



Published in final edited form as:

Neuroimage. 2009 July 15; 46(4): 1066–1081. doi:10.1016/j.neuroimage.2009.01.056.

Space-Time Event Sparse Penalization for Magneto-/Electroencephalography

Andrew Bolstad^{*,1}, Barry Van Veen, and Robert Nowak

Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53715 USA

Abstract

This article presents a new spatio-temporal method for M/EEG source reconstruction based on the assumption that only a small number of events, localized in space and/or time, are responsible for the measured signal. Each space-time event is represented using a basis function expansion which reflects the most relevant (or measurable) features of the signal. This model of neural activity leads naturally to a Bayesian likelihood function which balances the model fit to the data with the complexity of the model, where the complexity is related to the number of included events. A novel Expectation-Maximization algorithm which maximizes the likelihood function is presented. The new method is shown to be effective on several MEG simulations of neurological activity as well as data from a self-paced finger tapping experiment.

Keywords

M/EEG source localization; Spatio-temporal priors; M/EEG Inverse problem

1 Introduction

The ill-posed nature of the M/EEG inverse problem is widely recognized (Baillet et al., 2001), and unique solutions are usually obtained by placing prior expectations on the nature of the brain activity associated with the measured data. In this paper we present an approach for solving the inverse problem that assumes the underlying brain activity is due to a small number or sparse set of space-time events. Each space-time event corresponds to activity that is localized spatially and in a temporal subspace. Sets of spatial and temporal basis functions are employed to represent the activity due to each space-time event and are chosen to match the characteristics of expected source activity. Seeking to explain the measured data in terms of a small number of space-time events leads to a convex optimization problem, which we solve via a novel Expectation-Maximization (EM) algorithm.

* Corresponding author. akbolstad@wisc.edu (Andrew Bolstad). URL: homepages.cae.wisc.edu/~bolstad (Andrew Bolstad).

¹This work supported in part by the NIBIB under NIH award R21EB005473 and MIT Lincoln Laboratory under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

⁴In ML estimation, the M-Step is to maximize the Q-function alone; however, in the case of MAP estimation, the log likelihood of Θ is included as well (see Dempster et al. (1977)).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Common approaches to the M/EEG inverse problem may be classified as either parametric or imaging methods (Baillet et al., 2001). Parametric or scanning methods use a small number of dipoles (Scherg and von Cramon, 1985), multipoles (Mosher et al., 1999; Nolte and Curio, 2000), or cortical patches (Limpiti et al., 2006; Wagner et al., 2002, 2000) and scan over locations to find the best set of sources to represent the data. Examples of scanning methods include MUSIC (Mosher et al., 1992), beamforming (Van Veen et al., 1997), and maximum likelihood estimation (Dogandžić and Nehorai, 2000). One persistent challenge for scanning methods is choosing the number of sources in the solution.

The method proposed in this paper may be classified as an imaging technique. Imaging methods tessellate the brain using a large number of dipoles and attempt to simultaneously solve for all dipole amplitudes. This leads to the following linear model:

$$\mathbf{Y}=\mathbf{H}\mathbf{X}+\mathbf{N} \quad (1)$$

where \mathbf{Y} is an $M \times T$ matrix of spatio-temporal measurements from M sensors at T time points, \mathbf{H} is an $M \times 3D$ (or $M \times D$ in the case of known dipole moment orientation) lead field matrix relating sensor measurements to D discrete dipole locations (voxels) on the cortex, \mathbf{X} is the $3D \times T$ signal matrix representing the unknown time courses and orientations of each dipole on the cortical surface, and \mathbf{N} is a matrix of additive noise. The goal of imaging techniques is to estimate each element of \mathbf{X} simultaneously given \mathbf{Y} and \mathbf{H} . Typically M is on the order of several hundred, while D exceeds several thousand, so Eq. (1) is severely underdetermined. Furthermore, the effective rank of \mathbf{H} is often significantly less than M .

The problem of nonuniqueness with imaging techniques is often reduced by use of a Bayesian prior to favor some solutions over others. The prior is incorporated as a penalty term in a cost function which is minimized to estimate the neural activity. Under the assumption of white noise, these cost functions take the form:

$$\sum_{t=1}^T \|\mathbf{Y}_t - \mathbf{H}\mathbf{X}_t\|_2^2 + \lambda_t f(\mathbf{X}_t) \quad (2)$$

Equation (2) is expressed in terms of individual time samples since the optimization, in most previous work, is solved separately at each time point, and solutions are often found only at specific time points selected by the user (such as signal peaks). The first term of Eq. (2) ensures that solutions explain the data well, while f acts as a penalty term. The constants λ_t balances the contribution of both terms. A common choice of penalty is simply $f(\mathbf{X}_t)=\|\mathbf{W}\mathbf{X}_t\|_2^2$, which corresponds to a Gaussian prior on \mathbf{X}_t , and the minimum norm estimate (MNE) is obtained with $\mathbf{W} = \mathbf{I}$. MNE has a simple closed form solution, but tends to blur activity across the cortex. The well-known LORETA technique sets \mathbf{W} to an approximation of the Laplacian operator (Pascual-Marqui et al., 1994). Like the MNE, LORETA gives blurry, low resolution reconstructions of neuronal activity. In an effort to reduce the spatial blurring of the MNE and LORETA, the FOCUSS algorithm reinforces strong sources while deemphasizing weak ones by iteratively updating \mathbf{W} (Gorodnitsky and Rao, 1997). FOCUSS does produce more focal estimates, but is sensitive to the initialization of \mathbf{W} as well as to noise (Baillet et al., 2001).

Another set of methods aimed at producing focal solutions to the inverse problem use the ℓ_1 norm of \mathbf{X}_t as a regularizer. The corresponding solutions tend to be sparse, i.e., most entries in \mathbf{X} are zero, and can be found efficiently via linear programming. ℓ_1 penalties generally

correspond to an exponential prior on \mathbf{X}_p . Techniques differ in how to incorporate unknown dipolar orientation into the ℓ_1 penalty. The selective minimum norm (SMN) technique of Matsuura and Okabe (1995, 1997) penalizes the absolute sum of the components of each dipole. Later work by Uutela et al. (1999) uses the MNE to determine the orientation of dipolar sources, then solves the ℓ_1 penalized problem with these fixed orientations. A very recent approach for EEG called Sparse Source Imaging (SSI) replaces the ℓ_1 norm with the sum over all dipoles of the ℓ_2 norm of the three components of each dipole and solves the problem using second-order cone programming (Ding and He, 2007). In this way, SSI selects dipole orientation automatically based on the data. The criticism of ℓ_1 regularization is that it produces solutions which are “spiky” in space and inconsistent with physiological expectations. To remedy the “spiky” nature of minimum ℓ_1 solutions, another recent twist on ℓ_1 regularization in MEG, known as VESTAL, uses singular value decomposition of both the lead field matrix \mathbf{H} and the data matrix \mathbf{Y} (Huang et al., 2006). Here the dipole orientations are not fixed, but constrained to an octant via a diagonal sign matrix in the ℓ_1 penalty. Researchers have also proposed ℓ_p norm regularization with $0 < p < 1$ and $1 < p < 2$ as well (Auranen et al., 2005; Jeffs et al., 1987). The optimization problem is not convex for $0 < p < 1$. In general these optimization problems are more difficult to solve than $p = 1, 2$.

Temporal priors and sparsity have also been used to combat the ill-posed nature of the M/EEG inverse problem. Baillet and Garnero (1997) assume dipole magnitudes vary slowly with respect to sampling rate and thus penalize large changes between time samples. A similar approach is taken by Daunizeau et al. (2006) in which the second derivative of dipolar temporal evolution is penalized. Penalization of roughness encourages solutions that are low dimensional or sparse, that is, solutions implicitly described in terms of a small number of smooth bases. Dalal et al. (2008) present a parametric method for localizing cortical activity in time, frequency, and space. Localization in time and frequency implicitly assumes cortical activity is sparse in the time-frequency plane. A number of researchers have used wavelet temporal basis functions to explicitly find sparse representations of M/EEG activity. Efferen et al. (2000) present a method for representing event related potentials (ERPs) recorded via EEG (or invasive cortical sensors) using a sparse set of wavelet bases. Similarly, a wavelet shrinkage procedure is used to select a small number of coefficients representing M/EEG signals prior to a Variational Bayes inversion in Trujillo-Barreto et al. (2008).

The “event sparse penalty” (ESP) regularization we present here is closely related to ℓ_1 regularization. The ESP approach seeks a solution composed of a small number of “space-time events” (STEs). Each STE is a spatio-temporal signal defined in terms of a group of basis functions. Bases are a flexible and powerful tool for balancing prior expectations of source characteristics, such as focal or bandlimited source activity, against uncertainty in exact signal behavior, such as dipole orientation or temporal evolution. The STE is constrained to lie in the space spanned by the bases, but is unconstrained within this space. Example spatial bases include x-, y-, and z- components of a dipolar source, or multipoles and patch bases for extended sources. Example temporal basis functions include Fourier bases that represent specific frequency bands, wavelets that represent desired time-frequency characteristics, and singular value decomposition (SVD) derived bases which attempt to capture temporal characteristics of the signal from the data itself.

Use of spatial and temporal bases in the ESP regularizer effectively combines the spatial and temporal aspects of the source activity estimation problem. We present ESP regularization as well as a novel Expectation-Maximization (EM) algorithm for solving the resulting optimization problem. A simple criterion for guiding the selection of an appropriate regularization parameter λ is incorporated into the optimization procedure. The effectiveness of the method is illustrated on simulations as well as human subject data from a self-paced finger tapping experiment.

The remainder of the paper is organized as follows. In Materials and Methods, we first introduce the concept of spatio-temporal basis function expansions of \mathbf{X} in Eq. (3). This expansion allows us to define STEs and develop the ESP regularization technique for finding solutions which are sparse in STEs. Next, we present an EM algorithm for solving the ESP regularized least squares problem. The Results section describes several MEG simulations, including a set of randomized source simulations, and application of ESP regularization to MEG recordings of a subject performing self-paced finger tapping. We conclude with a discussion of the ESP method and possible extensions.

Matrices and vectors are represented by upper and lower case boldface letters, respectively. The notation \mathbf{e}_i represents the vector of all zeros except for a one in the i^{th} position. Matrix inversion is represented by superscript -1 , while transposition is represented by superscript T . Classic vector p norms are denoted $\|\cdot\|_p$, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. We denote the Kronecker product using the \otimes operator while $\text{vec}(\cdot)$ concatenates columns of a matrix to produce a column vector. For a review of algebra involving \otimes and $\text{vec}(\cdot)$, see Brewer (1978).

2 Materials and Methods

2.1 Event Sparse Penalty Regularization

The ESP approach to the M/EEG inverse problem relies on representing the cortical signal \mathbf{X} as a linear combination of physiologically meaningful basis functions from carefully crafted spatial and temporal dictionaries. In general, a spatio-temporal expansion of \mathbf{X} can be written:

$$\mathbf{X} = \mathbf{S}\mathbf{\Theta}\mathbf{T}^T \quad (3)$$

where the columns of \mathbf{S} are a dictionary containing spatial bases, the columns of \mathbf{T} are a dictionary containing temporal bases, and the $(k, l)^{\text{th}}$ element of coefficient matrix $\mathbf{\Theta}$ specifies the contribution associated with the k^{th} spatial basis modulated by the l^{th} temporal basis. If both \mathbf{S} and \mathbf{T} are invertible then there is a one-to-one mapping between $\mathbf{\Theta}$ and \mathbf{X} , that is $\mathbf{\Theta} = \mathbf{S}^{-1}\mathbf{X}\mathbf{T}^{-T}$. However, in general \mathbf{S} and \mathbf{T} are chosen to intentionally restrict the signal to lie in specific spatial and/or temporal subspaces in order to incorporate prior expectations into the solution of the M/EEG inverse problem, and thus are not invertible.

Given a spatio-temporal expansion in the form of Eq. (3), many methods can be used to solve the reformulated problem for $\mathbf{\Theta}$. In particular, regularized least squares problems analogous to Eq. (2) can be formulated with some function of $\mathbf{\Theta}$ as the regularization term. However, improved quality solutions are possible by designing the bases and the penalty term concurrently.

Our approach assumes that M/EEG signals are the result of a small number of space-time events (STEs) which occur over limited time spans and occupy contiguous, local areas of cortex. This assumption is reflected in the following spatio-temporal expansion:

$$\mathbf{X} = \mathbf{S}\mathbf{\Theta}\mathbf{T}^T = \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \dots & \mathbf{S}_I \end{bmatrix} \begin{bmatrix} \mathbf{\Theta}_{1,1} & \mathbf{\Theta}_{1,2} & \dots & \mathbf{\Theta}_{1,J} \\ \mathbf{\Theta}_{2,1} & \mathbf{\Theta}_{2,2} & \dots & \mathbf{\Theta}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Theta}_{I,1} & \mathbf{\Theta}_{I,2} & \dots & \mathbf{\Theta}_{I,J} \end{bmatrix} \begin{bmatrix} \mathbf{T}_1^T \\ \mathbf{T}_2^T \\ \vdots \\ \mathbf{T}_J^T \end{bmatrix} = \sum_{i=1}^I \sum_{j=1}^J \mathbf{S}_i \mathbf{\Theta}_{i,j} \mathbf{T}_j^T \quad (4)$$

Here the spatial and temporal dictionaries, as well as the coefficient matrix, have been partitioned into space-time events. The columns of each submatrix \mathbf{S}_i are an orthogonal basis for measurable spatial activity, such as that from a local region of cortex. Likewise, the columns of each \mathbf{T}_j are an orthogonal basis for a temporal subspace, such as a frequency band or time interval. Cortical activity that lies in the vector space defined by a single spatial submatrix \mathbf{S}_i and a single temporal submatrix \mathbf{T}_j defines a single STE and is represented by the matrix of coefficients Θ_{ij} . We seek parsimonious sets of spatial and temporal bases to describe candidate STEs and formulate a criterion for finding solutions that involve a small number of non-zero Θ_{ij} .

The choice of spatial and temporal bases in Eq. (4) allows users a great deal of flexibility in both signal model and assumptions about cortical activity. The temporal bases in \mathbf{T}_j may be chosen using wavelets, Gabor functions, Fourier bases, etc., to span time intervals and bandwidths consistent with the temporal characteristics of the signals of interest. Possible spatial bases \mathbf{S}_i for events that activate an extended area of cortex include cortical patches (Limpiti et al., 2006; Wagner et al., 2002) and multipolar expansions (Mosher et al., 1999; Nolte and Curio, 2000). The spatial bases are also selected consistent with the spatial attributes of interest. Dipolar source bases may be used in the ESP framework.

The motivation behind Event Sparse Penalty (ESP) regularization is the assumption that the activity of interest is a collection of a small number of space-time events. This assumption, combined with appropriate spatial and temporal dictionaries, implies that the true underlying signal should have $\Theta_{ij} = \mathbf{0}$ for the majority of pairs (i, j) . To encourage such a sparse solution, we seek to solve an optimization problem which balances measurement error with the number of nonzero blocks Θ_{ij} . One way to do this is to penalize the least squares problem based on the number of STEs. However, this results in a nonconvex problem that requires a computationally prohibitive exhaustive search over all possible combinations of active blocks.

Many recent results have suggested that relaxing nonconvex counting functions to the convex ℓ_1 norm in regularized least squares problems provides a very good approximation to the desired regularization and results in sparse solutions (Candes and Tao, 2005; Donoho, 2006; Tropp, 2006). In the spirit of these results, we propose the ℓ_1 block penalty regularized problem:

$$\min_{\Theta} \|\mathbf{Y} - \mathbf{H}\Theta\mathbf{T}^T\|_F^2 + \lambda \sum_{i,j} \|\text{vec}(\Theta_{i,j})\|_p \quad (5)$$

where $p \geq 1$ is chosen to capture the magnitude of the contribution of each block. Obvious choices are $p = 1, 2$, or ∞ . If $p = 1$, then the block structure of Eq. (4) is lost because the penalty in Eq. (5) becomes $\|\text{vec}(\Theta)\|_1$. This encourages a solution which is sparse in the number of nonzero coefficients in Θ , but not necessarily sparse in terms of the number of STEs. If $p = 2$ or $p = \infty$, then the penalty corresponds to the ℓ_2 or ℓ_∞ norm, respectively, of each block Θ_{ij} . This results in solutions that involve a small number of STEs, yet can accommodate uncertainty in the event characteristics by placing minimal restrictions on the coefficients in Θ_{ij} .

We prove in (Bolstad et al., 2007) that when one STE is present, the $p = 2$ case with lead field normalization (see Section 2.7) provides a higher probability of activating the correct STE than the $p = \infty$ case. Using $p = 2$ in Eq. (5) and noting that $\|\text{vec}(\Theta_{i,j})\|_2 = \|\Theta_{i,j}\|_F$ leads to estimate $\hat{\mathbf{X}} = \mathbf{S}\Theta^*\mathbf{T}^T$ where:

$$\hat{\Theta} = \arg \min_{\Theta} \|\mathbf{Y} - \mathbf{H}\mathbf{S}\Theta^T\|_F^2 + \lambda \sum_{i,j} \|\Theta_{i,j}\|_F \quad (6)$$

This objective function is convex since it is a sum of convex functions. At least one minimizer exists since the objective function is positive (i.e., bounded below) and convex for all Θ . Section 2.3 describes a novel EM algorithm for finding $\hat{\Theta}$. Next we discuss possible spatio-temporal dictionary constructions.

2.2 Spatio-Temporal Dictionaries for the Inverse Problem

The ESP approach is not limited to a specific set of space-time basis functions. Degenerate choices for the basis functions in Eq. (3) transform ESP regularization into existing regularization techniques. For example, ESP can reconstruct activity independently over all time points by setting each submatrix \mathbf{T}_j to the vector \mathbf{e}_j , $j = 1, 2, \dots, T$. Alternatively, reconstruction at a single time point t is obtained by setting $\mathbf{T} = \mathbf{e}_t$. Degenerate choices for the spatial dictionary \mathbf{S} combined with $\mathbf{T} = \mathbf{e}_t$ results in classical ℓ_2 or ℓ_1 regularization. If $\mathbf{S} =$

$$\hat{\Theta}_t = \arg \min_{\Theta_t} \|\mathbf{Y}_t - \mathbf{H}\Theta_t\|_2^2$$

$\mathbf{S}_1 = \mathbf{I}$ and $\mathbf{T} = \mathbf{e}_t$, the estimate given by Eq. (6) at time t becomes $\hat{\Theta}_t = \arg \min_{\Theta_t} \|\mathbf{Y}_t - \mathbf{H}\Theta_t\|_2^2 + \lambda \|\Theta_t\|_2$, which is equivalent to the MNE². Similarly, setting $\mathbf{S} = \mathbf{S}_1$ to an approximation of the surface Laplacian implements LORETA. If we assume dipole orientation is known (so that \mathbf{H} is $M \times D$) and use a spatial submatrix for each dipole on the cortical surface, i.e. $\mathbf{S}_i = \mathbf{e}_i$, the resulting estimate is equivalent to ℓ_1 regularization. If dipole orientation is unknown (now \mathbf{H} is $M \times 3D$) and \mathbf{S}_i is chosen as a $3D \times 3$ matrix with a 3×3 identity in the position corresponding to the components of the i^{th} dipole in \mathbf{H} , then ESP corresponds to the recently proposed SSI formulation of the EEG inverse problem (Ding and He, 2007)³. In this case ESP encourages a solution with a small number of dipoles while identifying the unknown dipole orientations.

The power of the ESP approach is that it allows for more general and flexible spatial and temporal representations of expected source activity than these degenerate examples. Temporal bases can be used, for example, to isolate a desired frequency band, or may be grouped into different frequency bands such that, for example, \mathbf{T}_1 spans the α -band, \mathbf{T}_2 the β -band, \mathbf{T}_3 the γ -band and so on, reflecting the expectation that different source activity occupies different bands. Temporal bases can also be used to isolate time ranges. Alternatively wavelets or other time-frequency bases can be used to identify source activity based on time-frequency characteristics analogous to Dalal et al. (2008). Similarly, the spatial bases can represent spatially extended source activity. The expectation that signals occupy extended areas rather than single dipoles is discussed, for example, in (Baillet and Garnero, 1997; Daunizeau and Friston, 2007). Representations for multipoles and cortical patches have been proposed to accommodate extended source activity (Limpiti et al., 2006; Mosher et al., 1999; Wagner et al., 2002, 2000) and are easily incorporated as bases, as we illustrate in Section 3.

2.3 Expectation-Maximization Algorithm

The gradient of $\sum_{i,j} \|\Theta_{i,j}\|_2$ is discontinuous so classical gradient-based minimization approaches to Eq. (6) can not be used without modification. Instead, we have developed an Expectation-Maximization (EM) algorithm which treats the objective of Eq. (6) as a negated

²It can be shown that the minimizer of $\|\mathbf{Y}_t - \mathbf{H}\mathbf{X}_t\|_2^2 + \lambda \|\mathbf{X}_t\|_2$ is also the minimizer of $\|\mathbf{Y}_t - \mathbf{H}\mathbf{X}_t\|_2^2 + \tilde{\lambda} \|\mathbf{X}_t\|_2^2$ for some $\tilde{\lambda}$.

³Although the problem formulation is the same, we solve the problem with a different algorithm.

Bayesian log-likelihood function where the penalty can be interpreted as the logarithm of a prior distribution on Θ . EM finds the maximum a posteriori (MAP) estimate of Θ . The approach described here is a generalization of the EM algorithm developed by Figueiredo and Nowak (2003). The pioneering work on EM was published by Dempster, Laird, and Rubin (Dempster et al., 1977).

The EM algorithm monotonically increases the Bayesian likelihood $p(\mathbf{Y}|\Theta)p(\Theta)$ over Θ .

Assuming white Gaussian noise and the density $p(\Theta) \propto \exp\left\{-\frac{\lambda}{2\sigma^2} \sum_{i,j} \|\Theta_{i,j}\|_F\right\}$, the negated

$$\frac{\|\mathbf{Y} - \mathbf{H}\Theta\mathbf{T}^T\|_F^2}{2\sigma^2}$$

Bayesian log-likelihood is proportional to $-\frac{\lambda}{2\sigma^2} \sum_{i,j} \|\Theta_{i,j}\|_F$. Multiplying through by $2\sigma^2$, we see that the EM algorithm is monotonically decreasing the function

$$\|\mathbf{Y} - \mathbf{H}\Theta\mathbf{T}^T\|_F^2 + \lambda \sum_{i,j} \|\Theta_{i,j}\|_F, \text{ which is the objective of Eq. (6).}$$

The key aspect of solving an estimation problem via the EM algorithm is to identify a so-called ‘‘hidden data set’’ which, if known, would make the problem easy to solve. EM alternates between solving the easy problem given an estimate of the hidden data and updating the estimate of the hidden data. We identify hidden data \mathbf{Z} for the ESP problem based on the decomposition:

$$\mathbf{Z} = \Theta + \alpha \mathbf{N}_1 \quad (7)$$

$$\mathbf{Y} = \mathbf{H}\mathbf{S}\mathbf{Z}\mathbf{T}^T + \mathbf{N}_2 \quad (8)$$

Here the noise matrix \mathbf{N} of Eq. (1) is decomposed into two independent random matrices \mathbf{N}_1 and \mathbf{N}_2 with distributions $\text{vec}(\mathbf{N}_1) \sim N(\mathbf{0}, \mathbf{I} \otimes \mathbf{I})$ and $\text{vec}(\mathbf{N}_2) \sim N(\mathbf{0}, \sigma^2 \mathbf{I} - \alpha^2 \mathbf{T}\mathbf{T}^T \otimes \mathbf{H}\mathbf{S}\mathbf{S}^T \mathbf{H}^T)$. The scalar α^2 must satisfy $\alpha^2 < \sigma^2 \|\mathbf{T}\mathbf{T}^T\|^{-1} \|\mathbf{H}\mathbf{S}\mathbf{S}^T \mathbf{H}^T\|^{-1}$ so that a valid covariance matrix exists for \mathbf{N}_2 . With this decomposition, estimating Θ given \mathbf{Z} amounts to a classic denoising problem (Donoho, 1995) in which the density of the desired signal Θ is characterized by the penalty function $\sum_{i,j} \|\Theta_{i,j}\|_F$. Equation (8) is then used to update the estimate of \mathbf{Z} .

The EM algorithm alternates between an Expectation Step (E-Step) and a Maximization Step (M-Step). The E-Step calculates the Q-function: $Q(\Theta|\Theta^{(k-1)}) \equiv E[\log p(\mathbf{Y}, \mathbf{Z}|\Theta)|\mathbf{Y}, \Theta^{(k-1)}]$ where the expectation is over \mathbf{Z} and $\Theta^{(k-1)}$ is the estimate from the previous iteration. The M-Step maximizes $Q(\Theta|\Theta^{(k-1)}) + \log p(\Theta)$ over Θ to produce $\Theta^{(k)}$. This produces a series of iterates $\Theta^{(k)}$ which are guaranteed to converge to a stationary point provided that $Q(\mathbf{A}|\mathbf{B})$ is continuous in both \mathbf{A} and \mathbf{B} (Wu, 1983).

In the E-Step, we calculate the Q-function using the measured data and the estimate $\Theta^{(k-1)}$:

$$\begin{aligned} Q(\Theta|\widehat{\Theta}^{(k-1)}) &= E[\log p(\mathbf{Y}, \mathbf{Z}|\Theta)|\mathbf{Y}, \widehat{\Theta}^{(k-1)}] \\ &= K - \frac{\|\Theta - \widehat{\mathbf{Z}}^{(k)}\|_F^2}{2\alpha^2} \end{aligned} \quad (9)$$

where K is not a function of Θ and:

$$\begin{aligned}\widehat{\mathbf{Z}}^{(k)} &= E[\mathbf{Z} | \mathbf{Y}, \widehat{\Theta}^{(k-1)}] \\ &= \widehat{\Theta}^{(k-1)} + \frac{\alpha^2}{\sigma^2} \mathbf{S}^T \mathbf{H}^T (\mathbf{Y} - \mathbf{H} \mathbf{S} \widehat{\Theta}^{(k-1)}) \mathbf{T} \end{aligned} \quad (10)$$

Thus, specifying $Q(\Theta | \Theta^{(k-1)})$ amounts to calculating $\mathbf{Z}^{(k)}$ using Eq. (10).

The M-Step is a generalization of the M-Step for an ℓ_1 penalized least squares problem (see Figueiredo and Nowak (2003)) since the penalty in Eq. (6) is a generalization of the ℓ_1 norm. We seek to solve the following problem:

$$\begin{aligned}\widehat{\Theta}^{(k)} &= \arg \max_{\Theta} Q(\Theta | \widehat{\Theta}^{(k-1)}) + \log p(\Theta) \\ &= \arg \min_{\Theta} \left\{ \frac{\|\Theta - \widehat{\mathbf{Z}}^{(k)}\|_F^2}{2\alpha^2} + \frac{\lambda}{2\sigma^2} \sum_{i,j} \|\Theta_{i,j}\|_F \right\} \end{aligned} \quad (11)$$

This is the maximum a posteriori estimate of Θ given $\mathbf{Z}^{(k)}$ assuming

$p(\Theta) \propto \exp\left(-\frac{\lambda}{2\sigma^2} \sum_{i,j} \|\Theta_{i,j}\|_F\right)$. The key feature facilitating the solution of Eq. (11) is the separability of the problem into blocks corresponding to the blocks of Eq. (4):

$$\begin{aligned} & \frac{\|\Theta - \widehat{\mathbf{Z}}^{(k)}\|_F^2}{2\alpha^2} \\ & + \frac{\lambda}{2\sigma^2} \sum_{i,j} \|\Theta_{i,j}\|_F \\ & = \sum_{i,j} \frac{\|\Theta_{i,j} - \widehat{\mathbf{Z}}_{i,j}^{(k)}\|_F^2}{2\alpha^2} \\ & + \frac{\lambda}{2\sigma^2} \|\Theta_{i,j}\|_F \end{aligned} \quad (12)$$

We solve Eq. (11) by finding the $\Theta_{i,j}$ that minimizes $\|\Theta_{i,j} - \widehat{\mathbf{Z}}_{i,j}^{(k)}\|_F^2 + \frac{\alpha^2 \lambda}{\sigma^2} \|\Theta_{i,j}\|_F$ for every pair (i, j) . The single block problem for the $(i, j)^{th}$ block is rewritten in simplified notation using

$\mathbf{z} = \text{vec}(\widehat{\mathbf{Z}}_{i,j}^{(k)})$ and $\theta = \text{vec}(\Theta_{i,j})$ as $f(\theta) = \|\theta - \mathbf{z}\|_2^2 + \frac{\alpha^2 \lambda}{\sigma^2} \|\theta\|_2$. Since $f(\theta)$ is strictly convex, it has a unique minimizer θ^* . We show in Appendix A that:

$$\theta^* = \begin{cases} \left(1 - \frac{\alpha^2 \lambda}{2\sigma^2 \|\mathbf{z}\|_2}\right) \mathbf{z}, & \text{if } \|\mathbf{z}\|_2 > \frac{\alpha^2 \lambda}{2\sigma^2} \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

We summarize the k^{th} iteration of the EM algorithm as:

$$\begin{aligned}\widehat{\mathbf{Z}}^{(k)} &= \widehat{\Theta}^{(k-1)} + c\mathbf{S}^T\mathbf{H}^T(\mathbf{Y} - \mathbf{H}\mathbf{S}\widehat{\Theta}^{(k-1)}\mathbf{T}^T)\mathbf{T} \\ \widehat{\Theta}^{(k)} &= \arg \min_{\Theta} \left\{ \|\Theta - \widehat{\mathbf{Z}}^{(k)}\|_F^2 + c\lambda \sum_{i,j} \|\Theta_{i,j}\|_F \right\}\end{aligned}\quad (14)$$

where c acts as a step size parameter. Setting α^2 at the upper bound $\alpha^2 = \sigma^2\|\mathbf{T}\mathbf{T}^T\|^{-1}\|\mathbf{H}\mathbf{S}\mathbf{S}^T\mathbf{H}^T\|^{-1}$ leads to $c = \|\mathbf{T}\mathbf{T}^T\|^{-1}\|\mathbf{H}\mathbf{S}\mathbf{S}^T\mathbf{H}^T\|^{-1}$. This step size guarantees a decrease in the objective function at each iteration; however, following recent work by Wright et al. (2008), we add an inner loop to the algorithm which starts with a larger step size, then reduces c until the objective function decreases. This simple modification speeds up convergence of the algorithm by one to two orders of magnitude.

As described in (Wu, 1983), continuity of the Q function and the existence of a lower bound to Eq. (6) implies convergence to a stationary point. Since Eq. (6) is convex, the only stationary points are global minima, and $\widehat{\Theta}^{(k)}$ converges to a global minimizer. Various stopping criteria may be used to terminate the algorithm. To ensure a correct solution has been reached, we run the algorithm until the subgradient condition (i.e., the subdifferential of Eq. (6) at $\widehat{\Theta}$ contains $\mathbf{0}$) is satisfied to within a small tolerance.

2.4 EM Continuation

An advantage of our EM algorithm approach is that it provides an easy means to calculate solutions $\widehat{\Theta}(\lambda)$ for a wide variety of λ via a ‘‘continuation’’ approach. Since a small change in λ does not typically produce a large change in $\widehat{\Theta}(\lambda)$, the solution $\widehat{\Theta}(\lambda_1)$ can be used to initialize the algorithm when solving for $\widehat{\Theta}(\lambda_2)$. If $\lambda_2 \approx \lambda_1$, then this initialization point will be close to $\widehat{\Theta}(\lambda_2)$. Furthermore, we show in Appendix B that $\widehat{\Theta}(\lambda) = \mathbf{0}$ for $\lambda \geq \lambda_{MAX}$, where:

$$\lambda_{MAX} = 2\|\max_{i,j} \mathbf{S}_i^T \mathbf{H}^T \mathbf{Y} \mathbf{T}_j\|_2 \quad (15)$$

depends on the data \mathbf{Y} . Thus, we solve for a range of λ by setting $\widehat{\Theta}(\lambda_{MAX}) = \mathbf{0}$, and then solving for $\lambda = (1 - \varepsilon)\lambda_{MAX}$ using $\mathbf{0}$ as the initial guess with ε small, e.g. 0.01. Next we use $\widehat{\Theta}((1 - \varepsilon)\lambda_{MAX})$ as a starting point to find $\widehat{\Theta}((1 - 2\varepsilon)\lambda_{MAX})$ and continue in this manner until we arrive at the desired λ . Selection of λ is discussed in Section 2.6.

2.5 EM Algorithm Performance

The EM algorithm described above solves Eq. (6) efficiently compared to second-order cone programming (SOCP) methods often used to minimize cost functions similar to Eq. (6) (e.g., Ding and He (2007); Malioutov et al. (2005); Ou et al. (2009)). The most computationally expensive calculation in EM is matrix multiplication, namely $(\mathbf{H}\mathbf{S})\widehat{\Theta}^{(k-1)}\mathbf{T}^T$ and $(\mathbf{H}\mathbf{S})^T\mathbf{R}\mathbf{T}$ (where $\mathbf{R} = \mathbf{Y} - \mathbf{H}\mathbf{S}\widehat{\Theta}^{(k-1)}\mathbf{T}^T$). This requires $O((K + T)ML)$ operations (note that significantly fewer operations are typically required since $\widehat{\Theta}^{(k-1)}$ is mostly zeros). In contrast SOCP methods require $O((M + K)L^3)$ operations per iteration when a single temporal block is used, i.e. $\mathbf{T} = \mathbf{T}_1$ (Ou et al., 2009). The cubic complexity of SOCP can be prohibitive in large problems.

There are also qualitative differences between solutions found by EM and those found by SOCP. Solutions found by SOCP are not sparse; that is, many of the coefficients are very small, but not zero. This means users must define some threshold to decide which blocks of coefficients should be set to zero. In contrast EM returns truly sparse solutions, with most $\widehat{\Theta}_{i,j} = \mathbf{0}$, eliminating the threshold selection problem. The fact that most $\widehat{\Theta}_{i,j} = \mathbf{0}$ follows from

the M-Step shrinkage procedure (see Eq. (13)). It is known that exact solutions to ℓ_1 penalized regularization problems are truly sparse (see e.g. Donoho (2006)). Also note that our EM approach efficiently finds solutions for multiple λ , while SOCP must start from scratch for each λ of interest.

The main criticism of EM algorithms in general is that they converge slowly when the cost function is relatively flat. This slow convergence is resolved in three ways in the ESP procedure. First, our modified EM algorithm uses a variable step size to speed up convergence in flat areas (see Sec. 2.3). Second, EM typically identifies nonzero $\Theta_{i,j}^{\wedge}$ quickly, but takes longer to adjust the coefficients within in each block. Since debiasing is typically applied to the EM result (see Sec. 2.8), the algorithm may be stopped once the “sparsity pattern” has stabilized. Third, use of EM continuation (see Sec. 2.4) provides a sequence of good initial guesses. Examples of convergence behavior are given in Fig. (15).

2.6 Penalty Weight Parameter Selection

The quality of the estimate Θ^{\wedge} given by Eq. (6) depends on the penalty weight parameter λ . Intuitively, large λ corresponds to heavy reliance on the prior and gives a very sparse (or all zero) estimate, while small λ corresponds to heavy reliance on the noisy data and gives solutions with activity spread all over the cortical surface. Clearly, both of these extremes are undesirable. In general, user expertise may provide the best results for selecting λ , and in many cases it may be desirable to consider a range of solutions from multiple values of λ . However, automatic selection of λ is necessary for objective performance evaluation and to provide a starting point for user based selection.

We have developed a simple heuristic to automate selection of λ based on the number of nonzero $\Theta_{i,j}^{\wedge}$, that is, the cardinality of $A^{\wedge}(\lambda) = \{(i, j) | \Theta_{i,j}^{\wedge}(\lambda) \neq \mathbf{0}\}$. The intuition behind our heuristic is the following. Suppose no signal is present, i.e. $\mathbf{Y} = \mathbf{N}$, and we solve Eq. (6). Since white noise is spread roughly equally through all subspaces, no single block $\Theta_{i,j}^{\wedge}$ can account for a significant portion of the measured data \mathbf{Y} . Thus, as λ decreases, the number of blocks in $A^{\wedge}(\lambda)$ increases rapidly to reduce the error term in Eq. (6). In contrast, when a signal described by a small number of STEs is present, the signal portion of the measured data can be explained by the corresponding blocks, and initially the cardinality of $A^{\wedge}(\lambda)$ increases relatively slowly as λ decreases. Once the active blocks are able to explain the signal component of the data, then additional blocks attempt to represent noise and the number of active blocks will increase rapidly as λ is decreased further. This behavior is illustrated in Fig. 1. Note that the number of STEs increases rapidly as λ decreases for the noise only case, while it initially increases very slowly when signal components are present.

These observations suggest a simple strategy analogous to the L-curve method of regularization in which we select the value of λ where the number of STEs begins to increase rapidly with further decrease in λ . We implement one such strategy by comparing the cardinality of $A^{\wedge}(\lambda)$, which we denote $\#A(\lambda)$, with a piecewise approximation based on two curve segments: a linear function for large λ and quadratic function for small λ . The linear function connects the points $(\lambda_{MAX}, 0)$ and $(\lambda, \#A(\lambda))$. The quadratic function is uniquely defined by two fixed endpoints since it is purely quadratic (i.e., has no linear term). The endpoints used for the quadratic are $(\lambda, \#A(\lambda))$ and $(0, \#A^{\wedge}(0))$, where $\#A^{\wedge}(0)$ is a rough estimate of $\#A(0)$ found by dividing the rank of the forward model by the number of bases in each block. This is the number of active blocks required to produce an estimate which fits any measured data exactly. We vary λ to find the value that minimizes the squared error between $\#A(\lambda)$ and the piecewise approximation. This value of λ is used to identify the approximate point at which the transition between fitting signal and noise occurs.

This method of λ selection requires that $\#A(\lambda)$ be calculated for small enough λ such that the characteristic shape of Fig. 1 is observed before applying the heuristic. Two examples of the heuristic λ selection are shown in Fig. 2. These two examples suggest that the λ selected with this heuristic is typically smaller than the “optimal” λ . Hence, this heuristic provides a good lower bound on λ .

2.7 Non-white Noise and Lead Field Normalization

A key assumption behind the ESP method as presented so far is that the signal is contaminated with white noise. We accommodate colored noise by first whitening the data with the noise covariance matrix. Suppose the noise in Eq. (1) is distributed $\mathbf{N} \sim N(\mathbf{0}, \mathbf{R}_N)$. Taking the matrix square root $\mathbf{R}_N = \mathbf{R}_N^{1/2} \mathbf{R}_N^{T/2}$, the whitened data is:

$$\begin{aligned} \tilde{\mathbf{Y}} &= \mathbf{R}_N^{-1/2} \mathbf{Y} = \mathbf{R}_N^{-1/2} \mathbf{H} \mathbf{X} \\ &\quad + \mathbf{R}_N^{-1/2} \mathbf{N} = \tilde{\mathbf{H}} \mathbf{X} + \tilde{\mathbf{N}} \end{aligned} \quad (16)$$

where $\tilde{\mathbf{N}} \sim N(\mathbf{0}, \mathbf{I})$. ESP is applied to the whitened system. In practice \mathbf{R}_N is usually unknown, so whitening requires that \mathbf{R}_N be estimated from available data.

After whitening the data and expanding the signal with ESP basis functions, the signal model is:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{H}} \mathbf{S} \mathbf{\Theta} \mathbf{T}^T + \tilde{\mathbf{N}} \quad (17)$$

We normalize the lead fields using diagonal matrix \mathbf{W} whose entries equal the ℓ_2 norm of the corresponding column of $\tilde{\mathbf{H}} \mathbf{S}$. This eliminates bias toward sources that are closest to the sensors. Thus we have:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{H}} \mathbf{S} \mathbf{W}^{-1} \mathbf{W} \mathbf{\Theta} \mathbf{T}^T + \tilde{\mathbf{N}} \quad (18)$$

Here the columns of the matrix $\tilde{\mathbf{H}} \mathbf{S} \mathbf{W}^{-1}$ are unit norm due to the normalization. The columns of \mathbf{T} may be normalized as well, though typically the \mathbf{T}_j are constructed so that $\mathbf{T}_j^T \mathbf{T}_j = \mathbf{I}$ and thus normalization of \mathbf{T} is not necessary. The EM algorithm solves for $\mathbf{\Theta}' = \mathbf{W} \mathbf{\Theta}$.

2.8 Debiasing

The penalty in Eq. (6) tends to “shrink” the coefficient matrices $\mathbf{\Theta}_{ij}$ and produces biased amplitude estimates. This bias can be removed by using ESP to estimate the set of active STEs: $A = \{(i, j) | \mathbf{\Theta}_{ij} \neq \mathbf{0}\}$ and then solving a least squares problem on this set to produce a refined estimate of $\mathbf{\Theta}_{ij}$, for all $(i, j) \in A$. We employ a minimum norm solution to the inverse problem defined on active blocks A to obtain the refined estimate. This process is referred to as debiasing in $(x02113)_1$ regularized inverse problems since it removes some of the bias introduced by the penalty term in Eq. (6). Note that debiasing does not alter the spatial or temporal extent of the ESP solution e.g., introduce blurring, since the set of active STEs does not change. The debiasing step may be performed without lead field normalization. Also, different basis expansions for the active blocks may be used in debiasing than those used for the original ESP

regularization. The debiasing step is not necessary for the ESP regularization approach, especially if source localization is the primary goal.

3 Results

We illustrate examples of ESP reconstruction in this section. We begin by describing the specific spatio-temporal bases used for the examples. Next, we present a simulation in which the true activity consists of three space-time events and the measurements are contaminated by realistic sensor noise, followed by results from a randomized simulation involving two cortical signals. Finally, we apply ESP reconstruction to measurements collected during a self-paced finger tapping experiment.

3.1 Space- Time Basis Function Selection for Examples

As noted previously there are many possible ways to define STEs. We illustrate the attributes of the ESP reconstruction approach using STEs defined spatially by cortical patches and temporally by fixed time windows and a common bandwidth. Use of cortical patches to design spatial bases results in event bases that are locally supported and particularly parsimonious. This significantly reduces the dimension of the inverse problem. The basis construction procedure described here exemplifies several key attributes of appropriate bases and is derived from the cortical patch bases of Limpiti et al. (2006).

Assume the i^{th} cortical patch is tessellated with a set of dipoles whose lead fields and amplitudes are concatenated into a patch lead field matrix \mathbf{H}^i and amplitude vector α^i , respectively. Hence, the signal \mathbf{y}^i measured at the sensors due to the i^{th} patch is $\mathbf{y}^i = \mathbf{H}^i \alpha^i$. The bases representing events in this patch should only describe the space of possible activity distributions α^i that contribute to the measurement. To this end, take the singular value decomposition of $\mathbf{H}^i = \mathbf{U}^i \Sigma^i \mathbf{V}^{iT}$ where the columns of \mathbf{U}^i and \mathbf{V}^i are the left and right singular vectors, respectively, and Σ^i is a diagonal matrix of singular values ordered from largest to smallest. Typically \mathbf{H}^i is low rank (Limpiti et al., 2006) so a small number of the singular values are significant.

The columns of \mathbf{V}^i are a natural basis for the patch signal α^i . Hence, we may express

$\alpha^i = \sum_k a_k \mathbf{v}_k^i$ where \mathbf{v}_k^i is the k^{th} column of \mathbf{V}^i and a_k are the expansion coefficients. The corresponding signal measured at the sensor is:

$$\mathbf{Y}^i = \sum_k a_k \mathbf{u}_k^i \sigma_k^i \quad (19)$$

where \mathbf{u}_k^i is the k^{th} column of \mathbf{U}^i and σ_k^i is the k^{th} largest singular value. The strength of the k^{th} component of the measured signal is proportional to σ_k^i , so spatial bases \mathbf{v}_k^i associated with small singular values make small contributions to the measured signal. Given the typically low signal to noise ratio of M/EEG, the measurable spatial component of an STE from the i^{th} patch is adequately represented by the right singular vectors corresponding to the largest singular values. Hence, we construct the spatial basis \mathbf{S}_i from the “significant” right singular vectors associated with the i^{th} patch.

Figure 3 depicts the singular values for two sets of patches computed for the 275 channel CTF MEG system. The singular values shown in Figure 3(a) correspond to all 393 patches of geodesic radius 20 mm which tile both hemispheres of a subject with 50% overlap. Note that in each case the 7th and higher singular values are an order of magnitude smaller than the largest singular value. Furthermore, the first two or three singular values are five times larger than the

higher order singular values. The number of significant singular values is related to the size of the patch. Figure 3(b) shows singular values corresponding to 57 patches of geodesic radius 50 mm which again tile both hemispheres with 50% overlap. The singular values for the larger patches decay more slowly; the 10th and higher singular values are an order of magnitude smaller than the largest singular value in a given patch.

Figure 4 illustrates the spatial pattern associated with several right singular vectors for a typical 20 mm patch by depicting each element of \mathbf{v}_k at the corresponding dipole location in the patch. The singular vectors associated with the largest singular values tend to represent low resolution features within the patch. The level of detail or fine structure represented by a singular vector increases as the corresponding singular values decrease. Selecting the number of significant singular values and vectors thus involves a tradeoff between representing the details of the spatial distribution of activity within the patch and having adequate SNR to estimate the detail. Relatively high SNR is required to reliably identify the fine structure associated with small singular values. Similar tradeoffs exist with other approaches to choosing spatial bases. For example, the contributions of higher order terms in a multipole expansion also decay rapidly, so the SNR dictates the number of terms that can be reliably estimated and used as spatial bases.

For the examples that follow, we choose patches of 20 mm geodesic radius on the cortical surface extracted from an MRI of the subject. Note that geodesic distances calculated using the edges of triangles representing the cortical surface are greater than the true geodesic distance. Figures 13 and 14 illustrate typical 20 mm patches. We choose three spatial bases per patch for each STE since the majority of measurable activity in any patch is represented by the first three singular vectors, as indicated in Fig. 3.

We represent the temporal aspect of STEs for our examples by dividing the total time extent of interest into time windows that overlap by 50%. A wavelet basis that spans the frequency band of interest (0-100 Hz) is used for each window. We have found that wavelet bases result in smooth transitions at the window edges. We use 32 wavelets per time interval, so the number of coefficients associated with each STE is 96 ($\Theta_{i,j}$ is 3 by 32).

3.2 Simulated Example 1

We test the ESP approach using simulated data from a 275 channel CTF MEG system. The simulated cortical signal consists of three STEs: activation of a patch in both the left and right primary auditory cortex with a slight offset in time followed by a patch of activity in the motor cortex. All three simulated patches have geodesic radii of roughly 10 mm. The patches used as simulated sources do not correspond to patches used in the ESP procedure; the simulated source patches are smaller and have different center locations. Furthermore, the simulated spatial activity pattern within each patch (raised cosine) does not correspond to the singular vectors used to represent any patch. Lastly, the temporal evolution of the signal, although primarily low-pass, does not correspond directly to the wavelet bases used in reconstruction. The measurements are simulated using Eq. (1) for 60 trials, with \mathbf{N} consisting of spatially colored noise drawn from a $N(\mathbf{0}, \mathbf{I} \otimes \mathbf{R})$ distribution where \mathbf{R} was calculated from the prestimulus portion of an MEG experiment. The noise level is set to achieve an average SNR (defined as $\|\mathbf{H}\mathbf{X}\|_F^2 / E[\|\mathbf{N}\|_F^2]$) of -6 dB after averaging over all epochs. Figure 5 illustrates the time courses and spatial extents of the three space-time events, as well as the noisy measurements.

After whitening the data as described in Section 2.7 using a noise covariance matrix estimated by removing the mean from each trial, we apply ESP reconstruction to the simulated signal using the STEs described in Section 3.1. We use 7 overlapping time windows of 64 ms each.

Combined with the 393 spatial patches, this results in a total of 2751 candidate STEs, each with 96 coefficients.

The heuristic of Section 2.6 is used to select λ . The result of debiasing the ESP reconstruction for this λ is shown in Figs. 6 and 7. All three STEs are recovered at the correct locations in space and time. Activity is spread spatially around the true location. Some of this spreading is due to the use of larger patches in the ESP reconstruction, while some spreading is inherent to the ill-posed nature of the inverse problem. We also point out that some temporal spreading occurs as a result of noise lying in the subspace defined by fixed time windows.

This is evident between 97 and 105 ms, where the motor cortex exhibits low-level activity in the reconstruction, though the true motor activity starts at 106 ms. The fixed time window in which the motor cortex is active begins at 97 ms. The reconstructed cortical activity in Fig. 7 (d) – 7(f) is consistent with the true sources depicted in Fig. 5(d).

We now illustrate the flexibility of ESP by employing different sets of temporal basis functions for the \mathbf{T}_j . We also show the advantage of assuming sparsity in the temporal domain. First, we use the same seven overlapping time windows, but perform singular value decomposition (SVD) of the noisy data in each window and keep the two most significant singular vectors for each \mathbf{T}_i – at most two singular values are visibly much larger than the others in the SVDs of the data in each window. The technique of using SVD of the data to determine the temporal signal subspace has been used previously in M/EEG (Huang et al., 2006; Ou et al., 2009). The resulting reconstruction after heuristic λ selection 8(a). The result is similar to that obtained using wavelet bases, although there is some additional spatial blurring (not shown). Next, we repeat the experiment with temporal bases derived from the SVD of the entire time period of the experiment (e.g. 0 to 256 ms). We use the three most significant singular vectors to form $\mathbf{T} = \mathbf{T}_1$. The result after heuristic λ selection and debiasing is shown in Figs. 8(b) and 9. In this case, the reconstructed activity extends throughout the measurement period due to the influence of noise on the SVD bases, even though the true signal becomes negligible after about 200 ms. Also, Fig. 9 shows that each cortical area involved in the signal estimate has activity which persists throughout the measurement period. This example shows the advantage of assuming sparsity in the temporal domain; when time windows are used, ESP can automatically set the estimate $\hat{\mathbf{X}}_i$ to zero when there is not sufficient evidence of activity during a particular time period, and identify the time intervals when each source is active.

In our final variation on this example, we illustrate an alternative way to exploit sparsity in the temporal domain. Here we consider a scenario involving a “slow” event which spans a long period of time and two “fast” events spanning much shorter time periods. These signals are illustrated in Fig. 10. Simulated measurements are generated using Eq. (1) assuming white noise at 0 dB. We apply ESP using multiscale temporal blocks. For \mathbf{T}_1 we use 32 low frequency, large scale wavelets which occupy the entire time period. In contrast, \mathbf{T}_2 and \mathbf{T}_3 consist of 32 wavelets each spanning, respectively, the first and second half of the 256 ms measurement period. Thus the temporal blocks implement a time-frequency partition of the data. We illustrate the resulting reconstruction (after heuristic λ selection and debiasing) in Fig. 11. ESP estimates the “slow” portion of the data using the large scale wavelets (\mathbf{T}_1). The “fast” components are reconstructed with the shorter scale, higher frequency bases of \mathbf{T}_2 . \mathbf{T}_3 is identified by ESP as inactive. This example illustrates the use of temporal sparsity to identify events spanning different time scales. Similarly, by appropriate basis selection, temporal sparsity can be used to identify events spanning different frequency bands.

3.3 Simulated Example 2

We now conduct a series of randomized simulations to illustrate the performance of the ESP procedure under various signal conditions. Two spatially extended cortical sources of activity

are present in each simulation. Each source follows a damped cosine time course as shown in Fig. 12. Trials with less than about 25 ms between the onset of source 1 and source 2 exhibit temporal overlap at the sensors. For example, the onset of source 2 is 14 ms later than that of source 1 in Fig. 12 and significant temporal overlap is present. The location, size, and starting time of each source is selected at random to provide a rich variety of source configurations. Specifically, the center point of each event is selected uniformly at random from all 24514 voxels on the cortical surface. The geodesic radius of each source is selected uniformly at random between 8 and 40 mm. The amplitude of activity of each dipole in a source is given

by $\cos\left(\frac{1.5\Delta}{r}\right)$ where Δ is the geodesic distance from the center of the source and r is the geodesic radius of the source. Finally, the onset of activity in each source is selected uniformly at random between 0 and 75 ms. The measurement period is 128 ms and the average SNR is set to 0 dB.

ESP reconstruction is performed using the 20 mm radius spatial bases and 64 ms duration wavelet temporal bases of Sec. 3.1 with three overlapping time windows. We calculate the distance error between the center of the true source and the center of the debiased ESP estimate of that source at the peak time sample. The center is found by taking a signal energy weighted average of all active voxels. When sources overlap in time, a weighted k-means algorithm is used to separate voxels into two groups (one for each source). The source one and two SNR is defined as $\|\mathbf{HX}_1\|_F^2/E[\|\mathbf{N}\|_F^2]$ and $\|\mathbf{HX}_2\|_F^2/E[\|\mathbf{N}\|_F^2]$, respectively. These values depend on both the source spatial extent and the source location. Note that the individual source SNRs do not sum to the overall SNR of 0dB because of spatio-temporal interaction between the sources. The heuristic of Section 2.6 is used to select λ in each trial. Table 1 summarizes the results of 10 trials.

Figure 13 illustrates ESP spatial reconstruction for both sources in Trial 6. Some spatial spreading occurs around both sources. ESP identifies low level activity in the region of source 2 before the activity begins. This is a result of noise present between the edge of the active time window and the onset of activity. Figure 13(d) shows the ESP reconstruction at the time of the peak source 2 activity. This time point corresponds roughly to the second temporal peak of the source 1 activity (see Fig. 12), so activity associated with source 1 is visible.

Spatial reconstructions for Trial 8 are shown in Fig. 14. Here again spatial spreading is evident. Though the correct location is included in the solution for both sources, the peak of activity in both cases is biased.

We conclude this example with an illustration of EM algorithm convergence performance. Figure 15 shows the number of EM iterations required to find each $\hat{\Theta}(\lambda)$ for all ten trials of Table 1. For nearly all trials, fewer than 50 iterations are required for $\lambda \geq 0.3\lambda_{MAX}$. When $\lambda = 0.1\lambda_{MAX}$, about 500 iterations are required. This increase in iterations corresponds to an increase in the number of active STEs in the solution (see Fig. 1).

3.4 Self-Paced Finger Tapping Experiment

We also applied ESP to MEG measurements of a subject performing self-paced finger tapping. The data were collected with a Magnes 3600 whole head system at a sampling rate of 678 Hz and band pass filtered to lie between 2 and 40 Hz. The measured data were grouped into 80 epochs of 755 ms duration (512 samples) referenced to the tap at time 0 detected via a switch closure. Whitening was performed via Eq. (16) where \mathbf{R}_N was estimated by removing the average response from all epochs and calculating a sample covariance matrix by averaging over time and epochs. The averaged sensor measurements before and after whitening are shown in Fig. 16. Cortical patches of 20 mm geodesic radius are used to construct the spatial bases \mathbf{S}_i . Seven possible temporal events are represented by dividing the 755 ms measurement

interval into overlapping 189 ms (128 samples) intervals. Each interval is represented using 32 wavelet bases.

The debiased reconstruction shown in Fig. 17 is calculated using the heuristic of Section 2.6 to select λ . The estimate of activity begins about 200 ms pre-tap and persists through the end of the measurement period. The earliest activity occurs in the left motor cortex, which corresponds with the action of initiating the finger tap. The next region to become active spans the somatosensory and posterior parietal cortex. The somatosensory activity likely reflects the sensation of pushing down the switch. The posterior parietal cortex is involved in spatial relationships, body image, and movement planning (Bear et al., 2001). The parietal lobe activity begins at low levels about 100 ms before the switch closes and continues throughout the measurement period. It is possible that the early onset of activity in the parietal lobe is a result of noise in the same temporal subspace as an active time window. Beginning at about 0 ms the ESP reconstruction includes a small patch on the left frontal lobe. Some weak activity is also identified in the right hemisphere, primarily in the parietal lobe. This is consistent with earlier studies' finding that the right parietal cortex plays a role in spatial processing (Culham et al., 2006).

4 Discussion

The ESP approach searches for a small number of STEs which best explain the measurements. An appropriately chosen basis function representation for each STE reduces the number of parameters in the inverse problem by only representing activity that is measurable at the M/EEG sensors. Thus, our approach can incorporate the inherent limitations of the forward problem to formulate a more tractable inverse problem. The spatial bases can be chosen to represent extended sources such as patches or multipoles. The temporal basis functions naturally constrain the solution based on temporal features of interest, such as the time-frequency characteristics. Use of both spatial and temporal bases allows ESP to automatically determine both where and when significant activity occurs. A similar joint spatio-temporal philosophy based on beamforming has been proposed recently by Dalal et al. (2008).

We have illustrated the effectiveness of the ESP approach on simulated data and MEG recordings from a finger tapping paradigm. The simulated example in Sec. 3.2 shows the benefit of exploiting spatio-temporal sparsity in a three source scenario. The randomized two source scenarios in Sec. 3.3 indicate that ESP is capable of recovering almost all cortical sources with the majority of centroid localization errors less than 1 cm. The centroid localization errors are generally inversely proportional to the source SNR. In Trials 3 and 7, the weaker of the two sources was not detected using ESP due to the very low SNRs associated with these sources. In the self-paced finger tapping experiment of Sec. 3.4, ESP recovers spatio-temporal activity consistent with known brain behavior.

Our examples include source activity of varying complexity. In some cases, individual events have similar powers. Others include a weak source and a strong source. The spatial separation and temporal overlap of sources also vary. Evaluating the degradation in performance of ESP (or any other approach) as a function of increasing source complexity is an important problem. Unfortunately, quantifying this degradation is difficult, as "source complexity" can include both the number of events and the pattern of spatial and/or temporal activity of each source. We believe the use of temporal priors in the ESP approach renders it more robust to degradation due to source complexity than methods which use only spatial priors. Sources falling in different temporal subspaces are easier to separate with ESP than simple ℓ_1 spatial regularization at a single time point.

The use of sparse representations in the spatial and temporal/frequency domains for M/EEG is not unique to our approach. Sparsity in the spatial domain is enforced by limiting the number of dipolar, multipolar, patch based, or “meso-” sources used to represent measured data, usually via an ℓ_1 penalty (Auranen et al., 2005; Daunizeau and Friston, 2007; Ding and He, 2007; Gorodnitsky and Rao, 1997; Huang et al., 2006; Jeffs et al., 1987; Matsuura and Okabe, 1995, 1997; Ou et al., 2009; Uutela et al., 1999). Sparsity in the time domain has been used both implicitly and explicitly in M/EEG. One common implicit approach is to project measured data onto a few temporal basis functions derived from a singular value decomposition of the data (Huang et al., 2006; Ou et al., 2009). The more explicit approaches often involve finding a small number of wavelets which represent the signal well. This is accomplished via thresholding of wavelet coefficients (Effern et al., 2000; Trejo and Shensa, 1999) or wavelet shrinkage approaches (Trujillo-Barreto et al., 2008). A similar approach using damped sinusoids instead of wavelets is proposed in (Demiralp et al., 1998).

The ESP criterion involves the weighted combination of the squared fitting error and a penalty designed to encourage solutions consisting of a small number of STEs. The resulting cost function is convex and is minimized with a novel EM algorithm. The EM algorithm efficiently identifies a parameterized family of solutions that balance sparsity against squared error. By choosing appropriate bases and minimizing the number of STEs, ESP solutions avoid the “spiky” spatial and temporal artifacts typical of ℓ_1 methods that minimize the number of active dipoles at individual time instants.

We propose choosing the penalty weight by evaluating the number of STEs in the solution as the weight decreases from the maximum value. The number of STEs increases relatively slowly as the weight decreases when unrepresented signal components are present, and then increases very rapidly as the weight decreases when the STEs in the solution begin to represent noise. Hence, the weight value at which the number of STEs begins to increase rapidly is a good choice. We present one automated approach for finding this value of rapid increase based on fitting a curve to the number of STEs.

The ESP approach differs significantly from recent work on ℓ_1 regularization. In most existing ℓ_1 approaches, the regularization term penalizes every active dipole via $\|\text{vec}(\mathbf{X})\|_1$ or, if applied in the context of a basis expansion (Eq. (4)), penalizes every active coefficient via $\|\text{vec}(\Theta)\|_1$. In contrast, ESP gives a smaller penalty to active coefficients in the same block than to active coefficients in different blocks. This promotes solutions consisting of a few active spatio-temporal blocks, rather than a few active coefficients or dipoles, while accommodating uncertainty in the spatio-temporal evolution of each block. Given appropriate spatial and temporal bases for each STE, ESP favors solutions which are connected spatially and vary smoothly over time as opposed to the “spiky”, disconnected results often seen in standard ℓ_1 minimization.

The mesostate model recently proposed by Daunizeau and Friston (2007) also treats cortical activity as a collection of a small number of events, or mesosources, which are identified through a Variational Bayesian (VB) inversion. The mesostate approach differs in a number of ways from the ESP approach. First, the maximum number of meso-sources must be known a priori. In Daunizeau and Friston (2007), this number is between one and eight. Second, the VB inversion “switches off” mesostates with little or no evidence. The ESP approach, on the other hand, works by “switching on” STEs with strong evidence from a large number of candidate STEs. Third, each meso-source is described by a mean time course and spatial location. Dipoles comprising a meso-source are modeled as perturbations of this mean activity. An STE, on the other hand, describes any measurable activity occurring in a local region of space and time, and thus is not limited to representation of average activity.

Another recent source imaging method proposed by Ou et al. (2009) uses a cost function similar to our ESP cost function. Unlike our approach, Ou et al. (2009) do not exploit potential sparsity in the temporal and/or frequency domains to separate sources. Instead, the measured data is projected onto a few dominant singular vectors of the data, similar to the VESTAL approach. As a result, these bases span the entire measurement period under study. The main downside to this approach is that, unlike ESP, it can not resolve temporally localized activity (compare Figs. 6 and 9). Furthermore, the method proposed in Ou et al. (2009) decimates the mesh representing the cortex, using single dipolar sources to represent larger cortical areas. A third difference between ESP and the Ou et al. (2009) approach is the method by which the cost function is minimized. Ou et al. (2009) use second-order cone programming on problems with far fewer parameters. ESP uses an novel EM algorithm which scales linearly with problem dimensions, whereas SOCP scales cubically. SOCP approaches are computationally intractable for the problem sizes considered in this paper.

Several block penalty methods similar to our ESP approach have been proposed (Malioutov et al., 2005; Tropp, 2006; Turlach et al., 2005; Yuan and Lin, 2006), though none of them in the context of M/EEG. All of these aim to find solutions which are sparse spatially, but not temporally, i.e., are sparse in only one of the dimensions. Also, all but Yuan and Lin (2006) construct blocks containing only a single spatial basis repeated over multiple time samples. This corresponds to defining blocks based on single rows of \mathbf{X} . Malioutov et al. (2005) penalize the sum of the ℓ_2 norms of each row of the signal matrix in direction of arrival (DOA) estimation for radar. SOCP is used to perform the optimization. Both Turlach et al. (2005) and Tropp (2006) use the sum of ℓ_∞ norms as a penalty, but again consider only single rows of the signal matrix as blocks and do not employ a temporal decomposition. Turlach et al. (2005) solve a constrained version of the ESP problem via an interior point method. Tropp (2006) does not specify an optimization algorithm, but instead provides useful analysis of the properties of minimizers. Yuan and Lin (2006) suggest essentially fixed point iteration on the nonlinear functions comprising the subgradient condition for a minimizer of Eq. (6). However, they indicate that computational complexity limits their approach to small problems. None of these earlier works consider the cost function as a Bayesian likelihood, nor does there appear to be any published Expectation-Maximization approach to these “block-sparse” problems.

The choice $p = 2$ in Eq. (5) results in a convex optimization and appears to yield very good results. However, modifying the M-Step in Eq. (14) slightly allows solution of Eq. (5) for other values of p as well. In this case Eq. (11) is written:

$$\widehat{\Theta}^{(k)} = \arg \min_{\Theta} \left\{ \|\Theta - \widehat{\mathbf{Z}}^{(k)}\|_F^2 + c\lambda \sum_{i,j} \|\text{vec}(\Theta_{i,j})\|_p \right\} \quad (20)$$

Exact solutions to the M-Step represented by Eq. (20) can be obtained for the cases $p = 1$ and $p = \infty$ in addition to $p = 2$. The Generalized Expectation-Maximization approach (Gellman et al., 2003) can be used to solve Eq. (20) for other values of p .

Using a set of spatial basis functions based on cortical patches and wavelet temporal bases, we demonstrated the effectiveness of ESP regularization on simulated cortical signals as well as data collected from a self-paced finger tapping experiment. These specific basis functions are not required to use ESP. Other methods for constructing the \mathbf{S}_i and \mathbf{T}_j may also prove effective in specific problems. As one example, short time Fourier bases could be used to isolate activity in different frequency bands over the same time window. Similarly, the spatial bases could be

defined for patches associated with anatomically meaningful regions of the brain such as Brodmann areas.

An anatomical approach to patch design is used in the recently proposed MiMS technique (Cottreau et al., 2007). The spatial extent of each patch is chosen based on the local characteristics of the cortical surface. A multipolar expansion of fixed order is used to represent the activity from candidate patch. MiMS employs a coarse to fine strategy to identify active regions. It begins with large patches, finds a subset of active patches that best fit the data using cross validation, subdivides active patches and repeats the patch selection/subdivision process until a fitting criterion is satisfied. MiMS does not address selection of temporal areas of interest or solve for activity jointly over space and time.

The assumptions behind ESP apply to a wide range of experimental paradigms and can accommodate varying specific assumptions through the choice of basis functions. Simulated and actual MEG data suggest ESP is effective at localizing cortical activity in both space and time.

A Minimizer of Single Block Problem

We show that θ^* given by Eq. (13) is the minimizer of $f(\theta) = \|\theta - \mathbf{z}\|_2^2 + \frac{\alpha^2 \lambda}{\sigma^2} \|\theta\|_2$ using facts from the theory of subgradients (Rockafellar, 1970). A vector \mathbf{g} is a subgradient of a convex function $f(\theta)$ at θ_0 if $f(\theta) - f(\theta_0) \geq \mathbf{g}^T(\theta - \theta_0)$ for all θ . The set of all subgradients is called the subdifferential. The subdifferential generalizes the gradient; if there is only one subgradient of $f(\theta)$ at θ_0 , it is the gradient. A point θ^* is the global minimizer of a convex function if and only if $\mathbf{0}$ is a subgradient of $f(\theta)$ at θ^* .

It is easy to see that the subdifferential of $\|\theta\|_2$ contains only the gradient $\frac{\theta}{\|\theta\|_2}$ when $\theta \neq \mathbf{0}$, and any vector \mathbf{g} such that $\|\mathbf{g}\|_2 \leq 1$ when $\theta = \mathbf{0}$. Now consider two cases. Case 1: Suppose

$\|\mathbf{z}\|_2 > \frac{\alpha^2 \lambda}{2\sigma^2}$. Then $\theta^* = (1 - \frac{\alpha^2 \lambda}{2\sigma^2 \|\mathbf{z}\|_2}) \mathbf{z}$ from Eq. (13). Since $\theta^* \neq \mathbf{0}$, the subdifferential of $f(\theta)$ at

$$\begin{aligned} 2(\theta^* - \mathbf{z}) + \frac{\alpha^2 \lambda}{\sigma^2} \frac{\theta^*}{\|\theta^*\|_2} \\ = -\frac{\alpha^2 \lambda}{\sigma^2 \|\mathbf{z}\|_2} \mathbf{z} \\ + \frac{\alpha^2 \lambda}{\sigma^2} \frac{\mathbf{z}}{\|\mathbf{z}\|_2} = 0. \end{aligned}$$

θ^* contains only the gradient: Thus, θ^* is a minimizer of $f(\theta)$ when

$\|\mathbf{z}\|_2 > \frac{\alpha^2 \lambda}{2\sigma^2}$. Case 2: Suppose $\|\mathbf{z}\|_2 \leq \frac{\alpha^2 \lambda}{2\sigma^2}$. Then $\theta^* = \mathbf{0}$ from Eq. (13). Let $\mathbf{g}^* = \frac{2\sigma^2 \mathbf{z}}{\alpha^2 \lambda}$ and note that

\mathbf{g}^* is a subgradient of $\|\theta\|_2$ at $\mathbf{0}$ since $\|\mathbf{g}^*\|_2 \leq 1$. This means that $2(\theta^* - \mathbf{z}) + \frac{\alpha^2 \lambda}{\sigma^2} \mathbf{g}^*$ is a subgradient of $f(\theta)$ at $\mathbf{0}$. To complete the proof, note that $2(\theta^* - \mathbf{z}) + \frac{\alpha^2 \lambda}{\sigma^2} \mathbf{g}^* = -2\mathbf{z} + 2\mathbf{z} = 0$.

B Derivation of λ_{MAX}

We show that $\hat{\Theta}(\lambda) = \mathbf{0}$ for $\lambda \geq \lambda_{MAX}$ given in Eq. (15). We again use the theory of subgradients (Rockafellar, 1970). We first rewrite the objective function of Eq. (6) as

$$f(\theta) = \|\mathbf{Y} - \sum_{i,j} (\mathbf{T}_j \otimes \mathbf{HS}_i) \theta_{i,j}\|_2^2 + \lambda \sum_{i,j} \|\theta_{i,j}\|_2$$
, where θ is a vector concatenation of all the $\theta_{i,j}$,
 $\theta = [\theta_{1,1}^T \theta_{1,2}^T \dots \theta_{2,1}^T \dots \theta_{i,j}^T]^T$. The subdifferential of $f(\theta)$ at $\theta = \mathbf{0}$ contains only the vectors whose $(i,j)^{th}$ component is $-2(\mathbf{T}_j \otimes \mathbf{HS}_i)^T (\mathbf{y} - \mathbf{0}) + \lambda \mathbf{g}_{i,j}$ where $\|\mathbf{g}_{i,j}\|_2 \leq 1$ for all i, j . Let $\mathbf{g}_{i,j} = \frac{2}{\lambda} (\mathbf{T}_j \otimes \mathbf{HS}_i)^T \mathbf{y}$ and suppose $\lambda \geq 2 \max_{i,j} (\mathbf{T}_j \otimes \mathbf{HS}_i)^T \mathbf{y}$. Then $\|\mathbf{g}_{i,j}\|_2 \leq 1$ for all i, j and the subdifferential of $f(\theta)$ at $\mathbf{0}$ contains the all zero vector, implying that $\theta = \mathbf{0}$ minimizes $f(\theta)$.

Acknowledgments

The authors thank the MEG Lab at the University of Texas Health Science Center Houston for the finger tapping data.

References

- Auranen T, Nummenmaa A, Hämäläinen M, Jääskeläinen I, Lampinen J, Vehtari A, Sams M. Bayesian analysis of the neuromagnetic inverse problem with ℓ^p -norm priors. *NeuroImage* 2005;26(3):870–884. [PubMed: 15955497]
- Baillet S, Garnero L. A Bayesian approach to introducing anatomic-functional prior in the EEG/MEG inverse problem. *IEEE Trans Biomed Eng* 1997;44(5):374–385. [PubMed: 9125822]
- Baillet S, Mosher J, Leahy R. Electromagnetic brain mapping. *IEEE Signal Processing Magazine* 2001;18(6):14–30.
- Bear, M.; Connor, B.; Paradiso, M. *Neuroscience: Exploring the Brain*. 2nd. Lippincott, Williams and Wilkins; Philadelphia, PA: 2001.
- Bolstad, A.; Van Veen, B.; Nowak, R. *IEEE Statistical Signal Processing Workshop*. Madison, WI: Aug. 2007 Magneto-/electroencephalography with space-time sparse priors; p. 190–194.
- Brewer, J. *IEEE Transactions on Circuits and Systems cas-*. Vol. 25. 1978. Kronecker products and matrix calculus in system theory; p. 772–781.
- Candes EJ, Tao T. Decoding by linear programming. *IEEE Transactions on Information Theory* 2005;51(12):4203–4215.
- Cottareau B, Jerbi K, Baillet S. Multiresolution imaging of meg cortical sources using an explicit piecewise model. *NeuroImage* 2007;38(3):439–451. [PubMed: 17889564]
- Culham JC, Cavina-Pratesi C, Singhal A. The role of parietal cortex in visuomotor control: What have we learned from neuroimaging? *Neuropsychologia* 2006;44(13):2668–2684. [PubMed: 16337974]
- Dalal S, Guggisberg A, Edwards E, Sekihara K, Findlay A, Canolty R, Berger M, Knight R, Barbaro N, Kirsch H, Nagarajan S. Five-dimensional neuroimaging: Localization of the time-frequency dynamics of cortical activity. *NeuroImage* 2008;40(4):1686–1700. [PubMed: 18356081]
- Daunizeau J, Friston K. A mesostate-space model for EEG and MEG. *NeuroImage* 2007;38(1):67–81. [PubMed: 17761440]
- Daunizeau J, Mattout J, Clonda D, Goulard B, Benali H, Lina J. Bayesian spatio-temporal approach for eeg source reconstruction: Conciliating ecd and distributed models. *IEEE Transactions on Biomedical Engineering* 2006;53(3):503–516. [PubMed: 16532777]
- Demiralp T, Ademoglu A, Istefanopulos Y, Gülçür H. Analysis of event-related potentials (erp) by damped sinusoids. *Biological Cybernetics* 1998;78:487–493. [PubMed: 9711822]
- Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society B* 1977;39:1–38.
- Ding, L.; He, B. *Proceedings of NFSI & ICFBI*. Oct. 2007 Sparse source imaging in EEG.
- Dogandži´acacute A, Nehorai A. Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays. *IEEE Trans Signal Processing* 2000;48(1):13–25.
- Donoho D. De-noising by soft-thresholding. *IEEE Transactions on Information Theory* 1995;41(3):613–627.

- Donoho D. For most large underdetermined systems of linear equations, the minimal ℓ_1 norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics* 2006;59(6):797–829.
- Effern A, Lehnertz K, Fernández G, Grunwald T, David P, Elger C. Single trial analysis of event related potentials: non-linear de-noising with wavelets. *Clinical Neurophysiology* 2000;111:2255–2263. [PubMed: 11090779]
- Figueiredo M, Nowak R. An em algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing* 2003;12(8):906–916. [PubMed: 18237964]
- Gellman, A.; Carlin, J.; Stern, H.; Rubin, D. *Bayesian Data Analysis*. 2nd. CRC Press; Boca Raton, FL: 2003.
- Gorodnitsky, I.; Rao, B. *IEEE Transactions on Signal Processing*. Vol. 45. 1997. Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm.
- Huang M, Dale A, Song T, Halgren E, Harrington D, Podgorny I, Canive J, Lewis S, Lee R. Vector-based spatial-temporal minimum ℓ_1 -norm solution for meg. *NeuroImage* 2006;31:1025–1037. [PubMed: 16542857]
- Jeffs B, Leahy R, Singh M. An evaluation of methods for neuromagnetic image reconstruction. *IEEE Transactions on Biomedical Engineering* 1987;34:713–723. [PubMed: 3653912]
- Limpiti T, Van Veen B, Wakai R. Cortical patch basis model for spatially extended neural activity. *IEEE Transactions on Biomedical Engineering* 2006;53(9):1740–1754. [PubMed: 16941830]
- Malioutov D, Çetin M, Wilsky A. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Transactions on Signal Processing* 2005;53(8):3010–3022.
- Matsuura K, Okabe Y. Selective minimum-norm solutions of the biomagnetic inverse problem. *IEEE Transactions on Biomedical Engineering* 1995;42(6):608–615. [PubMed: 7790017]
- Matsuura K, Okabe Y. A robust reconstruction of sparse biomagnetic sources. *IEEE Transactions on Biomedical Engineering* 1997;44(8):720–726. [PubMed: 9254985]
- Mosher, J.; Leahy, R.; Shattuck, D.; Baillet, S. *Lecture Notes in Computer Science Vol 1613 Proc IPMI99*. Springer-Verlag GmbH; 1999. MEG source imaging using multipolar expansions.
- Mosher, J.; Lewis, P.; Leahy, R. *IEEE Trans Biomed Eng*. Vol. 39. 1992. Multiple dipole modeling and localization from spatio-temporal MEG data; p. 541-557.
- Nolte, G.; Curio, G. *IEEE Trans Biomed Eng*. Vol. 47. 2000. Current multipole expansion to estimate lateral extent of neuronal activity: A theoretical analysis; p. 1347-1355.
- Ou W, Hämäläinen M, Golland P. A distributed spatio-temporal eeg/meg inverse solver. *NeuroImage* 2009;44(3)
- Pascual-Marqui RD, Michel CM, Lehmann D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology* 1994;18:49–65. [PubMed: 7876038]
- Rockafellar, RT. *Convex Analysis*. Princeton University Press; Princeton, New Jersey: 1970.
- Scherg M, von Cramon D. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalogr Clin Neurophysiol* 1985;62:32–44. [PubMed: 2578376]
- Trejo L, Shensa M. Feature extraction of event-related potentials using wavelets: An application to human performance monitoring. *Brain and Language* 1999;66:89–107. [PubMed: 10080866]
- Tropp J. Algorithms for simultaneous sparse approximation, part ii: Convex relaxation. *Signal Processing* 2006;86:589–602.
- Trujillo-Barreto N, Aubert-Vázquez, Penny W. Bayesian m/eeg source reconstruction with spatio-temporal priors. *NeuroImage* 2008;39(1):318–335. [PubMed: 17904869]
- Turlach B, Venables W, Wright S. Simultaneous variable selection. *Technometrics* 2005;47(3):349–363.
- Uutela K, Hämäläinen M, Somersalo E. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* 1999;10:173–180. [PubMed: 10417249]
- Van Veen, B.; Van Drongelen, W.; Yuchtman, M.; Suzuki, A. *IEEE Trans Biomed Eng*. Vol. 44. 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering; p. 867-880.

- Wagner, M.; Fuchs, M.; Kastner, J. Proceedings of the 13th International Conference on Biomagnetism. Jena; Germany: Aug. 2002 Current density reconstructions and deviation scans using extended sources.
- Wagner, M.; Köhler, T.; Fuchs, M.; Kastner, J. An extended source model for current density reconstructions. In: Nenonen, J.; Ilmoniemi, R.; Katila, T., editors. Proceedings of the 12th International Conference on Biomagnetism; Espoo, Finland: Espoo: Helsinki Univ of Technology; Aug. 2000 p. 749-752.
- Wright, S.; Nowak, R.; Figueiredo, M. Sparse reconstruction by separable approximation. IEEE International Conference on Acoustics, Speech, and Signal Processing; Las Vegas, NV. 2008.
- Wu C. On the convergence properties of the em algorithm. *Annals of Statistics* 1983;11:95–103.
- Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society B* 2006;68:49–67.

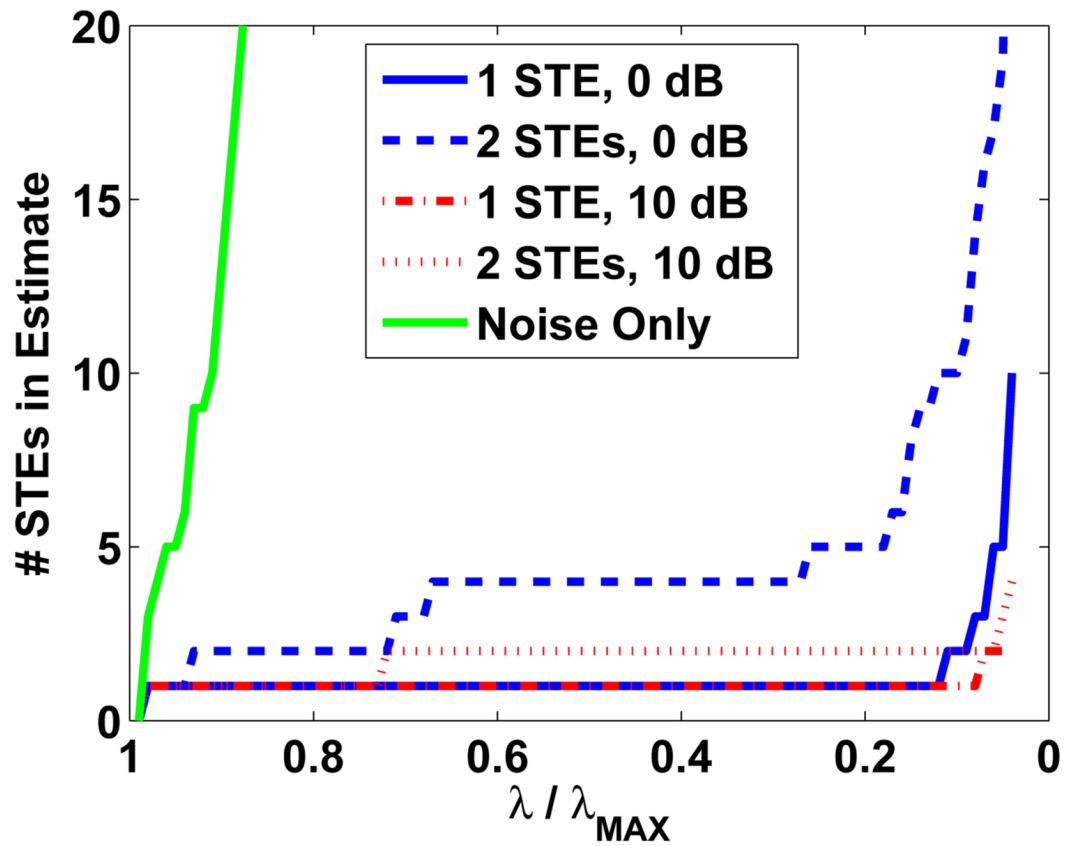


Fig. 1. Characteristic behavior of the number of STEs identified by ESP vs. λ / λ_{MAX} . The presence of source activity results in a plateau in the number of STEs selected as λ / λ_{MAX} decreases.

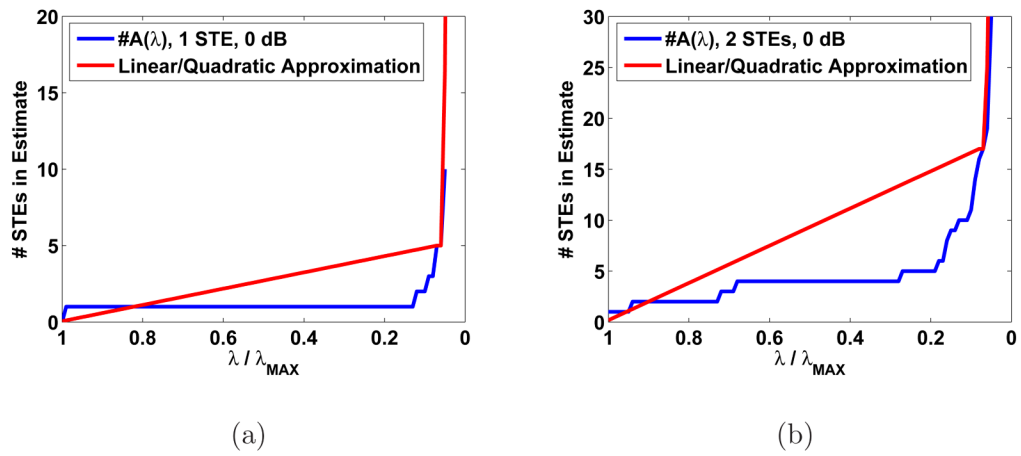


Fig. 2.
Example of heuristic applied to 0 dB simulated data with (a) 1 STE and (b) 2 STEs.

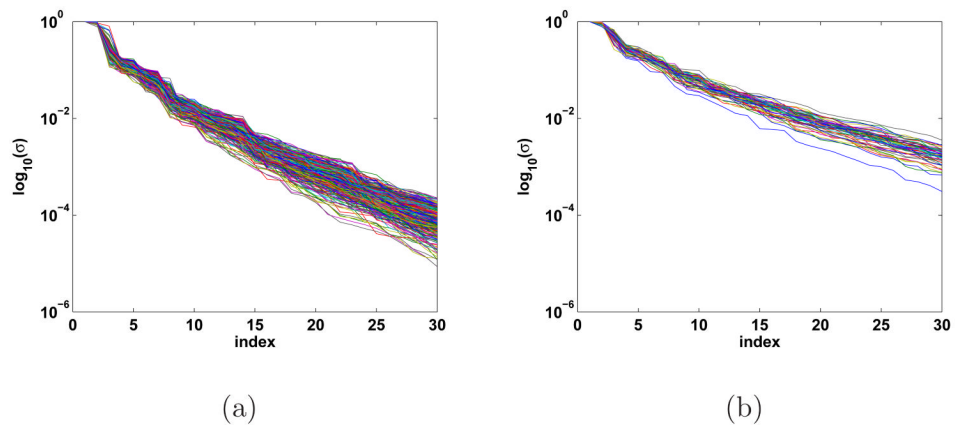


Fig. 3. Normalized singular values associated with patch lead field matrices from (a) 393 patches of 20 mm geodesic radius and (b) 57 patches of 50 mm geodesic radius which span the cortical surface with 50% overlap. The singular values decay rapidly in every patch.

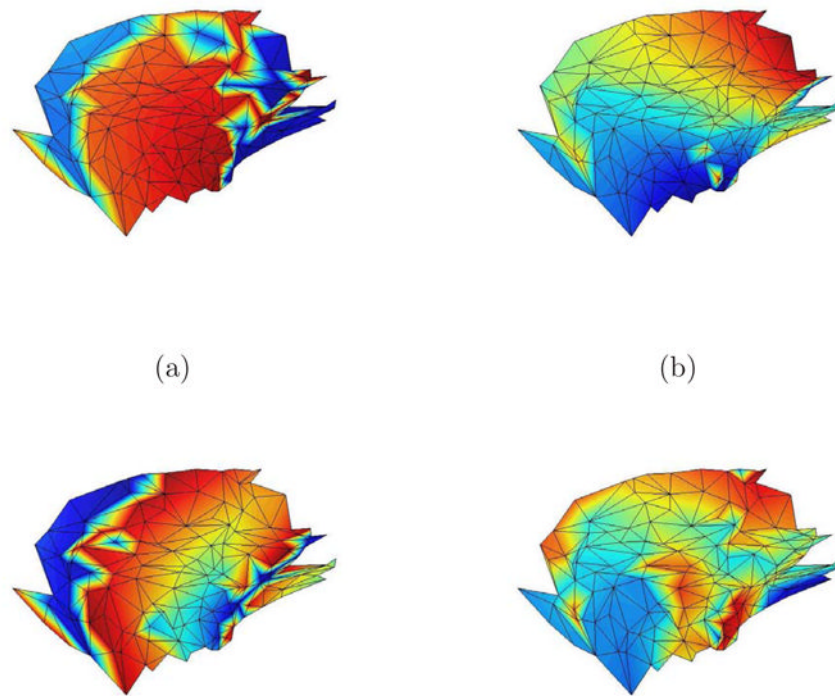


Fig. 4. Color cortical activity pattern for singular vectors \mathbf{v}_k of a representative patch. The vertices in the mesh represent dipole locations in the cortical tessellation of the patch. (a) \mathbf{v}_1 , largest singular value. (b) \mathbf{v}_3 , third largest singular value. (c) \mathbf{v}_5 , fifth largest singular value. (d) \mathbf{v}_7 , seventh largest singular value. The activity patterns associated with larger singular values tend to represent low resolution features within the patch, while those associated with smaller singular values represent fine structure.

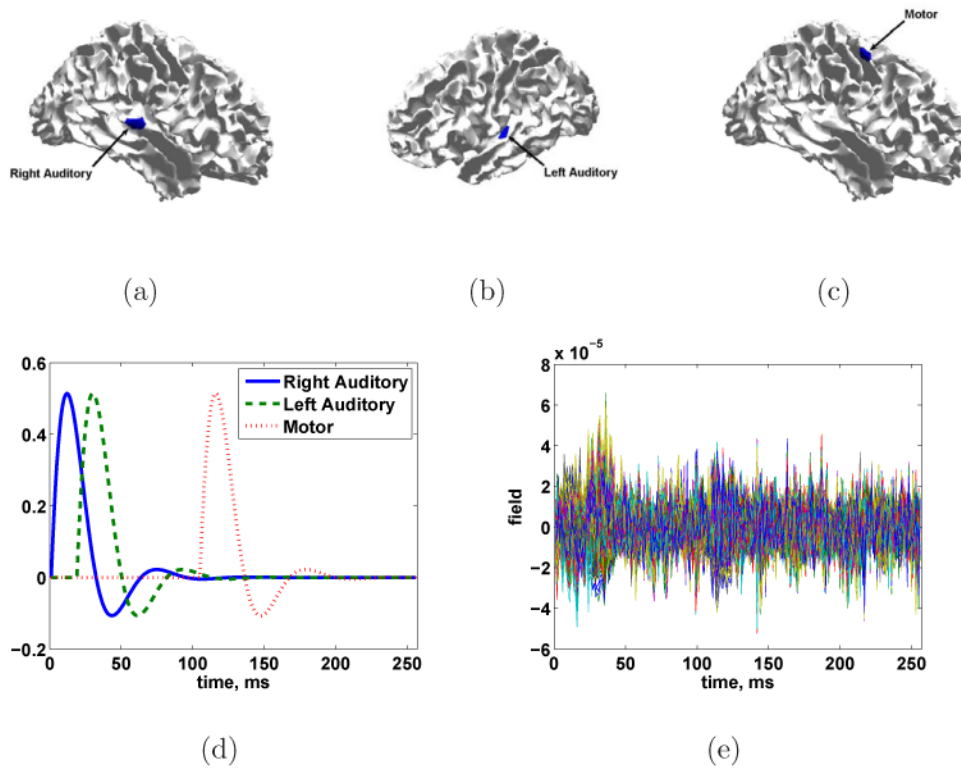


Fig. 5. Simulated three source scenario. True source location and spatial extent of sources at times (a) 12, (b) 30, and (c) 116 ms. (d) True source time courses. (e) Simulated signals in noise for SNR = -6 dB.

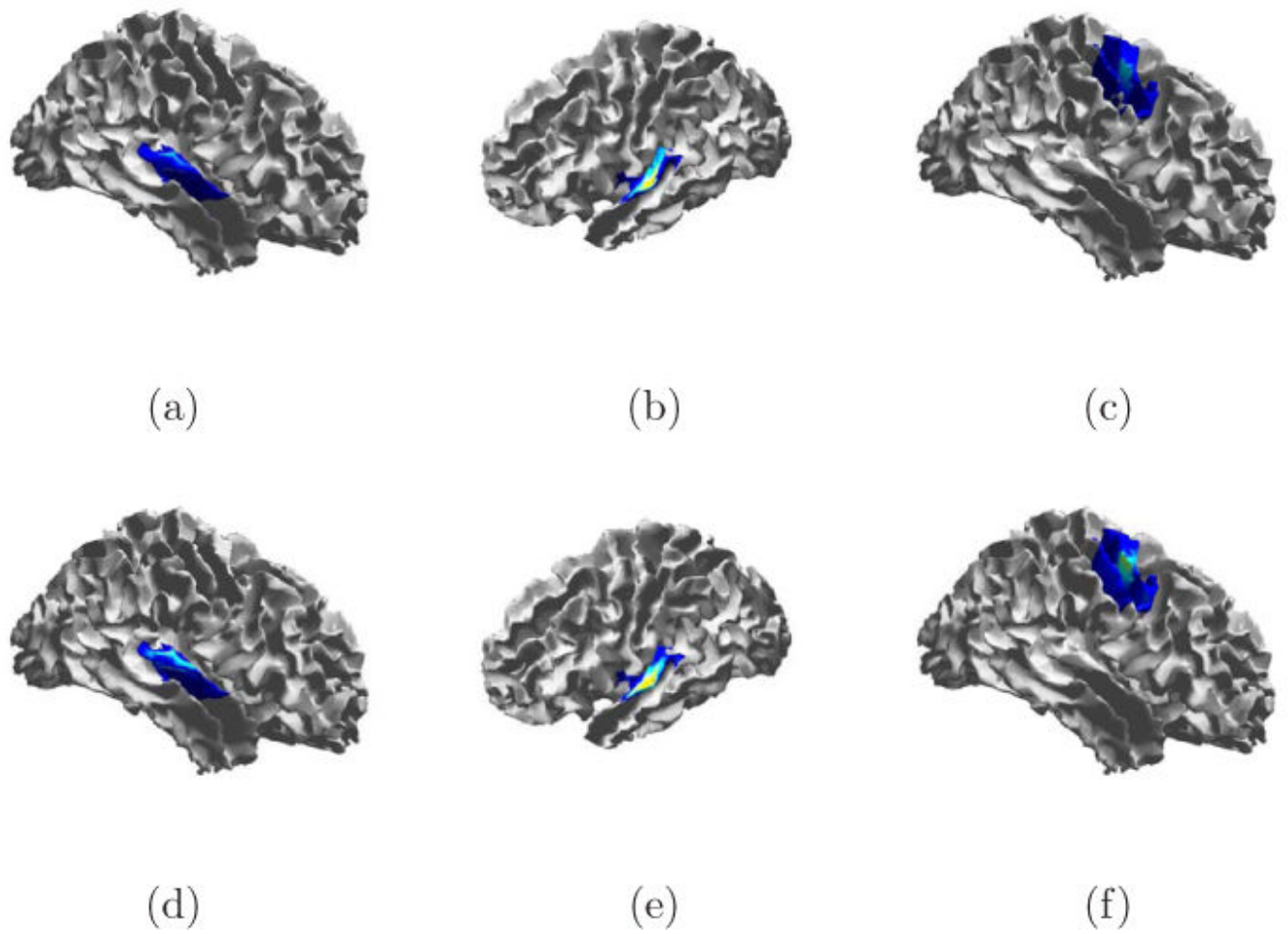
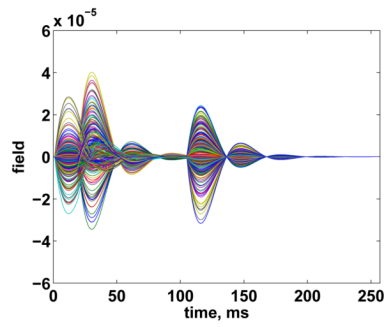
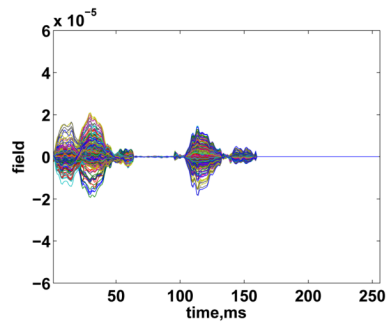


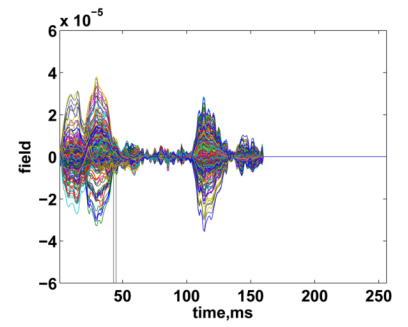
Fig. 6. Cortical signal estimate from ESP before (a – c) and after (d – f) debiasing. Estimates shown for peak signal times: (a,d) 12 ms, (b,e) 30 ms, and (c,f) 116 ms.



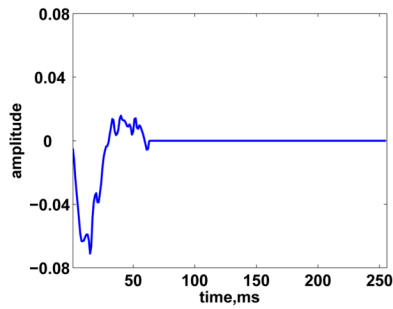
(a)



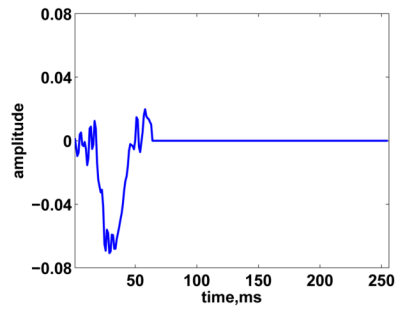
(b)



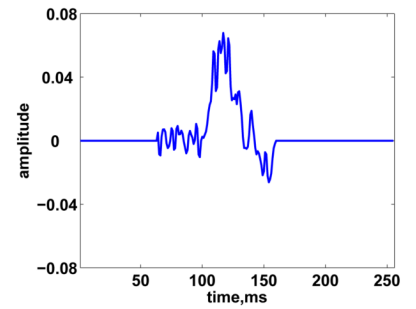
(c)



(d)



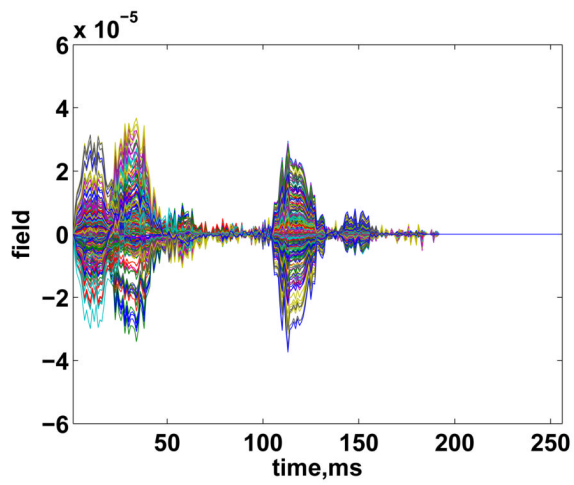
(e)



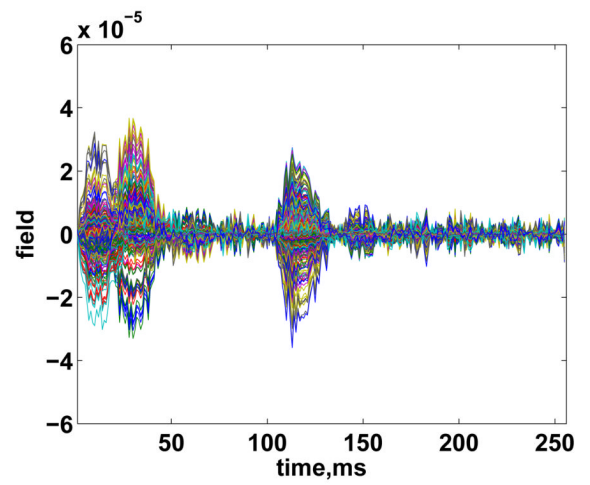
(f)

Fig. 7.

(a) Noise free simulated data (b) ESP reconstruction of sensor measurements before debiasing (c) ESP reconstruction of sensor measurements after debiasing. Estimated temporal evolution of: (d) a single dipole in the right auditory cortex, (e) a single dipole in the left auditory cortex, and (f) a single dipole in the left motor cortex.



(a)



(b)

Fig. 8.

(a) Reconstructed sensor measurements using multiple time windows and SVD temporal basis functions. (b) Reconstructed sensor measurements using single time window and SVD temporal basis functions.

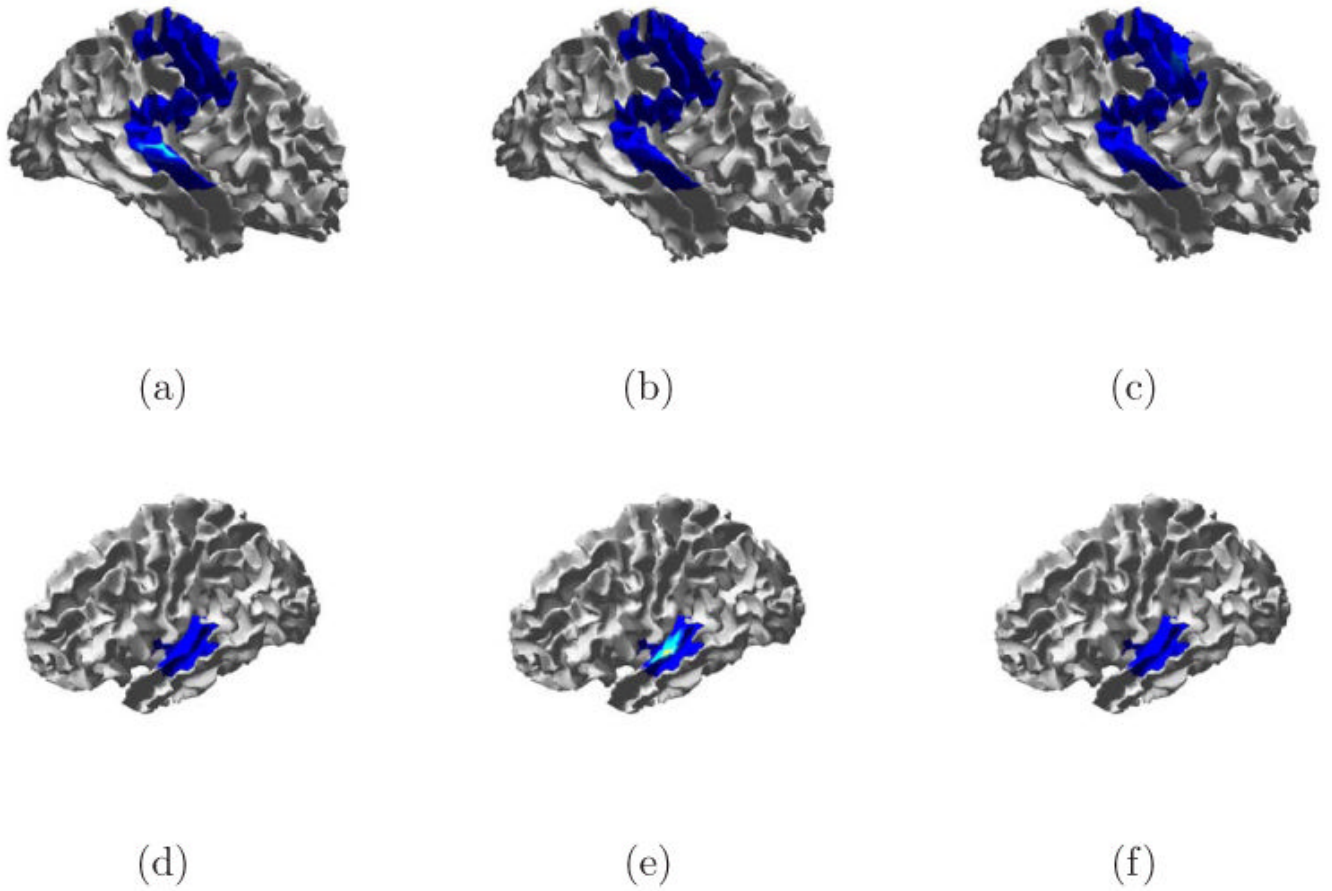


Fig. 9.
Reconstructed cortical signal using a single temporal block at time: (a,d) 12 ms (b,e) 30 ms
(c,f) 116 ms.

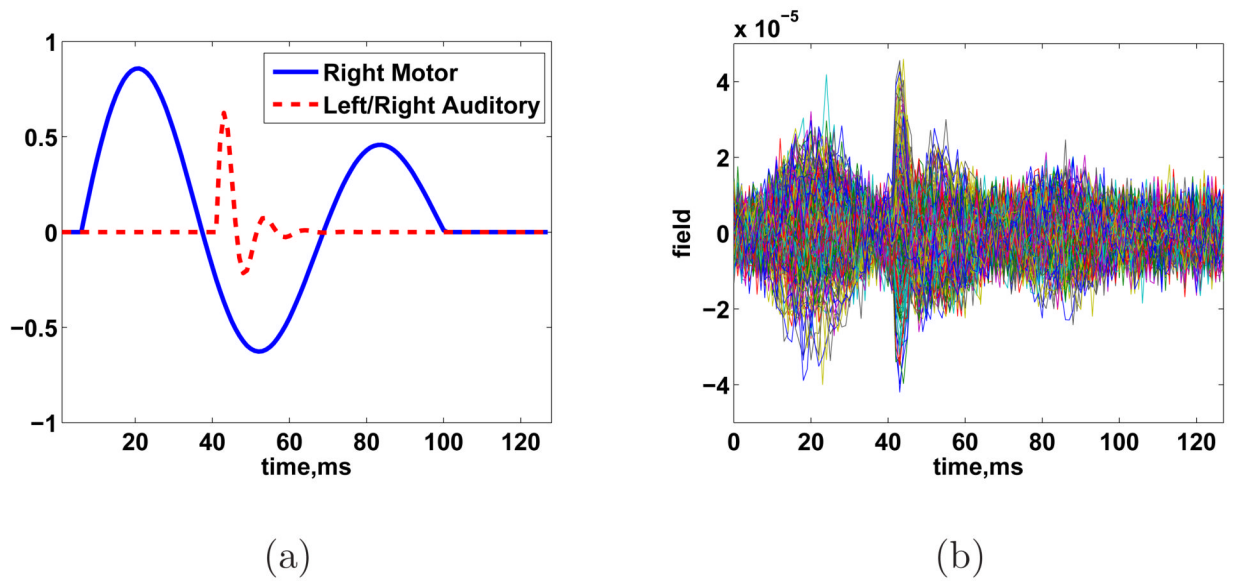


Fig. 10. (a) Temporal evolution of true activity. (b) Simulated noisy sensor measurements.

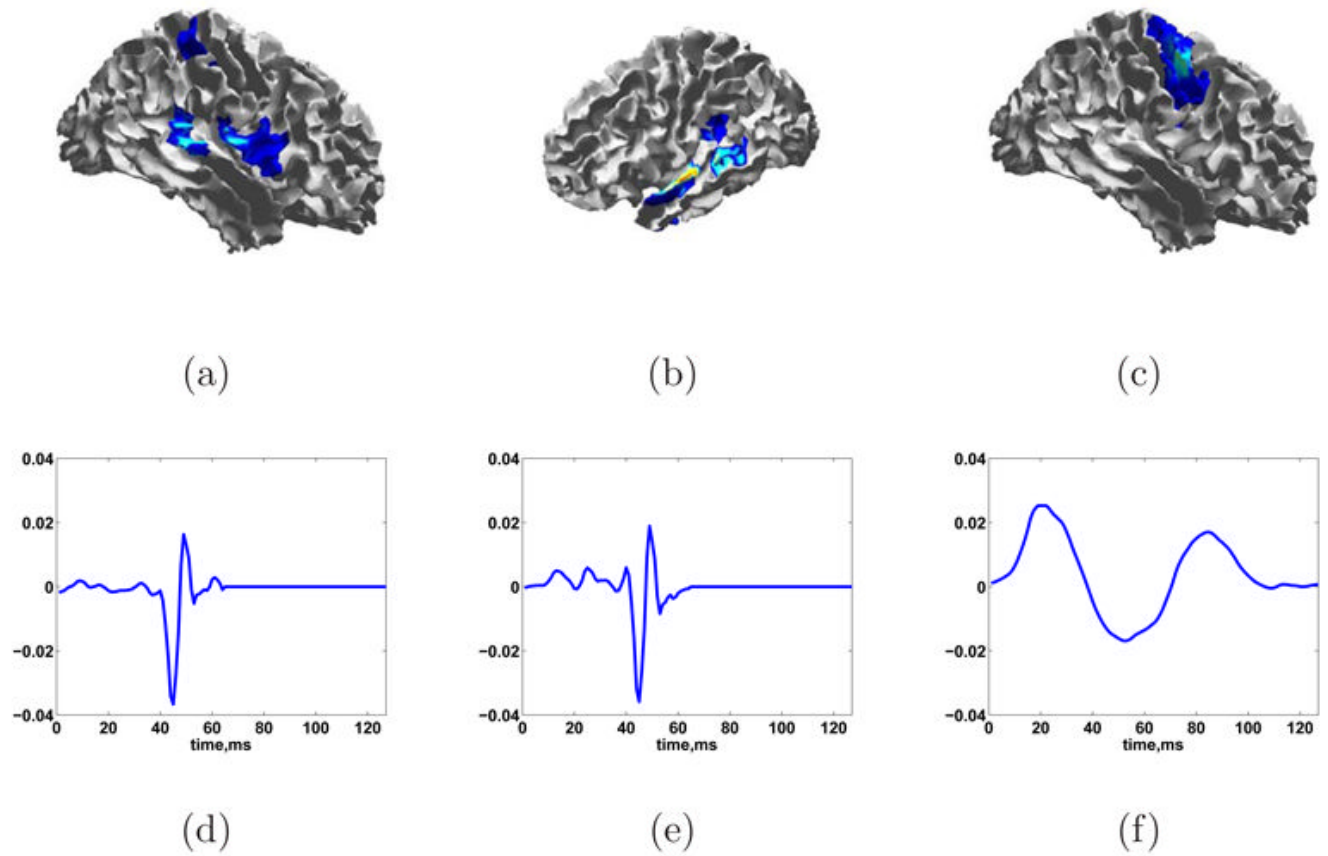


Fig. 11. Reconstructed cortical signal using temporal bases spanning multiple scales. Cortical activation and temporal evolution in the: (a,d) right auditory cortex, (b,e) left auditory cortex, and (c,f) right motor cortex.

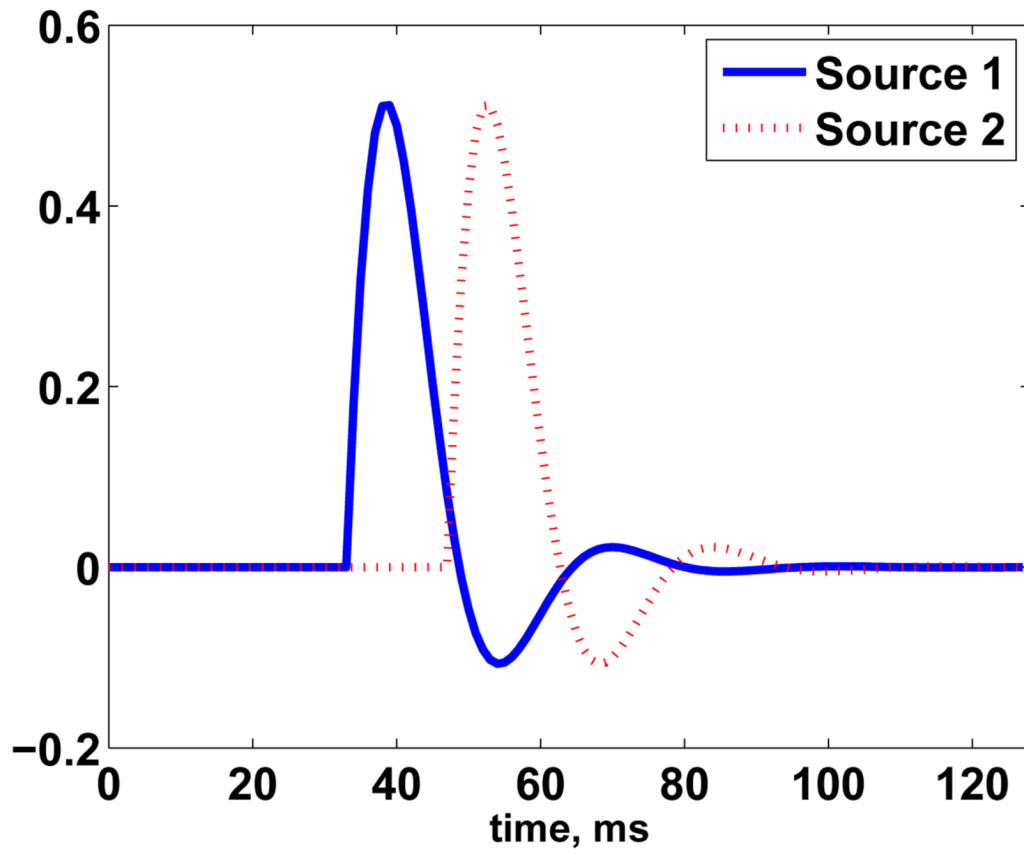


Fig. 12.
Example time courses (Trial 6) of cortical activity for simulations of Table 1.

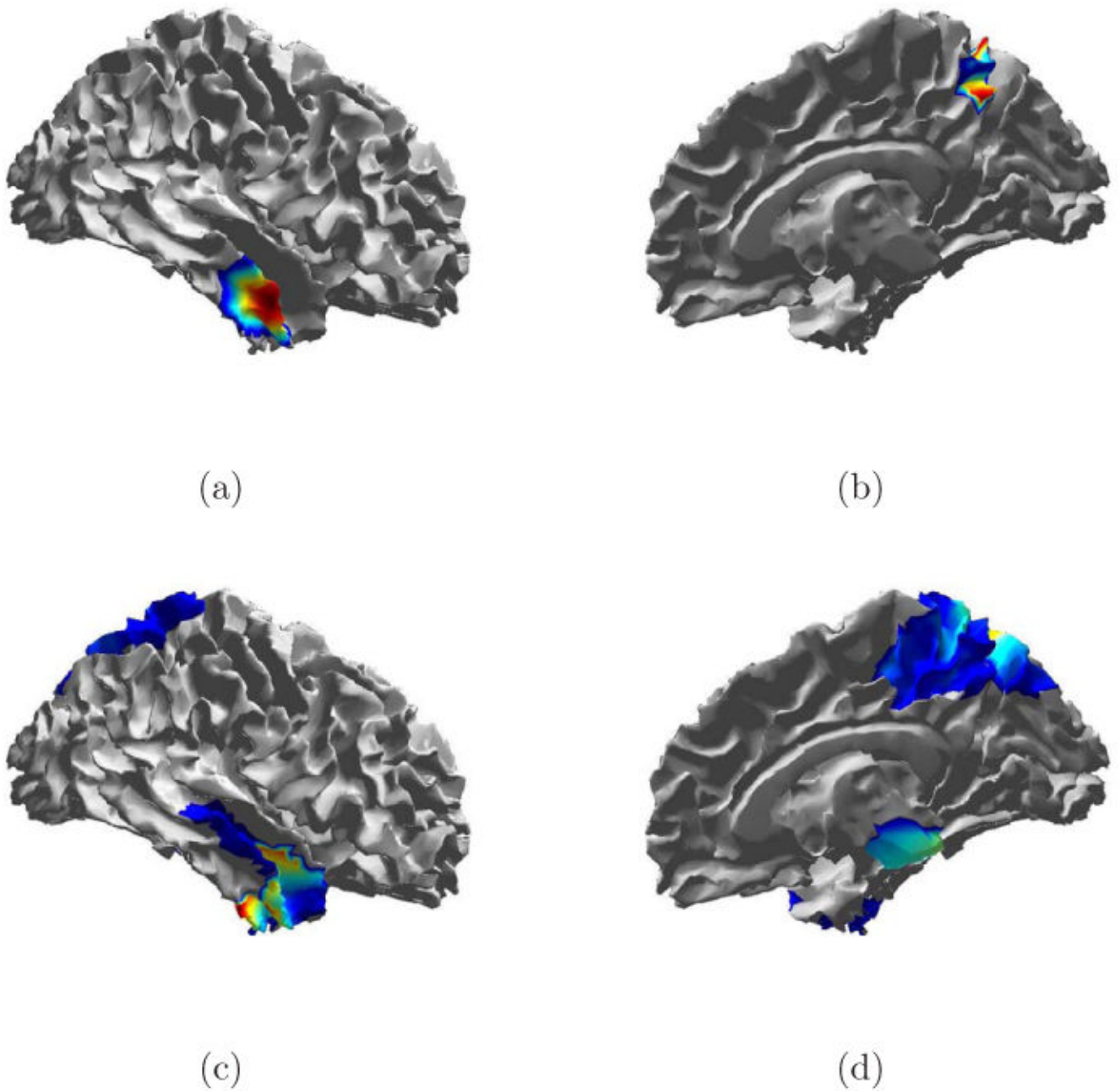


Fig. 13. True signal and debiased ESP reconstruction of Trial 6 from Table 1. (a) Source 1 spatial extent. (b) Source 2 spatial extent. (c) Debiased ESP reconstruction of source 1 at 41 ms. (e) Debiased ESP reconstruction of source 2 at 55 ms.

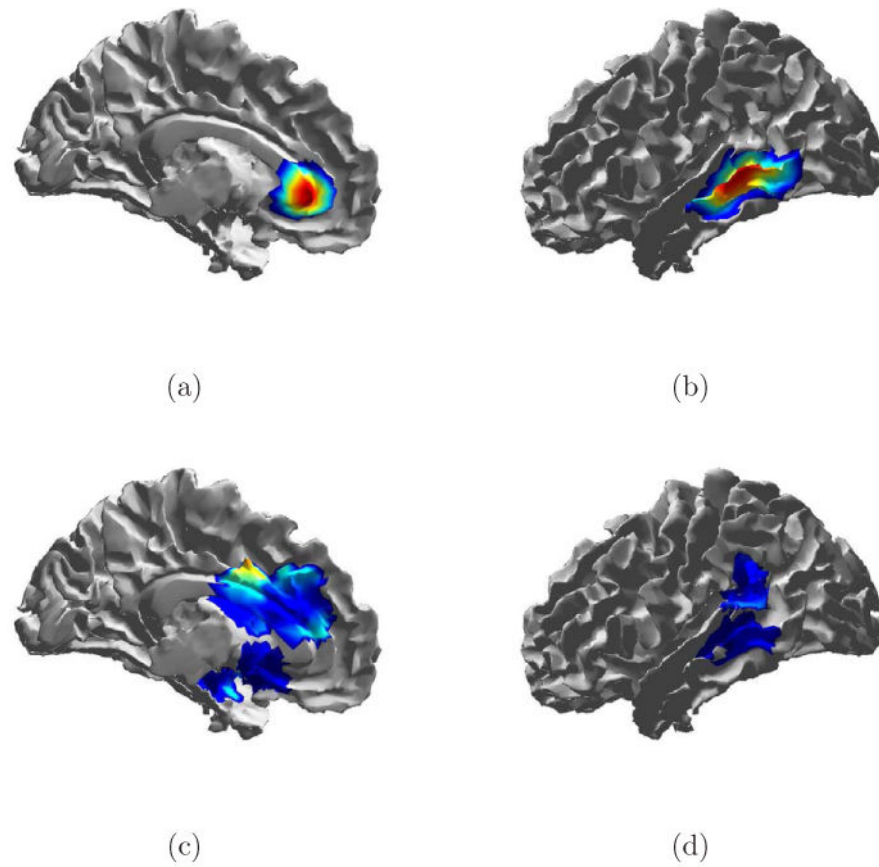


Fig. 14. True signal and debiased ESP reconstruction of Trial 8 from Table 1. (a) Source 1 spatial extent. (b) Source 2 spatial extent. (c) Debiased ESP reconstruction of source 1 at 66 ms. (e) Debiased ESP reconstruction of source 2 at 37 ms.

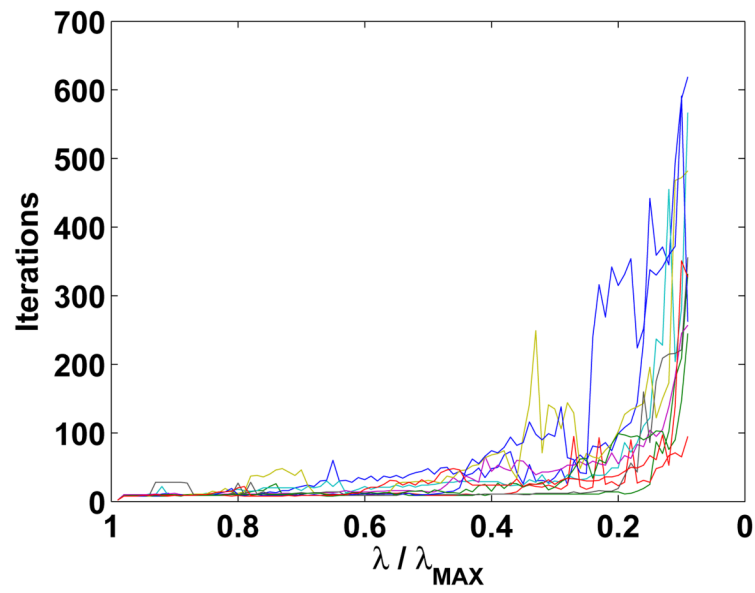
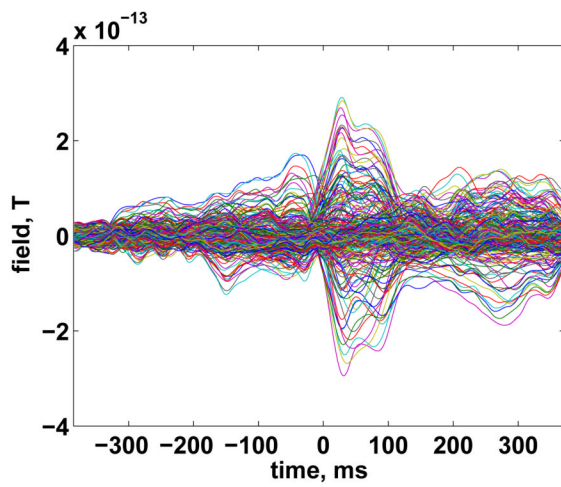
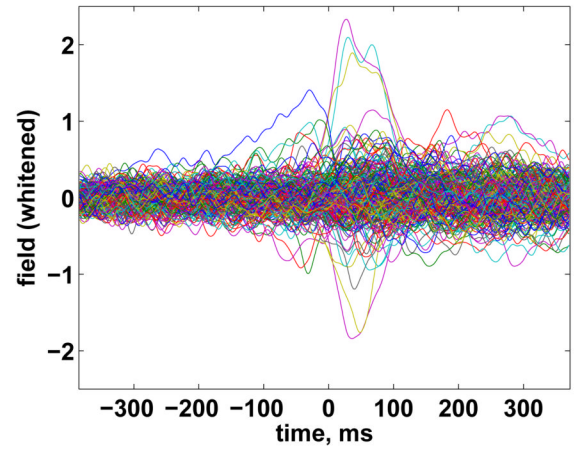


Fig. 15. Number of EM algorithm iterations required for each λ for all ten trials of Example 2 (Table 1). Solutions are calculated beginning with $\lambda = \lambda_{MAX}$ as described in Sec. 2.4.



(a)



(b)

Fig. 16. Recorded MEG from self-paced finger tapping. (a) Averaged data. (b) Whitened average data.

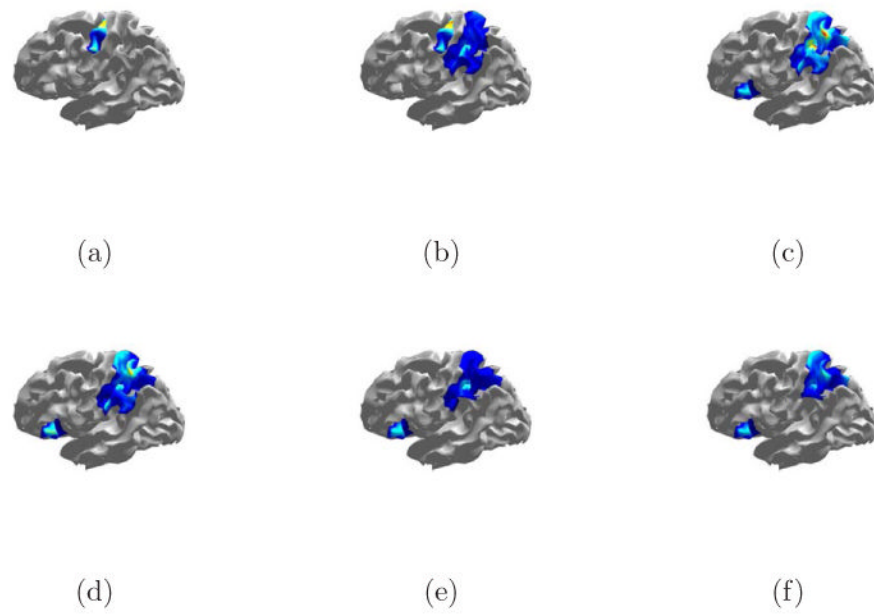


Fig. 17. Debiased ESP reconstruction of self-paced finger tapping with heuristic λ . (a) -105 ms. (b) -31 ms. (c) 28 ms. (d) 80 ms. (e) 176 ms. (f) 264 ms.

Table 1

Randomized source scenarios and patch center distance errors for Example 2.

Trial	Source	Hemisphere	Onset (ms)	Geo. Radius (mm)	Source SNR (dB)	Distance Error (mm)
1	1	Right	67	27.4	-2.61	12.5
	2	Left	57	23.6	-3.49	6.7
2	1	Right	14	32.8	-0.20	7.1
	2	Right	53	23.7	-13.57	19.3
3	1	Right	31	36.3	-0.07	4.0
	2	Right	16	16.7	-17.09	miss
4	1	Left	50	9.7	-1.27	9.7
	2	Right	23	15.3	-5.98	15.3
5	1	Right	42	22.5	-0.73	9.3
	2	Left	28	9.2	-8.02	5.6
6	1	Right	34	19.6	-4.55	8.1
	2	Right	48	15.4	-1.60	6.9
7	1	Left	48	33.8	-0.00	5.6
	2	Right	8	8.3	-29.67	miss
8	1	Left	14	18.1	-9.52	20.6
	2	Left	30	31.6	-0.52	10.4
9	1	Left	3	18.9	-9.48	16.0
	2	Left	41	31.5	-0.48	4.8
10	1	Right	73	12.2	-8.51	10.9
	2	Right	37	23.1	-0.67	8.8