standard high resolution COLIN head from which the lead field was obtained with a three concentric sphere model as implemented in the BRAINCRACKER software (NEURONIC SA). The cortical surface (white matter/gray matter frontier) was approximated with 5656 triangles. The vertices of the triangles were taken as the locations of the potential sources generators and their orientation as the average of the normals of all triangles sharing a vertex. A surface Laplacian was also obtained from the cortical surface for use as a roughness penalty. Both data and lead field were transformed to an average reference montage prior to STTONNICA computations. For this situation the estimation of a STTONNICA solution for a given set of regularization parameters took 4.2 s. Of course several sets of parameters had to be explored and such a search took about 15 min.

The ERP data for both conditions (FAMILIAR and UNFAMILIAR) were concatenated along the time axis. This was done to estimate common spatial signatures for both conditions. It was expected that the variation of temporal signatures would indicate the psycho-physiological relevance of the atoms extracted. The expectation was that a more frontal brain area involved with emotional processing would be sensitive to familiarity and a more occipital-temporal one with unspecific face recognition.

Two STTONNICA atoms were found to be optimal for describing this data. We shall designate these atoms as "Generators 1 and 2". Their spatial signatures are shown in Figure 7 and their temporal signatures are shown in Figure 8. It is interesting to note that Generator 1 has a more frontal distribution. Its time course is enhanced for

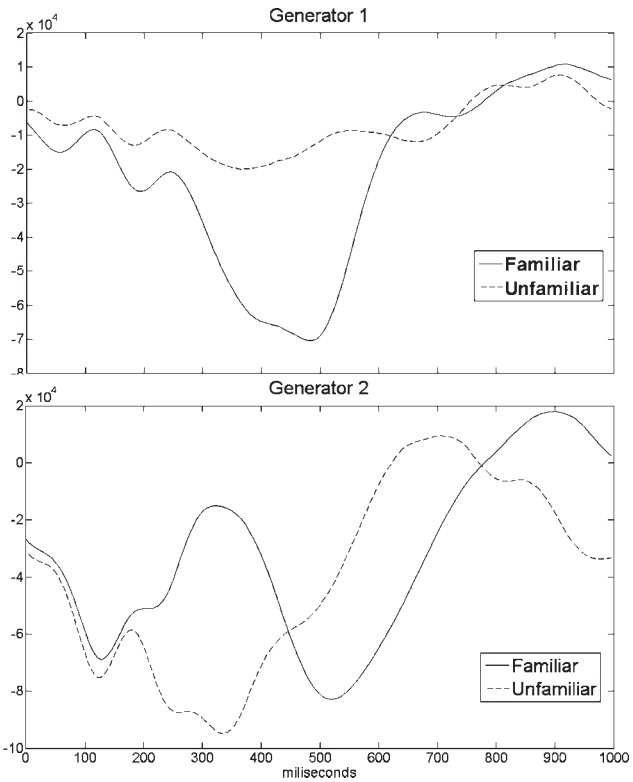

Figure 8.

Temporal Signatures of STTONICA atoms of ERPS to Familiar versus unfamiliar faces. The temporal signature for Generator 1 shows major deflections for only familiar faces. That of Generator 2 shows great variability for both familiar and unfamiliar faces.

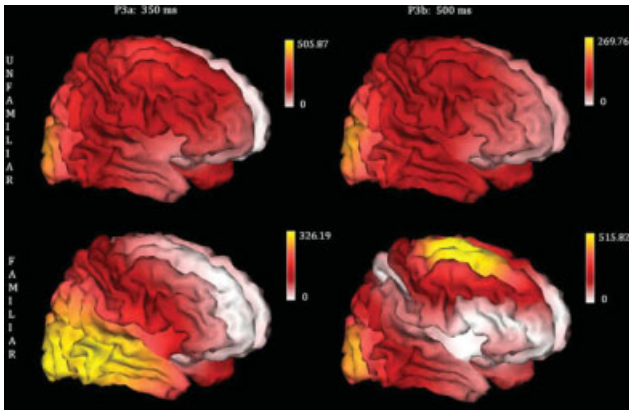

Figure 9.
Absolute value of ENET tomographic solutions for the two P3 latencies (300 and 500 ms) and each of the experimental conditions (FAMILIAR and UNFAMILIAR). Note the much more localized activation obtained by STTONNICA than for the other methods.

familiar faces but not for unfamiliar faces. By contrast Generator 2 is present in the occipital pole and notably in the fusiform area. It is to be noted that the frontal atom is only active for familiar faces whereas the more occipital atom responds to both types of stimuli. Thus, these results are in good correspondence with prior expectations.

These results can be compared to the sources found by an instantaneous method using the traditional single-latency analysis. In this data, the four topographies corresponding to latencies 350 and 500 ms for each condition (FAMILIAR and UNFAMILIAR) were subjected to source localization using Elastic Net. Figure 9 shows the solutions obtained. It can be seen that in addition to the more blurred activations, it is also more difficult to interpret these solutions in terms of the sources related to the processing of familiar or unfamiliar faces, as can be easily done with STTONNICA.

DISCUSSION AND CONCLUSIONS

This article introduces for the first time a fusion of independent component analysis and electrophysiological source imaging (ESI) to address spatio-temporal models. Spatio-temporal tomographic nonnegative ICA (STTONNICA) estimates each component or "atom" as the outer product of a spatial and temporal signature. This separation results in physiologically more interpretable estimates. Moreover, imposing constraints of smoothness, sparsity, positivity and orthogonality on the spatial signatures results in estimates of cortical activation that have several desirable properties. For example, a major distinction with regard to previous ICA techniques is that these components are estimated in the source space instead of in the data space as is usual [Onton et al., 2006; Tang et al., 2002]. In fact the spatial constraints of nonnegativity and orthogonality translate into spatial clustering of active voxels providing an automatic functional parcellation of the cortex.

The identification of non overlapping sources has recently been a subject of some interest. A similar objective has been recently stated as the principal of minimally overlapping components or MOCA [Marzetti et al., 2008]. These authors carry out the following steps:

1. Find spatial topographical components with the method known as PISA. This is a frequency domain method that operates on the imaginary part of the crosspectral matrix.
2. A linear inverse solution is applied to the real and imaginary parts of the eigenvectors of the PISA components found by PCA or sPCA (source PCA).
3. The source space is rotated so that these tomographic components are both orthogonal and have minimal fourth order statistics in the spirit of ICA.

MOCA is similar to our approach in that it posits non overlapping sources but differs in the pair of requirements to achieve it: non negativity and orthogonality for STTONNICA; minimal fourth order statistics and orthogonality for MOCA). More importantly, STTONNICA carries out the estimation in a single step-thus the term "tomographic ICA". We have recently speculated with one of the authors of MOCA (G. Nolte) that PISA-MOCA might be achieved at one step using a similar approach to ours-thus carrying out tomographic ICA instead of tomography on ICA.

The objectives of [Daunizeau et al., 2006] are also very similar to ours in that they propose a Bayesian hierarchical model with the outer product of spatial locations modeled as clusters of voxels (spatial soft K-means found by Mixtures of Gaussians) and a temporal component. As pointed out by [Ding et al., 2006a,b] the semi nonnegative factorization is equivalent to both soft and hard K-means clustering. Further work is needed to compare the relative theoretical and computational advantages of both techniques.

A question that arises is whether the hypothesis of non overlapping sources is physiologically meaningful. We do consider it a useful approximation since the EEG (MEG) reflects the activity of large neural masses and successive activation of different brain sites is plausible and is the basis of all current spatio-temporal models. However, let us consider two scenarios in which this assumption may be challenged:

- If two subpopulations of the same brain site are active with different temporal dynamics neither STTONNICA nor any other current model based on Eq. (5) would be able to separate these sub-populations without some additional temporal a priori information. The model for STTONNICA can be extended to

include temporal constraints in a relatively straightforward manner as will be discussed later on.
- Care must be taken with the time window analyzed. Too long time windows may lead to confounding several temporal dynamics into one single spatial source. Too short time windows may lead to overseparation of spatial sources for a common dynamics. The correct length of segments could be decided by means of information criteria such as BIC proposed earlier.

Despite these caveats both the simulations and analysis of actual data presented in this article show that STTONNICA is a promising approach. STTONNICA is able to recover sources with different degree of sparseness and spatial smoothness, as well as with rhythmic and ERP-like temporal dynamics (Figures 2–4). It also shows better reconstructions (much less ghost sources and cross-talk) than the tomography of components obtained by LORETA and ENET after ICA decomposition (see Fig. 5). The exhaustive comparison with instantaneous approaches such as ENET, LORETA, and LFUSION showed that STTONNICA exhibited the best compromise with respect to measures of quality of the reconstruction (see Fig. 6). Finally, STTONNICA was also able to produce more interpretable and physiologically meaningful results when applied to real ERP data than the traditional single-latency approach using the instantaneous ESI method ENET (Figs. 7–9).

In general, one interesting feature of the proposed technique is that there is no need for depth weighting as necessary for most distributed solutions. This may be due to the fact that the resulting models are very constrained and therefore not underdetermined. The algorithm selected (multiplicative updates) is easy to implement, fast and quite flexible. It is easy to foresee tailored "prior carpentry" to incorporate more detailed anatomical ad physiological prior information into ESI based on this heuristic.

In the present work we have not imposed constraints on the temporal signatures. One immediate extension is requiring a smooth temporal time course. Preliminary results indicate that this requirement does not slow down the algorithm too much. However similar results might be obtained by preprocessing the data with adequate temporal filtering. A possibly more useful extension would be to posit some sort of multivariate autoregressive model for the temporal signatures. This would transform this model into a state space model and allow the evaluation of causal relations between the atoms. Work along these lines has already begun. Another possible extension would be the analysis of complex valued time series as result from time/frequency decompositions. Finally it should be noted that STTONNICA is a new type of generative model for Bayesian ESI. As such, the particular estimation procedure used in this article does not need to be its only embodiment. Indeed it is easy to envisage its use with other Bayesian estimation methods such as E-M, variational methods, Bayesian Model averaging or MCMC [Nummemnaa et al., 2007; Trujillo-Barreto et al., 2004].

APPENDIX A

Derivation of the Multiplicative Update Rule for Estimation of the Spatial Signatures

As set out in [Ding et al., 2006b; Yoo and Choi, 2008] the multiplicative updates for minimizing an objective function with respect to a nonnegative $\mathbf{M}$ can be stated as: $\mathbf{M} = \mathbf{M}.* \sqrt{(\nabla_{\mathrm{N}}./(\nabla_{\mathrm{P}} + \tau))}$s, where .* and ./ denotes the dot-product and dot-division respectively and $\bigtriangledown_{\mathrm{N}}$ and $\bigtriangledown_{\mathrm{P}}$ are the negative and positive parts of the gradient of the objective function. In our case the objective function is $\|\mathbf{V} - \mathbf{KSMG}\|_2^2 + \lambda_{\mathrm{sparse}}\|\mathbf{M}\|_1 = \|\mathbf{V} - \mathbf{KSMG}\|_2^2 + \lambda_{\mathrm{sparse}} \times \mathbf{1}_{N_s}^T\mathbf{M}\mathbf{1}_k$. The gradient function with respect to $\mathbf{M}$ is $\nabla_{\mathbf{M}} = (\mathbf{KS})^T(\mathbf{V} - \mathbf{KSMG})\mathbf{G}^T + \lambda_{\mathrm{sparse}}\mathbf{1}_{N_s}\mathbf{1}_k^T$. Where 1_{N_s} is a N_s-length column vector of ones. As we also impose orthogonality on $\mathbf{M}$ we wish to carry out updates only on the Stiefel Manifold of matrices with orthonormal columns [Yoo and Choi, 2008; Edelman A, Arias TA, Smith ST, 1998]. For this reason the correct gradient to use is $\tilde{\nabla}_{\mathbf{M}} = \tilde{\nabla}_{\mathbf{M}} - \mathbf{M}\tilde{\nabla}_{\mathbf{M}}^T\mathbf{M}$. Additionally, to avoid drifts off the desired manifold we add a lagrangian term to modify gradient to enforce the original constraint [Plumbley, 2005]: $\tilde{\nabla}_{\mathbf{M}} = \tilde{\nabla}_{\mathrm{M}} - \mathbf{M}\tilde{\nabla}_{\mathbf{M}}^T\mathbf{M} + \mu\mathbf{M}(\mathbf{M}^T\mathbf{M} - \mathbf{I})$. As a consequence the positive $\bigtriangledown_{\mathrm{P}}$ and negative $\bigtriangledown_{\mathrm{N}}$ parts of the modified gradient are

$$\nabla_P = [(\mathbf{KS})^T\mathbf{V}\mathbf{G}^T]^- + [(\mathbf{KS})^T\mathbf{KSMGG}^T]^+ + [\mathbf{MG}^T\mathbf{KSM}]^+ + \lambda_{\mathrm{sparse}}\mathbf{1}_{N_s}\mathbf{1}_k^T + \mu\mathbf{M}(\mathbf{M}^T\mathbf{M} - \mathbf{I})$$

$$\nabla_N = [(\mathbf{KS})^T\mathbf{V}\mathbf{G}^T]^+ + [(\mathbf{KS})^T\mathbf{KSMGG}^T]^- + [\mathbf{MG}^T\mathbf{KSM}]^- + [\mathbf{MGG}^T\mathbf{M}^T(\mathbf{KS})^T\mathbf{KSM}]^+ + \lambda_{\mathrm{sparse}}\mathbf{M}\mathbf{1}_{N_g}\mathbf{1}_k^T\mathbf{M}$$

where $[\mathbf{A}^-]$ and $[\mathbf{A}^+]$ represent the negative and positive values of the matrix A.

ACKNOWLEDGMENTS

The authors thank Eduardo Aubert Vazquez for help with image processing and Javier Varona for help with the preparation of this manuscript. Discussions with G. Nolte were very helpful in clarifying the relation of our approach to that of MOCA.

REFERENCES

Baillet S, Garnero L (1997): A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. IEEE Trans Biomed Eng 44:374–385.

Beucker R, Schlitt HA (1996): On minimal lp-norm solutions of the biomagnetic inverse problem. Tech Rep KFA-ZAM-IB-9614 7:1–13.

Bobes MA, Quiñonez I, Perez J, Leon I, Valdés-Sosa M (2007): Brain potentials reflect access to visual and emotional memories for faces. Biol Psychol 75:146–153.

Bosch-Bayard J, Valdés-Sosa P, Virues-Alba T, Aubert-Vázquez E, Roy John E, Harmony T, Riera-Díaz J, Trujillo-Barreto NJ

(2001): 3D statistical parametric mapping of variable resolution electromagnetic tomography (VARETA). Clin Electroencephalogr 32:47–66.

Chen SS, Donoho DL, Saunders MA (2001): Atomic decomposition by basis pusuit. SIAM Rev 43:129–159.

Daunizeau J, Mattout J, Clonda D, Goulard B, Benali H, Lina JM (2006): Bayesian spatio-temporal approach for EEG source reconstruction: Conciliating ECD and distributed models. IEEE Trans Biomed Eng 53:503–516.

Ding C, Li T, Jordan MI (2006a): Convex and semi-nonnegative matrix factorizations. Available at: http://www.cs.berkeley.edu/∼jordan/papers/ding-li-jordan.pdf.

Ding C, Li T, Peng W, Park H (2006b): Orthogonal nonnegative matrix t-factorizations for clustering. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Philadelphia, PA, USA.

Edelman A, Arias TA, Smith ST (1998): The geometry of algorithms with orthogonality constraints. Siam Journal on Matrix Analysis and Applications 20:303–353.

Fan JQ, Li RZ (2001): Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 96:1348–1360.

Friston K (2008): Hierarchical models in the brain. Plos Comput Biol 4:e1000211.

Galka A, Yamashita O, Ozaki T, Biscay R, Valdes-Sosa P (2004): A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. Neuroimage 23:435–453.

Gorodnitsky IF, Rao BD (1997): Sparse signal reconstruction from limited data using FOCUSS: Are-weighted minimum norm algorithm. IEEE Trans Signal Process 45:600–616.

Gorodnitsky IF, George JS, Rao BD (1995): Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm. Electroencephalogr Clin Neurophysiol 95:231–251.

Hämäläinen MS (1993): Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys 65:413–497.

Hämäläinen MS, Ilmoniemi RJ (1994): Interpreting measured magnetic fields of the brain: estimates of current distributions. Med Biol Eng Comput 32:35–42.

Haufe S, Nikulin VV, Ziehe A, Müller KR, Nolte G (2008): Combining sparsity and rotational invariance in EEG/MEG source reconstruction. Neuroimage 42:726–738.

Hebiri M (2008): Regularization with the smooth-lasso procedure. ARXIV arXiv:0803.0668. http://prunel.ccsd.cnrs.fr/docs/00/33/09/11/PDF/Smooth_Lasso2.pdf

Hoerl AE, Kennard RW (2000): Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 42:80–86.

Koles ZJ, Lind JC, Soong ACK (1995): Spatiotemporal decomposition of the EEG—A general-approach to the isolation and localization of sources. Electroencephalogr Clin Neurophysiol 95:219–230.

Land S, Fiedman J (1996): Variable Fusion: A New Method of Adaptive Signal Regression. Stanford: Stanford University.

Marzetti L, Del Gratta C, Nolte G (2008): Understanding brain connectivity from EEG data by identifying systems composed of interacting sources. Neuroimage 42:87–98.

Matsuura K, Okabe Y (1995): Selective minimum-norm solution of the biomagnetic inverse problem. IEEE Trans Biomed Eng 42:608–615.

Matsuura K, Okabe Y (1997): A robust reconstruction of sparse biomagnetic sources. IEEE Trans Biomed Eng 44:720–726.

Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004): EEG source imaging. Clin Neurophysiol 115:2195–2222.

Miwakeichi F, Martinez-Montes E, Valdes-Sosa PA, Nishiyama N, Mizuhara H, Yamaguchia Y (2004): Decomposing EEG data into space-time-frequency components using parallel factor analysis. Neuroimage 22:1035–1045.

Mosher JC, Leahy RM (1998): Recursive MUSIC: A framework for EEG and MEG source localization. IEEE Trans Biomed Eng 45:1342–1354.

Nagarajan SS, Portniaguine O, Hwang D, Johnson C, Sekihara K (2006): Controlled support MEG imaging. Neuroimage 33:878–885.

Nummemnaa A, Auranen T, Hamalainen MS, Jaaskelainen IP, Lampinen J, Sams M, Vehtari A (2007a): Hierarchical Bayesian estimates of distributed MEG sources: Theoretical aspects and comparison of variational and MCMC methods. Neuroimage 35:669–685.

Nummenmaa A, Auranen T, Hamalainen MS, Jaaskelainen IP, Sams M, Vehtari A, Lampmen J (2007b): Automatic relevance determination based hierarchical Bayesian MEG inversion in practice. Neuroimage 37:876–889.

Onton J, Westerfield M, Townsend J, Makeig S (2006): Imaging human EEG dynamics using independent component analysis. Neurosci Biobehav Rev 30:808–822.

Ou WM, Hamalainen MS, Golland P (2009): A distributed spatio-temporal EEG/MEG inverse solver. Neuroimage 44:932–946.

Parra L, Spence C (2000): Convolutive blind separation of non-stationary sources. IEEE Trans Speech Audio Process 8:320–327.

Parra LC, Spence CD, Gerson AD, Sajda P (2005): Recipes for the linear analysis of EEG. Neuroimage 28:326–341.

Pascual-Marqui RD (2002): Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. Methods Find Exp Clin Pharmacol 24:5–12.

Pascual-Marqui RD, Michel CM, Lehmann D (1994): Low-resolution electromagnetic tomography—A new method for localizing electrical-activity in the brain. Int J Psychophysiol 18:49–65.

Pascual-Montano A, Carazo JM, Kochi K, Lehmann D, Pascual-Marqui RD (2006): Nonsmooth nonnegative matrix factorization (nsNMF). IEEE Trans Pattern Anal Machine Intell 28:403–415.

Plumbley MD (2005): Geometrical methods for non-negative ICA: Manifolds, lie groups and toral subalgebras. Neurocomputing 67:161–197.

Plumbley MD, Oja E (2004): A "nonnegative PCA" algorithm for independent component analysis. IEEE Trans Neural Netw 15:66–76.

Roweis S, Ghahramani Z (1999): A unifying review of linear gaussian models. Neural Comput 11:305–345.

Scherg M, Voncramon D (1986): Evoked dipole source potentials of the human auditory-cortex. Electroencephalogr Clin Neurophysiol 65:344–360.

Silvaa C, Maltezb JC, Trindadea E, Arriagac A, Ducla-Soares E (2004): Evaluation of L1 and L2 minimum norm performances on EEG localizations. Clin Neurophysiol 115:1657–1668.

Tang AC, Pearlmutter BA, Malaszenko NA, Phung DB, Reeb BC (2002): Independent components of magnetoencephalography: Localization. Neural Comput 14:1827–1858.

Tibshirani R (1996): Regression shrinkage and selection via the LASSO. J R Stat Soc B Met 58:267–288.

Tibshirani R, Saunders M, Rosset S, Zhu J, Knight RT (2005): Sparsity and smoothness via the fused LASSO. J R Stat Soc B 67:91–108.

Trujillo-Barreto NJ, Aubert-Vazquez E, Valdes-Sosa PA (2004): Bayesian model averaging in EEG/MEG imaging. Neuroimage 21:1300–1319.

Uutela K, Hämäläinen M, Somersalo E (1999): Visualization of magnetoencephalographic data using minimum current estimates. NeuroImage 10:173–180.

Valdes-Sosa PA, Sánchez-Bornot JM, Vega-Hernández M, Melie-García L, Lage-Castellanos A, Canales-Rodríguez E (2006): Granger causality on spatial manifolds: Applications to neuroimaging. In: Schelter B, Winterhalder M, Timmer J, editors. Handbook of Time Series Analysis. pp 461–485.

Vega-Hernández M, Martínez-Montes E, Sanchez-Bornot JM, Lage-Castellanos A, Valdes-Sosa PA (2008): Penalized least squares methods for solving the EEG inverse problem. Stat Sin 18:1535–1551.

Wipf D, Nagarajan S (2009): A unified Bayesian framework for MEG/EEG source imaging. Neuroimage 44:947–966.

Yamashita O, Galka A, Ozaki T, Biscay R, Valdes-Sosa P (2004): Recursive penalized least squares solution for dynamical inverse problems of EEG generation. Hum Brain Map 21:221–235.

Yoo C, Choi S (2008): Orthogonal nonnegative matrix factorization: multiplicative updates on stiefel manifolds. In: Fyfe C, Kim D, Lee S-Y, Hujun Yin H, editors. Intelligent Data Engineering and Automated Learning—IDEAL 2008, 9th International Conference. Daejeon, South Korea: Springer. pp 140–147.

Zou H, Hastie T (2005): Regularization and variable selection via the elastic net. J R Stat Soc B 67:301.

Zumer JM, Attias HT, Sekihara K, Nagarajan SS (2008): Probabilistic algorithms for MEG/EEG source reconstruction using temporal basis functions learned from data. Neuroimage 41:924–940.