Combined MEG and EEG Source Imaging by Minimization of Mutual Information

Sylvain Baillet,* Line Garnero, Gildas Marin, and Jean-Paul Hugonin

Abstract— **Though very frequently assumed, the necessity to operate a joint processing of simultaneous magnetoencephalography (MEG) and electroencephalography (EEG) recordings for functional brain imaging has never been clearly demonstrated. However, the very last generation of MEG instruments allows the simultaneous recording of brain magnetic fields and electrical potentials on the scalp. But the general fear regarding the fusion between MEG and EEG data is that the drawbacks from one modality will systematically spoil the performances of the other one without any consequent improvement. This is the case for instance for the estimation of deeper or radial sources with MEG.**

In this paper, we propose a method for a cooperative processing of MEG and EEG in a distributed source model. First, the evaluation of the respective performances of each modality for the estimation of every dipole in the source pattern is made using a conditional entropy criterion. Then, the algorithm operates a preprocessing of the MEG and EEG gain matrices which minimizes the mutual information between these two transfer functions, by a selective weighting of the MEG and EEG lead fields. This new combined EEG/MEG modality brings major improvements to the localization of active sources, together with reduced sensitivity to perturbations on data.

Index Terms— **Brain functional imaging, distributed sources models, electroencephalography (EEG), entropy quantification, fusion of data, magnetoencephalography (MEG), mutual information, source reconstruction.**

I. Introduction

GIVEN a model of the electromagnetic and geometrical properties of the head tissues (basically the cortex, cerebrospinal fluid (CSF), skull, and scalp), magnetoencephalography (MEG) and electroencephalography (EEG) source imaging consists in finding the current dipole(s) that best explain(s) the data in accordance with the experimental protocol.

But as time resolution is the strongest asset of EEG/MEG brain functional imaging, it has been commonly assumed that both of these two modalities offer poor localization performances when no constraints are enforced within the source estimation procedure. EEG is considered as highly sensitive to the geometry and the conductivity properties of the head tissues which is not as much the case for magnetic fields—though intracranial inhomogeneity such as ventricular cavities may have strong influence on the field pattern outside the head [39]. MEG, however, fails to locate deeper cortical sources because of the rapid fall of the magnetic field with depth [12]. Numerous experiments for interictal spike source localization with MEG and EEG illustrate this point [43]–[46]. Furthermore, as head shapes are still extensively modeled with single or multiple concentric spheres, MEG is unable to locate so-called radial sources, whereas EEG seems sensitive to every source orientation. All of these points tend to confirm the general belief that a cooperative processing of EEG and MEG signals would improve the solution to the source estimation problem [7], [19], [24]. Unfortunately, few studies are dedicated to the systematic validation of this assumption. The very recent appearing of devices that offer simultaneous MEG and EEG recordings with large arrays of detectors may explain the relatively sparse "fusion" literature.

Wood *et al.* [29] were pioneers in this field with simultaneous EEG and single-channel MEG recordings. Sutherling *et al.* have also presented interesting points about a joint analysis of MEG and EEG data [23]. However, the study was done with single dipole models, and the "fusion" solution was computed from the average of MEG's and EEG's. These first works did not introduce any joint processing of a common data set made of both EEG and MEG. Stok *et al.* have shown in [22] that source estimation with 21 EEG channels and 21 MEG's produces "slightly better" results than with solely 21 EEG's or 21 MEG's. However this result can be mainly attributed to the increase of the number of sensors (from 21 to 42 on the whole). Dale and Sereno [5] and, more recently, Phillips [18] have confirmed these primary results, with distributed source models on realistic cortical anatomy. These simulation studies demonstrate the improvement of the spatial accuracy of the reconstruction methods when MEG and EEG data are gathered in a global data set (122 MEG magnetometers and 133 EEG electrodes in [18]). Very separate works by Pflieger *et al.* on the quantification of the information content of simultaneous EEG and MEG recordings describe them as "super-additive" in comparison to original EEG or MEG data [17]. This is another positive result in favor of the idea to operate a joint treatment of the two raw data sets for solving a common inverse problem.

Manuscript received July 8, 1997; revised October 22, 1998. *Asterisk indicates corresponding author*.

*S. Baillet is with Groupe Imagerie Cérébrale, Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, Hôpital de la Salpêtrière, CNRS—Université Paris VI, Paris, France. He is also with the University of Southern California, Signal and Image Processing Institute, Los Angeles, CA 90089-2564 USA (e-mail: silvin@sipi.usc.edu).

L. Garnero is with Groupe Imagerie Cérébrale, Laboratoire de Neurosciences Cognitives et Imagerie Cérébrale, Hôpital de la Salpêtrière, CNRS—Université Paris VI, Paris, France.

G. Marin and J.-P. Hugonin are with the Institut d'Optique Théorique et Appliquée, CNRS—Université de Paris-Sud XI, Faculté d'Orsay, Orsay, France.

Publisher Item Identifier S 0018-9294(99)03115-8.

The method described hereinafter is an attempt to go further than the simple combination of MEG and EEG data and gain matrices. The main question we want to address is the following: What kind of *optimization* in the combination of simultaneous MEG and EEG signals can be applied to improve the localization of the brain sources underneath?

The basic idea that will be developed here consists in preprocessing every lead field associated with a given distributed source pattern so as to minimize the redundancy of information between the MEG and EEG models. As a gain matrix can be considered as a transfer operator between the neural sources and the sensors, we then propose a method that selectively weights every lead field of each of the modality by minimization of mutual information (MI) between these two transfer functions. This procedure goes further than the global weighting technique which consists in equilibrating the rows and the columns of the MEG-EEG matrix to balance both the sensitivity of the sensors and the dipole contribution to the sensor array ([10], [16]). The latter will be hereinafter denoted as "raw" fusion.

The present article is divided as follows. The presentation of the MEG-EEG combination method by minimization of mutual information (CMMI) is in Section II. In Section III, the practical implementation of the method is discussed first in the general context of source estimation with regularization. Then as an example, we will stress on the adaptation of the nonlinear source estimator we developed for the inverse problem in a distributed source model: the ST-MAP method [1]. Simulation studies in Section IV use various source configurations with both spherical and real head models to discuss extensively the questions above for the validation of the new method proposed here.

II. Combination of MEG and EEG by Minimization of Mutual Information

A. Introduction

MEG/EEG source estimation may be seen as a true imaging problem. Indeed, it is possible to distribute some dipoles through the whole brain volume at fixed locations and to estimate their magnitudes according to data and some regularization priors—like the constraint of smoothness on the intensity gradient or Laplacian [16]. To go further into the refinement of the source model, dipoles may be constrained on the surface of the cortex—maybe on some restricted cortical regions—with orientations perpendicular to the cortical surface [4], [5], [18], [26]. This model descends from the fact that neurons in cortical layers are organized in macro columns oriented perpendicularly to the surface. As source orientations and locations are fixed, the only remaining degrees of freedom are dipole magnitudes.

We will see in Section III how a distributed source model can be well suited to take various *a priori* information into account that contributes to the estimation of the activity of multiple cortical areas.

If N dipole amplitudes—gathered in vector $\mathbf{J}$—are to be estimated from $\mathbf{M}$, a vector of size N_M, it is possible to connect sources to the sensors owing to the following system of equations:

$$\mathbf{M} = \mathbf{G}\,\mathbf{J} + \mathbf{b} \tag{1}$$

where $\mathbf{G}$ stands for the so-called gain matrix which columns are the MEG or EEG lead fields associated with each source. It contains global information about the head model, the source pattern (source orientations and positions), and sensor positions; $\mathbf{b}$ is a perturbation vector that stands for perturbations.

Operating fusion between EEG and MEG data consists in gathering them in a unique set of measurements, so as to solve a single inverse problem for simultaneous electromagnetic recordings. In a distributed source model with a single fusion gain matrix, the main issue here consists in avoiding any pollution of a given modality by the other one. For instance let us consider the case of radial dipoles in a spherical head model; or even with realistic models, sources pointing toward the head surface still produce small magnetic fields outside the head. Deeper sources in both spherical and realistic geometry tend to be MEG silent in the sense that they weakly contribute to data. Thus, their corresponding lead fields have much smaller norms than superficial or tangential sources. Further, if several of these sources are distributed along the cortical surface, as they have similar smeared contributions across the sensors, their lead fields are almost collinear. Therefore, estimating their amplitudes with MEG only is a very ill posed problem. Then, the source estimation algorithm has to rely on the EEG data only, i.e., on measurements with also a very smooth spatial pattern on the scalp. Practically, combining all of these lead-fields in a global linear operator generates a gain matrix that is ill conditioned [31]. Consequently, this contributes to the ill-posed nature of the MEG-EEG fusion inverse problem.

Ill conditioning of the linear operator generates solutions to the associated linear system, which are very sensitive to roundoff errors during computation and to small perturbation on data. Row and column equilibration of the linear operator is the simplest way to cure to badly scaled systems. We first recall that the condition number of a matrix according to the two-norm, is the ratio of its biggest singular value to its smallest [31]. It can be shown that row or column normalization is a way to approach the minimum value for the condition number of an operator when linearly operating on its rows or its columns. Actually, it can be shown that equalizing the two-norm of the rows of the $\mathbf{G}$ matrix produces a new linear operator which condition number is no more than a factor $\sqrt{N_M}$ away from the smallest condition number that can be achieved with a linear transformation of $\mathbf{G}$ [30]. A similar theorem stands also for column normalization. It is interesting to note that, though matrix equilibration has strong algebraic roots, the first motivations to do so in the MEEG inverse problem were based on physical considerations. The scaling of MEG and EEG data (like the normalization of the rows of the gain matrix, see [18] for instance), and the normalization of the lead fields were proposed to facilitate the recovering of deeper sources [16]. More generally, these *ad hoc* procedures based on physical considerations regarding the formation of data can

give correct starting points to investigate the enhancement of the condition of the linear operator as explained by Golub and Van Loan in [9]. The method we propose here is also driven by such physical considerations.

So the issue is the following: it seems very clear in principle that EEG and MEG are complementary, but how to *optimize* their combination as classical equilibration methods may not be sufficient for enhanced source estimation (see Section IV)?

Rather than a global lead field equilibration, we investigate the possibility to operate selective weighting of the contributions from either modality, which is twofold.

1) First for a given source, a preselection processing of the EEG and the MEG lead field will predict which of these two modalities is the most appropriate to estimate its amplitude.
2) Following this preselection, each of the least appropriate lead fields is then selectively attenuated. The associated weights are iteratively computed for optimization according to minimization of MI between the two weighted EEG and MEG gain matrices.

B. Combination of the Electromagnetic Data Set

According to (1), a common system can be written with both MEG and EEG

$$\begin{cases} \mathbf{B} = \mathbf{G}_B\mathbf{J} + \mathbf{b}_B \\ \mathbf{V} = \mathbf{G}_V\mathbf{J} + \mathbf{b}_V \end{cases} \tag{2}$$

$\mathbf{J}$ is the N-row vector of the dipole magnitudes of the distributed model; $\mathbf{B}$ (respectively, $\mathbf{V}$) is a N_B (respectively, N_V) row vector containing MEG (respectively, EEG) measurements for a given time sample. Finally, $\mathbf{G}_B$ (respectively, $\mathbf{G}_V$) is a $N_B \times N$ (respectively, $N_V \times N$) gain matrix. MEG and EEG data are corrupted by additive noises $\mathbf{b}_B$ and $\mathbf{b}_V$, respectively.

A preliminary phase consists in gathering the EEG and MEG data in a raw fusion data set $\mathbf{M}$ ($\mathbf{M} = \begin{bmatrix}\mathbf{B}\\ \mathbf{V}\end{bmatrix}$). Thus, and assuming for simplicity in notations noiseless measures, $\mathbf{M}$ is involved in the following linear system:

$$\mathbf{M} = \begin{bmatrix}\mathbf{G}_B \\ \mathbf{G}_V\end{bmatrix}\mathbf{J} = \mathbf{G}\,\mathbf{J}. \tag{3}$$

The method presented here consists in working on a global matrix $\hat{\mathbf{G}}$ descending from modifications of the original gain matrix $\mathbf{G}$, in the following way:

$$\hat{\mathbf{G}} = \begin{bmatrix}\mathbf{N}_B\mathbf{G}_B\mathbf{C}_B \\ \mathbf{N}_V\mathbf{G}_V\mathbf{C}_V\end{bmatrix} \cdot \mathbf{L}. \tag{4}$$

The new matrices that are introduced in this formulation are computed along the three following steps.

1) Equalization of EEG and MEG sensor responses: this is equivalent to the classical scaling of EEG and MEG data [18].
2) Local selective weighting of each MEG lead-field versus EEG for every dipole based on minimization of mutual information between $\mathbf{N}_B\,\mathbf{G}_B$ and $\mathbf{N}_V\,\mathbf{G}_V$. The motivation for the choice of the MI criterion, and the computation of the $\mathbf{C}_B$ and $\mathbf{C}_V$ matrices are discussed below.
3) Normalization of the resulting lead fields: this is a scaling step that contributes to the recovery of sources with weak lead fields norms [10], [16].

C. Normalization of Sensor Responses

In order to scale EEG and MEG, the rows of the MEG and EEG gain matrices are normalized. This is done owing to two diagonal matrices $\mathbf{N}_B$ and $\mathbf{N}_V$ which diagonal elements are: for $i \in [1, 2, \cdots, N_B]$

$$\mathbf{N}_B(i, i) = \|\mathbf{G}_B(i, :)\|^{-1}. \tag{5}$$

Where $\mathbf{G}_B(i, :)$ stands for the ith row of $\mathbf{G}_B$. An equivalent definition stands for $\mathbf{N}_V$. The data are then gathered in $\overline{\mathbf{M}} = \begin{bmatrix}\mathbf{N}_B \\ \mathbf{N}_V\end{bmatrix} \cdot \mathbf{M}$, and we note $\overline{\mathbf{G}}_B = \mathbf{N}_B\,\mathbf{G}_B$ and $\overline{\mathbf{G}}_V = \mathbf{N}_V\,\mathbf{G}_V$.

D. Local Selective Weighting of MEG and EEG Gain Vectors

We now introduce different weighting coefficients on each column of $\overline{\mathbf{G}}_B$ and $\overline{\mathbf{G}}_V$. If $\mathbf{C}_B$ and $\mathbf{C}_V$ are two $N \times N$ diagonal matrices (with real and positive diagonal components called $c_B(j)$ and $c_V(j)$, respectively, $j \in [1, 2, \cdots, N]$), and

$$\hat{\mathbf{G}} = \begin{bmatrix}\overline{\mathbf{G}}_B \cdot \mathbf{C}_B \\ \overline{\mathbf{G}}_V \cdot \mathbf{C}_V\end{bmatrix}. \tag{6}$$

The jth column of $\hat{\mathbf{G}}$ is

$$\hat{\mathbf{G}}(:, j) = \begin{bmatrix}\hat{\mathbf{G}}_B(:, j) \\ \hat{\mathbf{G}}_V(:, j)\end{bmatrix} = \begin{bmatrix} \left.\begin{matrix}\overline{\mathbf{G}}_B(1, j) \\ \overline{\mathbf{G}}_B(2, j) \\ \vdots \\ \vdots \\ \overline{\mathbf{G}}_B(N_B, j)\end{matrix}\right\} \times c_B(j) \\ \left.\begin{matrix}\overline{\mathbf{G}}_V(1, j) \\ \overline{\mathbf{G}}_V(2, j) \\ \vdots \\ \vdots \\ \overline{\mathbf{G}}_V(N_V, j)\end{matrix}\right\} \times c_V(j) \end{bmatrix}. \tag{7}$$

Thus, this writing allows to weight locally (i.e., for each source of the distributed model) the normalized contribution of MEG versus EEG. For instance, if $c_V(j) > c_B(j)$, the EEG data set will be considered as more reliable than MEG's for the estimation of the jth source (as for a radial source for example).

Now the issue consists in finding criteria that will drive this local selection via the determination of the $c_v(j)$ and $c_b(j)$ coefficients.

E. The Choice of Criteria Based on Conditional Entropy Quantification and MI Minimization

We first have to address the question of a strategy for the choice of the diagonal elements of $\mathbf{C}_B$ and $\mathbf{C}_V$ according to some constraints. The latter arise from a pre-evaluation of the respective merits of EEG and MEG to estimate a given source.

A possible criterion is the conditional entropy between the two original lead fields $\mathbf{B}j$ ($\mathbf{B}j = \overline{\mathbf{G}}_B(:, j)$) and $\mathrm{V}j$ ($\mathbf{V}j = \overline{\mathbf{G}}_V(:, j)$), $j \in [1 \cdots N]$. These vectors can be considered as two sets of realizations of two random variables. We call $E(\mathbf{B}j|\mathbf{V}j)$ and $E(\mathbf{V}j|\mathbf{B}j)$ the conditional entropy of $\mathbf{B}j$ given $\mathbf{V}j$, and of $\mathbf{V}j$ given $\mathbf{B}j$, respectively. The conditional entropy $E(\mathbf{B}j|\mathbf{V}j)$ is known to be a measure of the amount of uncertainty left in $\mathbf{B}j$ when knowing $\mathbf{V}j$ [36], [37]. This choice is motivated as low values for $E(\mathbf{B}j|\mathbf{V}j)$, for instance, are achieved when the $\mathbf{B}j$ lead field has very smeared contributions across sensors in comparison with $\mathbf{V}j$. This property fits very well to the discussion in Section I and is confirmed and illustrated by the studies in Section IV. Hence, for a given source j of the distributed model, we first estimate the respective conditional entropy of the associated MEG and EEG gain vectors, $E(\mathbf{B}j|\mathbf{V}j)$ and $E(\mathbf{V}j|\mathbf{B}j)$ (see the Appendix for practical implementation). The one with higher conditional entropy is designated as the best modality to evaluate the source amplitude.

This step is used to introduce some constraints on the c_b and c_v coefficients that are going to be computed in the optimization procedure described further below. As an example, if $E(\mathbf{B}j|\mathbf{V}j) > E(\mathbf{V}j|\mathbf{B}j)$, the global optimization procedure consists in setting $c_B(j) = 1$ and finding $c_V(j) < c_B(j)$ in [0, 1] according to the following global criterion on the modified gain matrices.

Following this preselection step, the basic idea consists in reducing the degree of redundancy or similarity between the two transfer functions. As discussed above, such a redundancy between several lead fields is in part responsible for the bad condition of the global operator. If we consider (gb, gv), a pair of elements of the MEG and EEG matrices as realizations of two random variables GB and GV, mutual information $I(GB, GV)$ is a measure of the degree of dependence between these two variables. $I(GB, GV)$ can also be seen as the amount of information that GB (respectively, GV) contains about GV [respectively, GB, as $I(GB, GV) = I(GV, GB)$]. Previous works have used MI maximization to estimate the best geometric transformation for multimodality image registration [37], [38]. Here, our goal is at the opposite of image registration. Hence, minimizing $I(GB, GV)$ by tuning some registration parameters [here the $c_V(j)$ and $c_B(j)$ coefficients] increases the global information content of the final fusion matrix owing to the reduction of redundancy between the two original modalities. Thus, we propose to scan all the lead fields ($\overline{\mathbf{G}}(:, j)$, $j = 1 \cdots N$) one after the other in order to find the best corresponding weights that will minimize MI between the modified matrices, $\hat{\mathbf{G}}_B$ and $\hat{\mathbf{G}}_V$ (see the Appendix for details of MI computation).

For clarity, the algorithm is summarized as follows:

1°) Initialization: C_B *and* C_V *are two* $N \times N$ *identity matrices.*
2°) For every source j
2-a) Estimation of conditional entropy:
$E(\mathbf{B}j|\mathbf{V}j)$ *and* $E(\mathbf{V}j|\mathbf{B}j)$
2-b) MI Minimization
If $E(\mathbf{B}j|\mathbf{V}j) > E(\mathbf{V}j|\mathbf{B}j)$

$$c_B(j) = 1,$$
$$c_V(j) = \min_{c_V(j)} \left\{ \mathrm{I}\left(\overline{\mathbf{G}}_B \mathbf{C}_B, \overline{\mathbf{G}}_V \mathbf{C}_V \right) \right\}$$

Else if $E(\mathbf{B}j|\mathbf{V}j) < E(\mathbf{V}j|\mathbf{B}j)$

$$c_V(j) = 1,$$
$$c_B(j) = \min_{c_B(j)} \left\{ \mathrm{I}\left(\overline{\mathbf{G}}_B \mathbf{C}_B, \overline{\mathbf{G}}_V \mathbf{C}_V \right) \right\}$$

End if
End MI Minimization
End for every source

F. Gain Vector Normalization at Each Source Location

Finally the diagonal matrix $\mathbf{L}$ is computed to equilibrate the columns of $\hat{\mathbf{G}}$. For $j \in [1, 2, \cdots, N]$

$$\mathbf{L}(j, j) = \left\| \hat{\mathbf{G}}(:, j) \right\|^{-1} \tag{8}$$

where $\hat{\mathbf{G}}(:, j)$ is the jth column of $\hat{\mathbf{G}}$.

It is shown in Section IV how this procedure produces better-equilibrated gain matrices with especially no dependency on the source orientation.

In the next section, the discussion regarding the practical resolution of the new linear system will take place in the general framework of the regularization of ill-posed inverse problems, in which most of the source estimation methods from distributed source models can be written.

III. The CMMI Method and the Source Estimation Problem

Source Estimation and Regularization: Regularization is a general framework for the estimation of source amplitudes as it necessitates the resolution of a linear system with an ill-conditioned operator [31]. Solving this kind of linear system without eliminating the oscillatory modes generated by the vectors associated with the smallest singular values produces source images with spurious sharp intensity gradients and large dipole amplitudes, which explain extremely well the data, but have no physiological sense [1]. Regularizing the inverse problem consists in attenuating or eliminating this behavior by introducing supplementary and *a priori* information on the sources to be estimated. Many articles and books describe the different approaches available here (see for instance [25], [33], and [34]).

Basically, the source estimate can be considered as an energy function minimizer

$$\hat{\mathbf{J}} = \min_{J} \left(U(\mathbf{J}) \right) \tag{9}$$

where

$$U(\mathbf{J}) = ||\mathbf{M} - \mathbf{G}\mathbf{J}||^2_{\mathbf{R}} + \lambda L(\mathbf{J}). \tag{10}$$

$\mathbf{R}$ is the variance-covariance matrix when a Gaussian noise component is assumed. λ is a positive scalar that balances the respective contributions to $U(\mathbf{J})$ of the data attachment term and the prior term $L(\mathbf{J})$. The $\mathbf{R}$-norm is defined as follows:

$$\|\mathbf{M}-\mathbf{GJ}\|^2_{\mathbf{R}} = (\mathbf{M}-\mathbf{GJ})^t \cdot \mathbf{R}^{-1} \cdot (\mathbf{M}-\mathbf{GJ}). \quad (11)$$

t stands for matrix transposition.

The regularization operator $L(\cdot)$ can be either quadratic or not, depending on the nature of the priors that one wishes to take into account. For instance, in the field of neuroimaging, this operator has been chosen either as the identity operator—thus producing source estimates with minimum norm priors, which are similar to solutions from pseudo inversion of the gain matrix; gradient or weighted Laplacian operator [16], or nonquadratic for nonlinear source estimators. In this latter case, one can choose priors in terms to the L1 norm of the source amplitudes that can avoid some of the smoothing properties of the L2 norm priors [41], [42]. Further, as extensively exposed in [1] and [32], the nonquadratic formulation of the priors may be linked to very numerous previous works in image restoration and reconstruction using Markov random fields models and associated stochastic algorithms [6]. The methods used in [1] and [32] consists in introducing explicit *a priori* information in order to recover sharp intensity gradients in the source image. It is assumed that the source magnitude pattern is made of areas with smooth intensity changes that may be separated by higher jumps in source amplitude: this situation occurs for instance between adjacent but functionally *nonrelated* cortical areas as the ones on both sides of a sulcus.

First, a system of neighborhood is designed between cortical areas. Then owing to a Bayesian formalism, the priors are easily integrated in the Gibbs formulation of a *a posteriori* density of probability. And finally the energy function $U(\mathbf{J})$ is minimized to produce a maximum *a posteriori* (MAP) estimate of the source pattern [8].

We will not go further in the general description of the different forms that can be adopted for the $L(\cdot)$ operator. Actually, the CMMI method can be adapted to many formulations for $L(\cdot)$.

A. The CMMI Method in the Context of Regularization

If we assume additive Gaussian noise on both EEG and MEG channels, the data attachment term in (10) is

$$\left\|\overline{\mathbf{M}}-\overline{\mathbf{G}}\,\mathbf{J}\right\|_{\overline{\mathbf{R}}}{}^2 = \left\|\overline{\mathbf{B}}-\overline{\mathbf{G}}_B\,\mathbf{J}\right\|_{\overline{\mathbf{R}}_B}{}^2 + \left\|\overline{\mathbf{V}}-\overline{\mathbf{G}}_V\mathbf{J}\right\|^2_{\overline{\mathbf{R}}_V} \quad (12)$$

where

$$\overline{\mathbf{R}}^{-1} = \begin{bmatrix} \overline{\mathbf{R}}_B^{-1} & 0 \\ 0 & \overline{\mathbf{R}}_V^{-1} \end{bmatrix}. \quad (13)$$

$\overline{\mathbf{R}}_V$ and $\overline{\mathbf{R}}_B$ are the variance/covariance matrices of the EEG and MEG noise, respectively [after row normalization by the $\mathbf{N}_B$ and $\mathbf{N}_V$ matrices, see (5)]. Hence, as it appears very clearly in (12), MEG and EEG data attachment terms are selectively weighted relatively to the different noise properties. Though these statistics might be driven from the data set, as proposed by Greenblatt [11] with the processing of prestimulus signals, we will for clarity consider both MEG and EEG noises as white and Gaussian with the same statistics (thus, with $\overline{\mathbf{R}}_B = \mathbf{I}_{N_B}$ and $\overline{\mathbf{R}}_V = \mathbf{I}_{N_V}$).

At this point, it is essential to note that the transformation from $\mathbf{G}$ to $\hat{\mathbf{G}}$ is nonlinear. Thus, the resolution of the linear system associated with the MI minimization transform necessitates an appropriate algorithm that we are now going to describe.

The iterative procedure is based on a rewriting of (3). Without any loss in generality, the additive noise components can be left out from now on. First let's define some notations: if $\mathbf{A}$ is a matrix with N columns, we note $\mathbf{A}_j$ as the jth column of the $\mathbf{A}$ matrix and $\tilde{\mathbf{A}}_j$ as the matrix made of the $N-1$ remaining columns. Then the linear system can be written

$$\overline{\mathbf{M}} = \overline{\mathbf{G}}\,\mathbf{J} = \mathbf{G}_j\,\mathbf{J}_j + \tilde{\mathbf{G}}_j\,\tilde{\mathbf{J}}_j. \quad (14)$$

$\mathbf{J}_j$ is the jth component of the source vector and $\tilde{\mathbf{J}}_j$ is a vector gathering the remaining $N-1$ source amplitudes.

The coefficients associated with $\mathbf{G}_j$, which have been computed during MI minimization can now be introduced

$$\begin{bmatrix} c_{bj}\left(\overline{\mathbf{B}}-\tilde{\mathbf{G}}_{Bj}\,\tilde{\mathbf{L}}_j\,\tilde{\mathbf{J}}_j\right) \\ c_{vj}\left(\overline{\mathbf{V}}-\tilde{\mathbf{G}}_{Vj}\,\tilde{\mathbf{L}}_j\,\tilde{\mathbf{J}}_j\right) \end{bmatrix} = \begin{bmatrix} c_{bj}\,\mathbf{G}_{Bj} \\ c_{vj}\,\mathbf{G}_{Vj} \end{bmatrix} \mathbf{L}_j\,\mathbf{J}_j \quad (15)$$

where $c_{bj} = c_b(j)$ and $c_{v_j} = c_v(j)$. We recall that according to preselection based on conditional entropy quantification, either c_{bj} or c_{vj} is equal to 1.

Please note that for simplicity, we have not introduced any new symbol for the dipole amplitudes despite the column normalization of the gain matrix by $\mathbf{L}$. It will then be necessary to multiply $\mathbf{J}$ by $\mathbf{L}$ at the end of the iterative estimation for correct scaling. We note

$$\begin{bmatrix} \hat{\mathbf{B}}j \\ \hat{\mathbf{V}}j \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{B}}-\tilde{\mathbf{G}}_{Bj}\tilde{\mathbf{L}}_j\tilde{\mathbf{J}}_j \\ \overline{\mathbf{V}}-\tilde{\mathbf{G}}_{Vj}\,\tilde{\mathbf{L}}_j\,\tilde{\mathbf{J}}_j \end{bmatrix}.$$

The iterative method estimates every $\mathbf{J}_j$ component sequentially, while taking the regularization term (which still corresponds to the prior on the global image) and the proper noise statistics into account in the following cost function

$$U(\mathbf{J}_j) = c^2_{bj}\left\|\hat{\mathbf{B}}_j - \mathbf{G}_{Bj}\,\mathbf{L}_j\,\mathbf{J}_j\right\|^2 + c^2_{vj}\left\|\hat{\mathbf{V}}_j - \mathbf{G}_{Vj}\,\mathbf{L}_j\,\mathbf{J}_j\right\|^2 + \lambda L(\mathbf{J}). \quad (16)$$

At this point, it is interesting to note that the respective data attachment terms of MEG and EEG are modulated by the coefficients descended from the MI minimization. For instance, if the local selective weighting coefficient c_{bj} is close to zero (hence $c_{vj}=1$), the estimation of the amplitude $\mathbf{J}_j$ will be mainly driven by EEG data. Zeroing the first derivative of $U(\mathbf{J}_j)$, we obtain

$$c^2_{bj}\,\mathbf{L}_j\,\mathbf{G}^t_{Bj}\,\hat{\mathbf{B}}_j + c^2_{vj}\,\mathbf{L}_j\,\mathbf{G}^t_{Vj}\,\hat{\mathbf{V}}_j = \left(c^2_{bj}\,\mathbf{G}^t_{Bj}\,\mathbf{G}_{Bj} - c^2_{vj}\,\mathbf{G}^t_{Vj}\,\mathbf{G}_{Vj}\right)\mathbf{L}_j\,\mathbf{J}_j - \lambda\frac{\partial L(\mathbf{J})}{2\partial \mathbf{J}_j} \quad (17)$$

and finally, as $c^2_{bj}\,L_j\,\mathbf{G}^t_{Bj}\,\mathbf{G}_{Bj}\,L_j + c^2_{bj}\,L_j\,\mathbf{G}^t_{Vj}\,\mathbf{G}_{Vj}\,L_j = 1$ [because of the normalization of the columns of $\hat{\mathbf{G}}$ (4)], we

obtain the following equation:

$$\mathbf{J}_j - \lambda \frac{\partial L(\mathbf{J})}{2\,\partial \mathbf{J}_j} = \mathbf{D}_j \tag{18}$$

where

$$\mathbf{D}_j = c_{bj}^2\,\mathbf{L}_j\,\mathbf{G}_{Bj}^t\,\hat{\mathbf{B}}_j + c_{vj}^2\,\mathbf{L}_j\,\mathbf{G}_{Vj}^t\,\hat{\mathbf{V}}_j. \tag{19}$$

Now let us consider different forms of the prior term $L(\mathbf{J})$.

1) $L(\mathbf{J})$ can be quadratic, with general form $L(\mathbf{J}) = \|\Gamma\mathbf{J}\|^2$, where Γ can take various forms. As $\|\Gamma\mathbf{J}\|^2 = \left\|\Gamma_j\,\mathbf{J}_j + \tilde{\Gamma}_j\,\tilde{\mathbf{J}}_j\right\|^2$, then $(\partial L(\mathbf{J})/2\,\partial \mathbf{J}_j) = \Gamma_j^t\left(\Gamma_j\,\mathbf{J}_j + \tilde{\Gamma}_j\,\tilde{\mathbf{J}}_j\right)$. And finally the estimate of $\mathbf{J}_j$ according to (18) is
$$\hat{\mathbf{J}}_j = \mathbf{L}_j \cdot \frac{\mathbf{D}_j - \Gamma_j^t\,\tilde{\Gamma}\,\tilde{\mathbf{J}}_j}{1 + \lambda\|\Gamma_j\|^2}. \tag{20}$$
This estimation is done sequentially for every dipolar source. Dipole scanning is repeated until convergence by continuous updating of the dipole moments with the new estimated values.
2) If $L(\mathbf{J})$ is nonquadratic, no useful formula for $\hat{\mathbf{J}}_j$ can be given, as it would strongly depends on the formulation of $L(\mathbf{J})$. Thus, as an example, we will stress in the next subsection on the adaptation of the ST-MAP estimation method.

Implementation of CMMI for the ST-MAP Source Estimation: As shown in [1], temporal regularization guidelines are powerful to enhance the results while processing data time series; however, we will not write them down here, as we want to emphasize the formulation of the multimodality operation. So $L(\mathbf{J})$ consists only in the spatial prior term, written as the sum of locally defined potential functions in terms of intensity gradients

$$L(\mathbf{J}) = \sum_{v=1}^{N_v} \Phi_v\left(\nabla\mathbf{J}_{|v}\right). \tag{21}$$

We denote ∇ as the gradient operator over the dipole amplitudes ($\nabla \in \Re_{N_v N}$, where $N_v = N \cdot N_n$, and N_n is the number of neighbors for each source j), $\nabla\mathbf{J}_{|v}$ is the vth element of the spatial gradient vector. We use

$$\Phi_v(u) = \frac{u^2}{1 + \left(\frac{u}{K_v}\right)^2} \tag{22}$$

where $u = \nabla\mathbf{J}_{|v}$ and K_v plays the role of a local detection threshold of intensity jumps in the source pattern. Actually for small gradients the local cost is quadratic, thus producing areas with smooth spatial changes in intensity, whereas for higher gradients, the associated cost is finite—$\Phi_v(u) \approx K_v$—thus allowing the preservation of discontinuities in the source image despite the needs for regularization.

The value given to the threshold where intensity jumps are not likely is much higher ($K_v = K_1$) than the one where there are *a priori* functional edges on the cortical surface ($K_v = K_2$, with $K_1 \gg K_2$). We choose $K_1 = 100 * K_2$, with $K_2 = 0.1$ nA.m (see [1] for further discussion on the choice of threshold values).

Dipole scanning at step j produces the following first derivative for $L(\mathbf{J})$ with respect to $\mathbf{J}_j$

$$\frac{\partial L(\mathbf{J})}{2\,\partial \mathbf{J}_j} = \nabla_j^t\,\Gamma\nabla\mathbf{J} = \nabla_j^t\,\Gamma\nabla_j\,\mathbf{J}_j + \nabla_j^t\,\Gamma\tilde{\nabla}_j\,\tilde{\mathbf{J}}_j \tag{23}$$

where

$$\Gamma = \operatorname{diag}\left(\frac{\Phi'\left(\nabla\mathbf{J}_{|v}\right)}{2\ \nabla\mathbf{J}_{|v}}\right), \quad v \in [1, N_v]$$

and Φ' is the first derivative of the Φ-function.

Then the estimation of $\mathbf{J}_j$ is computed iteratively as Γ depends on $\mathbf{J}_j$ [we can then write Γ as $\Gamma(\mathbf{J}_j)$]. At each iteration step k, we compute the following [see (18)]:

$$\hat{\mathbf{J}}_{j|_k} = \mathbf{L}_j \cdot \frac{\mathbf{D}_j - \lambda\nabla_j^t\,\Gamma(\mathbf{J}_{j|_{k-1}})\,\tilde{\nabla}_j}{1 + \lambda\nabla_j^t\,\Gamma(\mathbf{J}_{j|_{k-1}})\,\nabla_j} \tag{24}$$

where $\hat{\mathbf{J}}_{j|_k}$ is the estimation of $\mathbf{J}_j$ at the kth iteration. After convergence of the $\hat{\mathbf{J}}_{j|_k}$ estimate, then the algorithm proceeds the same way for $j = j+1$. Finally, the dipole scanning restarts from $j = 1$ on as described above for the quadratic case.

IV. Results

Simulations with both EEG and MEG have been carried out in spherical and realistic head models. These latter have been designed from MRI segmented images, and the boundary element method (BEM—number of elements; scalp: 3180, outer skull: 2018, inner skull: 1332 triangles). In the first set of simulations, we stress the validation of the CMMI in a very simple source model. We investigate the immunity of the method in presence of additive noise and we also check its ability to recover source patterns that EEG or MEG does not have (and especially radial sources). In the second set of simulations, the source configuration is more realistic. The source space is a piece of a segmented cortical surface, and here we investigate the ability of the CMMI method to recover more complex source configurations where simultaneous sources with various orientations are active.

A. Influence of the Noise and of the Source Orientations

Two whole-head arrays with 68 EEG and 68 MEG sensors (axial gradiometers) have been used for these simulations. The fusion data set is made of 34 equally spaced EEG sensors among the 68, and 34 MEG sensors.

We have chosen a rather simple cruciform anatomical model of the calcarine sulcus made of eight planes on which 64 dipole sources have been distributed [15] (eight dipoles per planes). We have voluntarily chosen a smaller number of sources than the number of sensors for not dealing with an underdetermined problem for this first step in the validation, and so as to stress on the respective merits of all four modalities (EEG, MEG, raw fusion, and CMMI) in a simple case. Each plane is a square patch of 1.5×1.5 cm with 8-mm distance between two parallel planes, and the cruciform model is located at about 6 cm below the scalp, which is rather deep. The dipoles on each plane are supposed to be functionally dependent (dipole amplitudes on the same plane are likely to be in the same range of intensity), whereas dipoles on two different planes are assumed to be uncorrelated (high intensity jumps are expected).

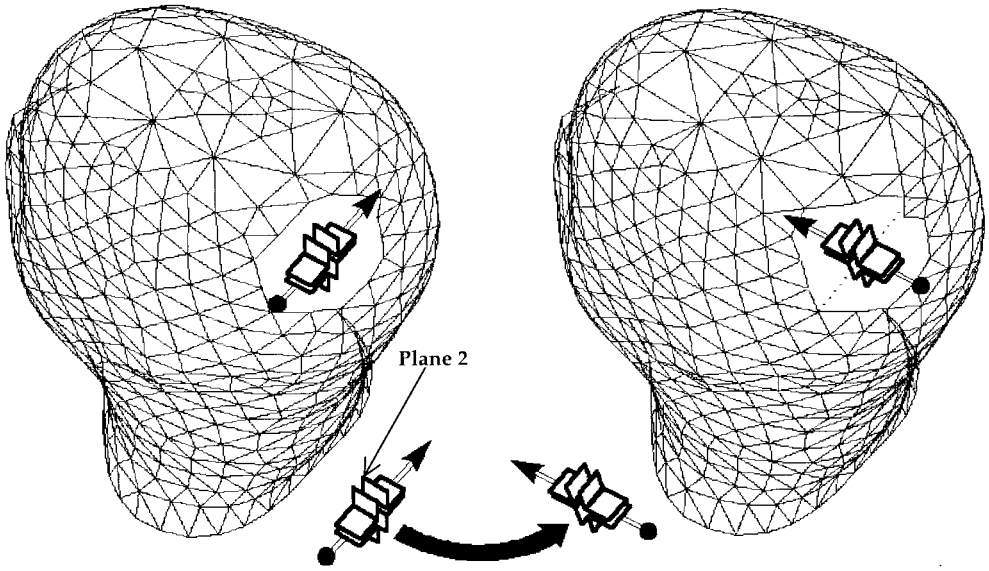


Fig. 1. Two extreme orientations of the cruciform: left = 0°—active dipoles are tangential, right = 90°—where many dipoles are almost radial (dipoles on plane #2 and on planes parallel to this plane).

TABLE I
SUMMARY OF THE RESPECTIVE SCORES OF EVERY MODALITY WITH THE ROTATING CRUCIFORM SIMULATIONS. THE AVERAGE ERROR AND STANDARD ERROR ARE COMPUTED ON ALL TEN ANGLES OF ROTATION. Min AND Max ARE, RESPECTIVELY, THE MINIMUM AND MAXIMUM VALUES TAKEN BY THE ERROR CRITERION ON THE 25 REALIZATIONS OF THE PSEUDORANDOM NOISE

	Spherical geometry				Realistic geometry			
	EEG	MEG	"Raw" fusion	Fusion with CMMI	EEG	MEG	"Raw" fusion	Fusion with CMMI
Average error (%)	27.5	15.3	10	1.8	15	3.2	8.3	1.7
Min (%)	10^{-3}	10^{-4}	0.2	0.2	0.5	10^{-3}	0.1	0.1
Max (%)	100	100	60.7	31.8	100	91.2	59	30
Average standard error (%)	±7.6	±4.8	±2.9	±0.4	±6.7	±1.4	±3	±0.4

The cruciform model is then rotated just as if it were following the circumvolutions of the sulcus from a 0° reference angle to 90° in ten equally spaced steps (Fig. 1). These simulations are simple enough to illustrate the respective performances of the different modalities both for:

- multiple source orientations in a distributed model;
- the ability for spatial discrimination between dipoles with same orientation on two close areas (distance between plane 1 and 2 is 8 mm, for instance).

1) Initial Source Configuration: Dipoles on plane 2 are then set to magnitude 1 (arbitrary units) for each of the orientation angle. At 0°, these dipoles are tangentially oriented, whereas for 90° they may be considered as radial.

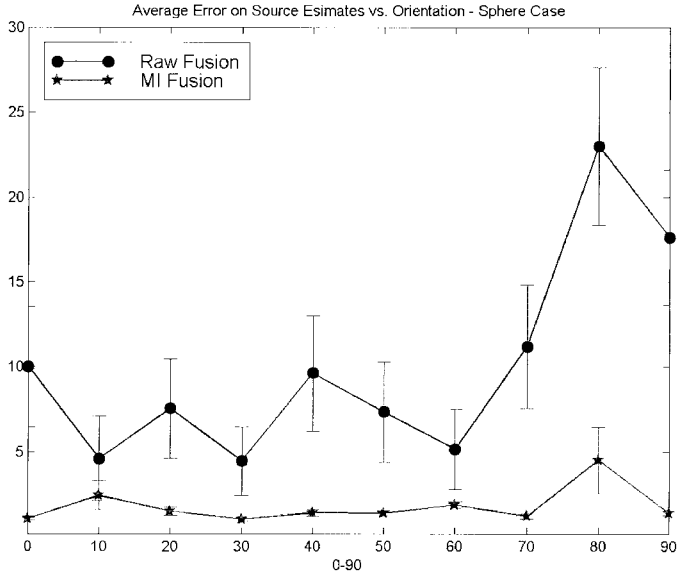
As shown in (12), both MEG and EEG noise characteristics can be taken into account for the processing of a real data set. Nevertheless and for simplicity, we will assume the same additive white Gaussian noise added to the data. The signal-to-noise ratio (SNR) is defined from the mean variance of the data, when the cruciform structure is turned in the ten directions. This ratio has been set to 20% for all the experiments.

As shown in [1], classical linear estimators fail to recover sharp edges in the initial source configuration presented here, whatever the source orientation may be. We then only discuss below the respective merits of the ST-MAP algorithm when it deals with the four available modalities.

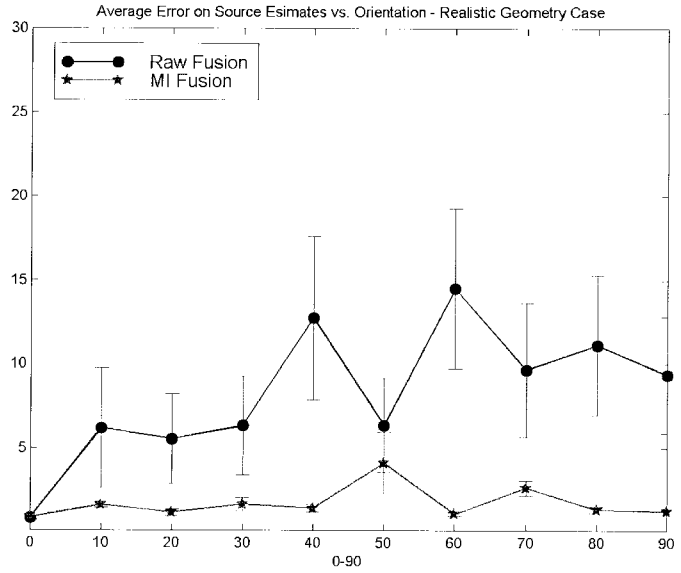
2) CMMI Versus the Classical Combination Method: We have first replaced the $\mathbf{C}_B$ and $\mathbf{C}_V$ matrices by identity matrices, thus no MI minimization is achieved. The associated simulation results, both with the spherical and the realistic head models are displayed in Fig. 2(a) and (b), respectively. It appears clearly that the combination of MEG and EEG without CMMI is much more sensitive to noise (please note the larger standard errors in Table I) with poorer performances (in the spatial discrimination between closely located planes). This is no more the case when the MI coefficients are introduced as modulations between MEG and EEG. One can notice the higher sensitivity to source orientation of the "raw" fusion method, especially in the spherical head model. Similar results have been obtained when for MEG/EEG fusion without CMMI, the iterative procedure—which is useless when $\mathbf{C}_B$ and $\mathbf{C}_V$ are identity operators—has been replaced by the standard regularized solution described in [1].

3) Comparison Between EEG, MEG, and CMMI: Results are discussed according to three criteria.

a) The dependence of the condition numbers of the three gain matrices (EEG, MEG, fusion) on the orientation angle. These numbers are presented as qualita-

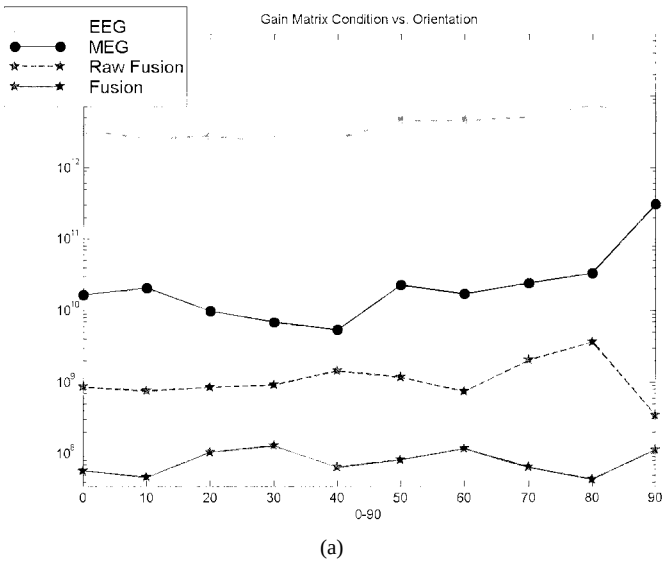


(a)

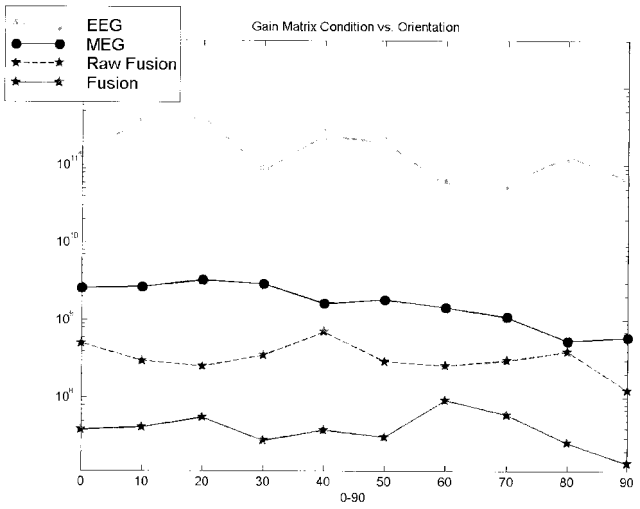


(b)

Fig. 2. Average reconstruction errors (in %) for the "raw" fusion modality (i.e., with equilibration of the rows and the columns of the gain matrix) and CMMI over 25 realizations of the pseudorandom noise. Error bars indicate the standard fluctuation around the average error: (a) spherical model and (b) realistic head model.



(a)



(b)

Fig. 3. Condition numbers of gain matrices plotted versus the sources orientations. The condition number of a matrix, according to the two-norm is the ratio of its largest singular value to its smallest: (a) with the spherical head model and (b) with the realistic head model.

tive indexes of the stability of the respective inverse problems, though the CMMI solution is computed iteratively with sequential introduction of the c_b and c_v coefficients.

b) The evolution of the respective average of the c_V and c_B weights with the orientation angle. These weights are computed relatively in percent of the sum $c_V + c_B$, for every angle. It will appear clearly that the respective merit of EEG versus MEG determined by CMMI slowly increases when the source pattern is rotated.

c) The average reconstruction error on dipoles over 25 realizations of the pseudorandom noise, defined as: $100 \times (\|\mathbf{J}_{\text{true}} - \mathbf{J}_{\text{estimated}}\|/\|\mathbf{J}_{\text{true}}\|)$, and the standard error of the fluctuations of the estimate will be indicated with error bars. This error quantification is severe, as small differences between sources will create large error values. For instance, a 50% error indicates that the estimated activity is spread over both the original active zone and "phantom ones" which are close in location and orientation; 100% error denotes source estimates in which the active plane is not even partly recovered. The regularization parameter is tuned for every noise realization so as to get the smallest error on the dipole sources.

4) Results with the Spherical Head Model:

a) Gain matrix condition numbers are shown in Fig. 3(a).

For each source orientation, the condition numbers of the EEG, MEG, and the combined gain matrices are computed. Whereas EEG condition does not depend

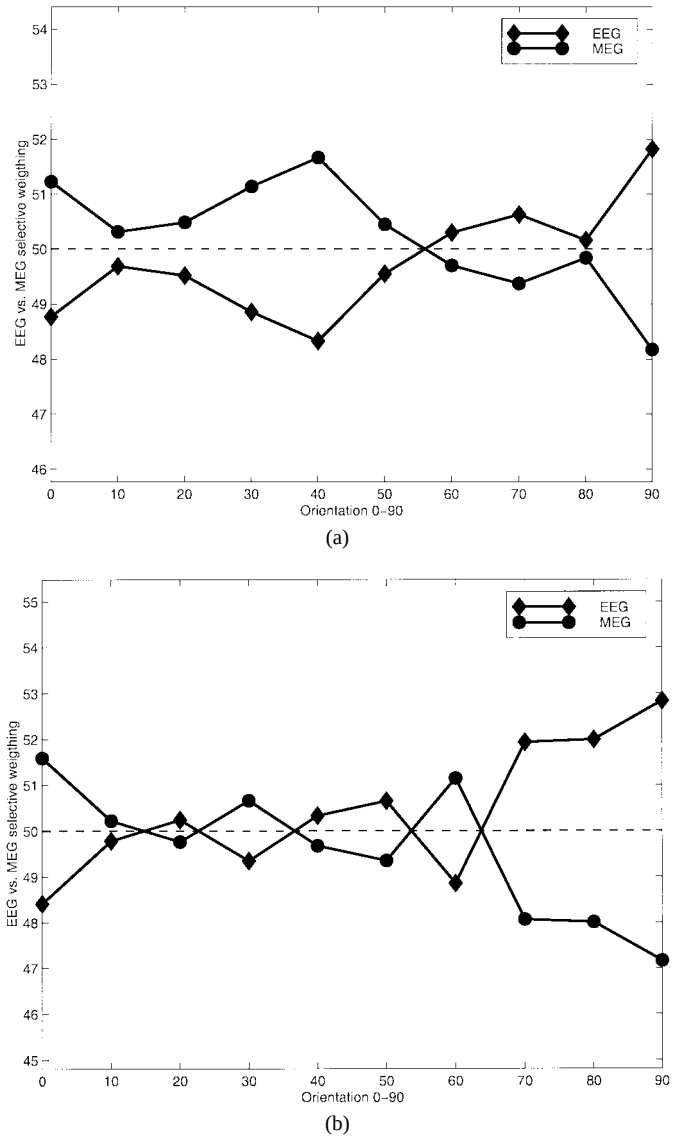


Fig. 4. MEG and EEG relative weights versus the orientation. This plot shows the relative faith accorded by the MI criterion to one modality in regard to the other. For each angle, the weights are averaged over those affecting the 64 sources: (a) spherical model and (b) realistic model.

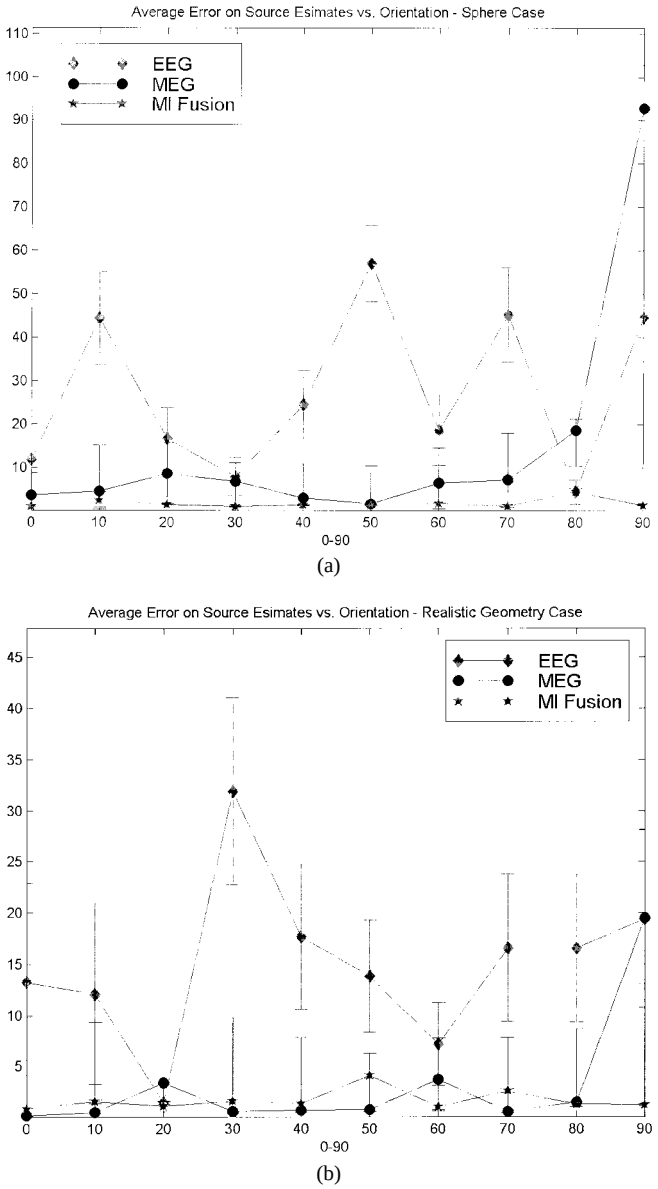


Fig. 5. Average reconstruction errors (in %) with the three modalities (MEG, EEG, and CMMI) over 25 realizations of the pseudorandom noise. Error bars indicate the standard error: (a) spherical model and (b) realistic head model.

on orientation, MEG's significantly increase from 0° to 90° orientation as it was expected. MEG-EEG combination with classical equilibration as proposed in [5] and [18] produces gain matrices with better condition, on the order of one size order relatively to MEG on the average. Finally, fusion with CMMI brings additional improvement on gain matrix condition for every orientation (up to five orders of magnitude toward EEG).

b) The relative average weighting of MEG versus EEG is plotted on Fig. 4(a). It is clear that whereas MEG is still relatively favored in comparison to EEG for most the most tangential source orientations, EEG is selected as "the best" modality for the last four angles. For these latter, more and more sources tend to become almost radial. This result show that preselection with conditional entropy makes the correct choice of the favorite modality.

c) Fig. 5(a) shows the reconstruction performances of the three modalities. This plot illustrates perfectly MEG's failure when more sources are becoming radial in the spherical model (90° of orientation). The EEG curve indicates error values that reflect its poor spatial resolution. More regularization would increase the reconstruction error, whereas more confidence to data leads to increased instability. Fusion succeeds in locating faithfully the active area for every source orientation, with less than 1.7% of error on the average (see full scores in Table I).

5) Results with the Realistic Head Model:

a) Condition numbers of all the three modalities are slightly improved in this source configuration when using the realistic head model [Fig. 3(b)]. This may be due to the breaking of symmetry in the conductor geometry that tended to produce almost collinear lead

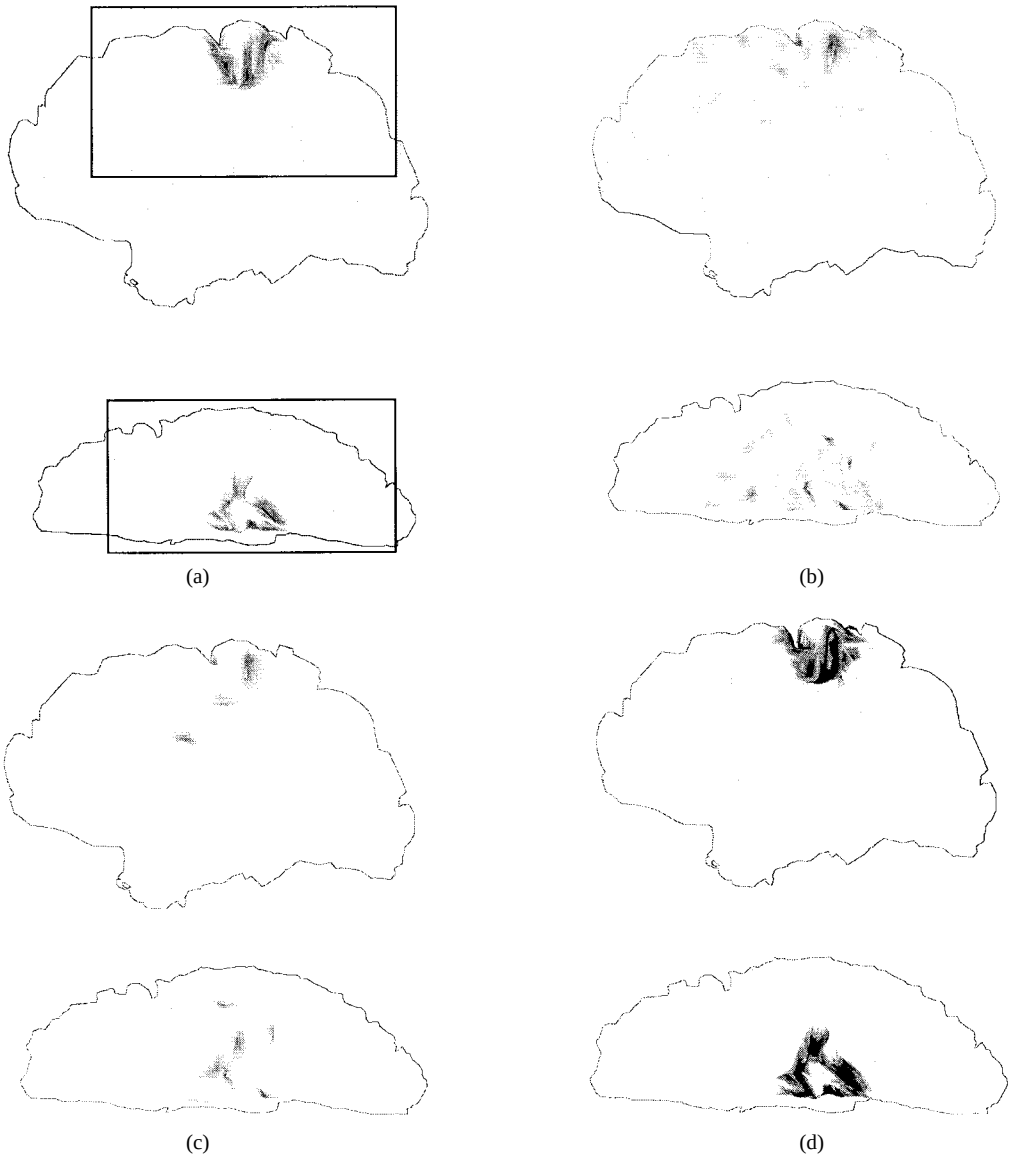


Fig. 6. Source estimations using the realistic head and cortical geometry. (a) The original source configuration with 22 dipoles set to value one. The black frame indicates the global source space of 242 sources; top: inner view of the right hemisphere, bottom, top view, (b) EEG source estimate, (c) MEG source estimate, and (d) CMMI source estimate.

fields in some source configurations. MEG for instance has gain matrices with constant condition for every orientation. Nevertheless, the fusion modality is still better conditioned, by up to two orders of magnitude in comparison to MEG.

b) The relative weights still depend on the orientation in the same way as for the spherical head model [see Fig. 4(b)]. This is explained by the fact that previous radial sources are now sources normal to the head surface, which is sufficient to be considered as quasi-radial sources [28], [39]. This is especially the case for dipoles located at the back of the head where the spherical model is a good approximation of the actual head anatomy. Furthermore, the entropy analysis of EEG and MEG matrices reveals the weakness of MEG for the three last angles, whereas the study of the condition numbers is not sufficient to give this conclusion.

c) Despite the increase in MEG performances, the error curves on Fig. 5(b) still clearly show the superiority of the fusion modality both in correct localization and robustness. Note that for 90° orientation, MEG still fails because of the pseudoradial orientation of half of the source pattern (Table I).

B. Simulations with a Realistic Cortical Geometry

Following these studies on small source configurations, we now discuss the ability of the CMMI method to discriminate some source patterns distributed along a realistic cortical geometry with a larger source space. In this purpose, 242 sources are distributed normally to a piece of cortical surface of the right hemisphere [see Fig. 6(a)]. These source are distributed among 11 families made of 22 dipoles each, which will be considered as functionally independent from each other (just as if they were some different planes from the previous cruciform model). The ST-MAP prior model

explicitly takes this partition into account. It should be noted that more sophisticated systems of anatomical and functional neighborhood should be taken into account to process real data sets.

The original active sources consist in 22 dipoles belonging to the same family with their amplitude set to one. The other dipole magnitudes are set to zero. Ten of these dipoles can be considered as pseudoradial when considering their orientations pointing toward a normal to the scalp surface. They are located at various depths ranging from 3.6 to 6.6 cm from the scalp surface (4.7 cm on the average). The MEG data are simulated on the 143 MEG sensor array of the CTF Inc. whole head system (axial gradiometers). The 143 EEG sensors are simulated electrodes which locations correspond to the projection of the MEG sensors on the scalp. The fusion sensors are made of 72 MEG and 71 EEG from the original locations, covering the whole head. The simulated data are corrupted with white and Gaussian noise, with 20% standard deviation. The original source pattern is represented in Fig. 6(a).

Results of EEG and MEG source estimation are displayed in Fig. 6(b) and (c), respectively. It appears very clearly that the EEG solution contains the original sources but combined with spurious activity all along a large extension of the hemisphere. The extension of the MEG solution on the cortex is smaller than in EEG, but the deepest sources and those with pseudoradial orientation are not properly estimated. These complementary behaviors cooperate extremely well in the CMMI estimation with ST-MAP [Fig. 6(d)] that recovers successfully the original source pattern with few scattering on the cortical surface.

V. Conclusions and Perspectives

A method for a joint EEG and MEG brain imaging modality has been described in this paper. This routine is based on a preprocessing of the EEG and MEG gain matrices. In a given distributed source model, a linear system of equation relates the source amplitudes to the sensor array. Due to a) EEG sensitivity to head tissue geometry and conductivity properties and b) MEG blindness toward deeply located or radial sources; solving this linear system is an ill-posed problem. From the algebraic point of view, it means that the gain matrices are numerically rank deficient, as some sources possess almost the same lead-fields.

A parallel between algebraic considerations and multimodality registration with MI led us to the design of a method that operates selective weighting of the MEG and EEG lead-fields so that each of these modified gain vectors contributes to the minimization of MI between the gain matrices. This preprocessing of the fusion global gain matrix produces a better conditioned inverse problem, which is solved owing to a regularized iterative procedure that can take various types of priors into account. It results in stability improvement and smaller reconstruction errors in the dipole amplitude estimation. Most important is the immunity of the fusion modality in regard to source orientation.

This first exploration step is promising and further experiments with increasing realism will be carried out to better evaluate the limits of this approach (with a real skull phantom for instance [35]), before it could be used as genuine brain functional imaging in cognitive neuroscience experiments [47].

Appendix
Conditional Entropy Quantification of Gain Vectors and the Computation of Mutual Information

A. Conditional Entropy Quantification of Gain Vectors

Entropy quantification of a continuous random variable (RV) from a limited set of data (here the elements of the $\mathbf{B}j$ and $\mathbf{V}j$ gain vectors are considered as the realization of continuous RV's defined below) necessitates the definition of a discrete code [36]. Thus, each of the MEG and EEG lead fields are first linearly scaled on the set of discrete values $[1, 2, \cdots, n]$, with

$$n = 2 \cdot M \cdot round(\sigma_{\overline{\mathbf{B}}_j}, \sigma_{\overline{\mathbf{V}}_j}) \tag{A1}$$

$\sigma_{\overline{\mathbf{B}}_j}$ (respectively, $\sigma_{\overline{\mathbf{V}}_j}$) is the standard deviation of the normalized lead field $\mathbf{B}_j$ (respectively, $\mathbf{V}_j$), and $M = \max\{N_B, N_V\}$.

$E(\mathbf{B}j|\mathbf{V}j)$ [respectively, $E(\mathbf{V}j|\mathbf{B}j)$] is the conditional entropy of RV's $\mathbf{B}j$ versus $\mathbf{V}j$ (respectively, $\mathbf{V}j$ versus $\mathbf{B}j$) which realizations are the elements of the respective gain vectors

$$E(\mathbf{B}j|\mathbf{V}j) = -\sum_{bj,\,vj} p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj)\,\log\frac{p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj)}{p_{\mathbf{V}j}(vj)} \tag{A2}$$

$$E(\mathbf{V}j|\mathbf{B}j) = -\sum_{bj,\,vj} p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj)\,\log\frac{p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj)}{p_{\mathbf{B}j}(bj)} \tag{A3}$$

$p_{\mathbf{B}j}(bj)$ and $p_{\mathbf{V}j}(vj)$ are the marginal, and $p_{\mathbf{B}j,\,\mathbf{V}j}(bj,\,vj)$ the joint probability distributions of, respectively, Bj and Vj.

Please note that when $N_B \neq N_V$ (like in the second set of simulations presented here), this computation is done using linearly interpolated lead-fields in order that they finally have the same number of elements. Hence, quantities such as $p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj)$ can be estimated.

These probability distributions can be estimated by normalization of the joint histogram $h(bj,\,vj)$, of the lead $\mathbf{B}j$ and $\mathbf{V}j$ fields. This histogram is computed by binning the pairs of values $(bj,\,vj)$ along the two gain vectors. Then

$$p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj) = \frac{h(bj,\,vj)}{\sum_{bj,\,vj} h(bj,\,vj)} \tag{A4}$$

and

$$p_{\mathbf{V}j}(vj) = \sum_{bj} p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj) \tag{A5}$$

$$p_{\mathbf{B}j}(bj) = \sum_{vj} p_{\mathbf{B}j\,\mathbf{V}j}(bj,\,vj). \tag{A6}$$

B. Computation of Mutual Information

In a very similar way, MI between the two RV's GB and GV associated with the gain matrices $\mathbf{G}_{\mathrm{B}}$ and $\mathbf{G}_{\mathrm{V}}$ also necessitates a preliminary scaling of the matrices to the range $[1, 2, \cdots, n]$, with

$$n = 2 \cdot M \cdot \mathrm{round}\left(\max_j (\sigma_{\overline{\mathbf{B}}_j}, \sigma_{\overline{\mathbf{V}}_j}) \right) \tag{A7}$$

where $\sigma_{\overline{\mathbf{B}}_j}$ (respectively, $\sigma_{\overline{\mathbf{V}}_j}$) is the standard deviation of the normalized lead field $\mathbf{B}_j$ (respectively, $\mathbf{V}_j$), and $M = \max\{N_B, N_V\}$.

As for the $\mathbf{B}_j$ and $\mathbf{V}j$ gain vectors above, both of the $\mathbf{G}_{\mathrm{B}}$ and $\mathbf{G}_{\mathrm{V}}$ matrices are seen as an ensemble of realizations of GB and GV, respectively, with marginal, $p_{GB}(gb)$ and $p_{GV}(gv)$, and joint, $p_{GB,GV}(gb, gv)$, probability distributions. The mutual information between these two RV's is:

$$I(GB, GV) = \sum_{gb, gv} p_{GB,GV}(gb, gv) \cdot \log\left(\frac{p_{GB,GV}(gb, gv)}{p_{GB}(gb) \cdot p_{GV}(gv)} \right). \tag{A8}$$

An estimation of MI can be computed when the two matrices are considered as two intensity images to be superimposed. From these pseudoimages, it is possible to compute a joint histogram $h(gb, gv)$ where gb (respectively, gv) can take all the values of the rescaled $\mathbf{G}_{\mathrm{B}}$ (respectively, $\mathbf{G}_{\mathrm{V}}$) matrix [37]. The joint histogram $h(gb, gv)$ is calculated by binning the intensity pairs (gb, gv) along the different element positions of the gain matrices.

Then, in a similar manner as for conditional entropy quantification

$$p_{GB,GV}(gb, gv) = \frac{h(gb, gv)}{\sum_{gb, gv} h(gb, gv)} \tag{A9}$$

and

$$p_{GV}(gv) = \sum_{gv} p_{GB,GV}(gb, gv) \tag{A10}$$

$$p_{GB}(gb) = \sum_{gv} p_{GB,GV}(gb, gv). \tag{A11}$$

Once again, when $N_B \neq N_V$ (like in the second set of simulations presented here), this computation is done using linearly interpolated lead-fields in order that both matrices finally have the same number of rows.

REFERENCES

[1] S. Baillet and L. Garnero, "A Bayesian framework to introducing anatomo-functional priors in the EEG/MEG inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 374–385, May 1997.

[2] P. Charbonnier, Laure Blanc-Féraud, and Michel Barlaud, "Deterministic edge-preserving regularization in computer imaging," *IEEE Trans. Image Processing*, vol. 5, Dec. 1996.

[3] D. Cohen, N. B. Cuffin, K. Yunokuchi, R. Maniewski, C. Purcell, G. R. Cosgrove, J. Ives, J. G. Kennedy, and D. L. Schomer, "MEG versus EEG localization test using implanted sources in the human brain," *Ann. Neurol.*, vol. 28, pp. 811–817, 1990.

[4] M. Fuchs, "Possibilities of functional brain imaging using a combination of MEG and MRT," in *Oscillatory Event-Related Brain Dynamics*, C. Pantev, Ed. New York: Plenum, pp. 435–457, Nov. 1994.

[5] A. M. Dale and M. I. Sereno, "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach," *J. Cogn. Neurosci.*, vol. 5, pp. 162–176, 1993.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine. Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[7] J. S. George, C. J. Aine, J. C. Mosher, D. M. Schmidt, D. Ranken, H. A. Schlitt, C. C. Wood, J. D. Lewine, J. A. Sanders, and J. Belliveau, "Mapping function in human brain with MEG, anatomical MRI and f-MRI," *J. Clin. Neurophysiol.*, vol. 5, no. 12, pp. 406–431, 1995.

[8] G. Gindi, M. Lee, A. Raganrajan, and I. G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Med. Imag.*, vol. 12, pp. 670–680, Dec. 1993.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: The John Hopkins Univ. Press, 1989, pp. 79–81.

[10] I. F. Gorodnistky, J. S. George, and B. D. Rao, "Neuromagnetic imaging with FOCUSS: A recursive weighted minimum norm algorithm," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, pp. 231–251, 1995.

[11] R. E. Greenblatt, "Probabilistic reconstruction of multiple sources in the bioelectromagnetic inverse problem," *Inverse Problems*, vol. 9, pp. 271–284, 1993.

[12] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. Lounasmaa, "Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Modern Phys.*, vol. 65, Apr. 1993.

[13] M. Hämäläinen and R. Ilmoniemi, "Interpreting measured magnetic fields of the brain: Estimates of current distributions," Helsinki Univ., Helsinki, Finland, Tech. Rep. TKK-F-A620, 1984.

[14] P. C. Hansen, "Regularization tools, a Matlab package for analysis and solution of discrete ill-posed problems," *Numer. Algorithms*, vol. 6, pp. 1–35, 1994.

[15] L. Kaufman, J. H. Kaufman, and J. Z. Wang, "On cortical folds and neuromagnetic fields," *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, pp. 211–226, 1991.

[16] R. D. Pascual-Marqui, C. M. Michel, and D. Lehman, "Low resolution electromagnetic tomography: A new method for localizing electrical activity of the brain," *Int. J. Psychophysiol.*, vol. 18, pp. 49–65, 1994.

[17] M. E. Pflieger, G. V. Simpson, J. J. Foxe, S. P. Ahlfors, H. G. Vaughan, Jr., J. Hrabe, R. J. Ilmoniemi, and G. Lantos, "Superadditive information from simultaneous MEG/EEG data," *Human Brain Mapping*, supp. 1, 25, 1995.

[18] J. W. Phillips, R. M. Leahy, and J. C. Mosher, "Imaging neural activity using MEG and EEG," *IEEE Eng. in Med. and Biol. Mag.*, vol. 16, no. 3, pp. 34–41, May/June 1997.

[19] M. Scherg and H. Buchner, "Somatosensory evoked potentials and magnetic fields: Separation of multiple source activity," *Physiological Meas.*, suppl. 4A, vol. 14, pp. A35–39, Nov. 1993.

[20] M. Scherg, "Fundamentals of dipole source potential analysis," in *Auditory Evoked Magnetic Fields and Potentials, Advances in Audiology*, F. Grandori, M. Hoke, and G. L. Roamni, Eds. Basel, Switzerland: 1990, vol. 6, pp. 40–69.

[21] B. Scholz and G. Schwierz, "Probability-based current dipole localization from biomagnetic fields," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 735–742, Aug. 1994.

[22] C. J. Stok, J. W. H. Meijs, and M. J. Peters, "Inverse solutions based on MEG and EEG applied to volume conductor analysis," *Phys. Med. Biol.*, vol. 32, no. 1, pp. 99–104, 1987.

[23] W. W. Sutherling, P. H. Crandall, T. M. Darcey, D. P. Becker, M. F. Levesque, and D. S. Barth, "The magnetic and electric fields agree with intracranial localizations of somatosensory cortex," *Neurol.*, vol. 38, pp. 1705–1714, Nov. 1988.

[24] C. D. Tesche, M. A. Uusitalo, R. J. Ilmoniemi, M. Huotilainen, M. Kajola, and O. Salonen, "Signal-space projections of MEG data characterize both distributed and well-localized neuronal sources," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, pp. 189–200, Sept. 1995.

[25] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems.* Washington, D.C.: Winston, 1977.

[26] P. Valdes-Sosa, F. Marti, F. Garcia, and R. Casanova, "Variable resolution electric-magnetic tomography," in *Proc. Biomag. '96*, Santa Fe, NM, to be published.

[27] I. Vajda, *Theory of Statistical Inference and Information.* Dordrecht, The Netherlands: Kluwer, 1989.

[28] J. P. Wikswo, Jr., A. Gevins, and S. J. Williamson, "The future of the EEG and MEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 87, pp. 1–9, 1993.

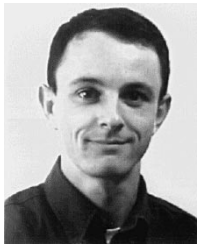
[29] C. C Wood, D. Cohen, B. N. Cuffin, M. Yarita, and T. Allison, "Electrical sources in human somatosensory cortex: Identification by combined magnetic and potential recordings," *Sci.*, vol. 227, pp. 1051–1053, Mar. 1985.
[30] A. van der Sluis, "Condition numbers and equilibration of matrices," *Numer. Math.*, vol. 14, pp. 14–23, 1969.
[31] J. H. Wilkinson, *Linear Algebra.* Englewood Cliffs, NJ: Prentice-Hall, 1963.
[32] J. W. Phillips, R. M. Leahy, and J. C. Mosher, "MEG-based imaging of focal neuronal current sources," *IEEE Trans. Med. Imag.*, vol. 16, pp. 338–348, June 1997.
[33] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech Signal Processing*, vol. 37, pp. 2024–2036, Dec. 1989.
[34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. New York: Cambridge Univ. Press, 1992.
[35] S. Baillet, G. Marin, F. Le Rudulier, and L. Garnero, "Evoked potentials from a real-skull phantom-head: An experimental step toward the validation of methods for solving the forward and inverse problems for EEG/MEG source imaging," submitted for publication.
[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.
[37] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, pp. 187–198, Apr. 1997.
[38] W. Wells, III, "Multimodal volume registration by maximization of mutual information," *Med. Imag. Anal.*, vol. 1, no. 1, pp. 33–51, 1996.
[39] J. Haueisen, C. Ramon, P. Czapski, and M. Eiselt, "On the influence of volume currents and extended sources on neuromagnetic fields: A simulation study," *Ann. Biomed. Eng.*, vol. 23, pp. 728–739, 1995.
[40] G. Marin, C. Guérin, S. Baillet, and L. Garnero, "Influence of skull anisotropy for the forward and inverse problem in EEG: Simulation studies using FEM on realistic head models," *Human Brain Mapping*, vol. 6, no. 4, pp. 250–269, 1998.
[41] K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 42, pp. 608–615, June 1995.
[42] ———, "A robust reconstruction of sparse biomagnetic sources," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 720–726, Aug. 1997.
[43] I. Merlet, R. Garcia-Larrea, L. Paetau, K. Uutela, M. L. Granström, and F. Mauguiere, "Apparent asynchrony between interictal electric and magnetic spikes," *Neurorep.*, vol. 5, no. 8, pp. 1071–1076, Mar. 1997.
[44] R. C. Knowlton, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Magnetoencephalography in partial epilepsy: Clinical yield and localization accuracy," *Ann. Neurol.*, vol. 42, no. 4, pp. 622–631, Oct. 1997.
[45] N. Mikuni, T. Nagamine, A. Ikeda, K. Terada, W. Taki, J. Kimura, H. Kikuchi, and H. Shibasaki, "Simultaneous recording of epileptiform discharges by MEG and subdural electrodes in temporal lobe epilepsy," *Neuroimag.*, vol. 5, no. 4, pp. 298–306, May 1997.
[46] K. Dy, C. Kufta, D. Scaffidi, and S. Sato, "Source localization determined by MEG and EEG in temporal lobe epilepsy: Comparison with electrocorticography: technical case report," *Neurosurg.*, vol. 42, no. 2, pp. 414–421, Feb. 1998.
[47] C. Eulitz, H. Eulitz, and T. Elbert, "Differential outcomes from magneto- and electroencephalography for the analysis of human cognition," *Neurosci. Lett.*, vol. 227, no. 3, pp. 185–188, May 1997.

Sylvain Baillet was born in France in 1970. From 1990 to 1994, he was a student in applied physics at Ecole Normale Supérieure of Cachan, France. He received the *agrégation* in applied physics in 1993 and majored in signal processing at the University of Paris-Sud and the Ecole Supérieure d'Electricité, Orsay, in 1994. In 1998, he received the Ph.D. degree in electrical engineering from the University of Paris-Sud. He has been completing the Ph.D. degree program at the Institute of Optics, Orsay, and at the Laboratory of Cognitive Neuroscience and Brain Imaging at La Salpêtrière Hospital, Paris.

He is now a postdoctoral Research Associate at the Signal and Image Processing Institute at the University of Southern California, Los Angeles and a Laureate of the Lavoisier Fellowship. His research interests primarily concern data processing and source modeling in electromagnetic functional imaging of the brain.

Line Garnero was born in France on September 4, 1955. She received the Ph.D. degree and the Doctorat d'Etat from the University of Paris XI, Orsay, in 1981 and 1987, respectively.

She is Chargé de Recherche at the CNRS (National Center of Scientific Research). She worked at the Institute of Optics in Orsay from 1981 to 1996. Her research field was the image reconstruction techniques for microwave, X-ray, and γ tomography. She is currently working in the laboratory of Cognitive Neuroscience and Brain Imaging Unit at La Salpêtrière Hospital, Paris, where her main research concerns the reconstruction of brain electrical activity from electroencephalographic (EEG) or magnetoencephalographic (MEG) data, and the fusion between IRMf, EEG, and MEG data.



Gildas Marin was born in Cholet, France, in 1970. From 1990 to 1994, he was a student at the Ecole Superieure d'Optique in Orsay. He majored in optics and received the Ph.D. degree in 1996, at the University of Paris-Sud. During his Ph.D. program, he was at the Institute of Optics in Orsay where he developed BEM and FEM realistic head models for the resolution of the EEG-MEG forward problem.

He is now with the research division of the Essilor Company in France.

Jean-Paul Hugonin received the Doctorat d'Etat from the University of Paris XI, Orsay, in 1983.

He has been an Assistant Professor of physics at the University of Paris VI since 1974 and joined the Institut d' Optique in Orsay at the same time. His research interests include image processing, digital photography and numerical theory of diffraction.