Improving Human Brain Mapping via Joint Inversion of Brain Electrodynamics and the BOLD Signal

Kevin S. Brown\textsuperscript{a,c}, Stephanie Ortigue\textsuperscript{b,c,d}, Scott T. Grafton\textsuperscript{a,b,c,d}, Jean M. Carlson\textsuperscript{a,c}

\textsuperscript{a}Department of Physics, University of California, Santa Barbara, CA 93106
\textsuperscript{b}Department of Psychology, University of California, Santa Barbara, CA 93106 USA
\textsuperscript{c}Institute for Collaborative Biotechnologies, University of California, Santa Barbara, CA 93106
\textsuperscript{d}UCSB Brain Imaging Center, University of California, Santa Barbara, CA 93106 USA

Abstract

We present several methods to improve the resolution of human brain mapping by combining information obtained from surface electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) of the same participant performing the same task in separate imaging sessions. As an initial step in our methods we used independent component analysis (ICA) to obtain task-related sources for both EEG and fMRI. We then used that information in an integrated cost function that attempts to match both data sources and trades goodness of fit in one regime for another. We compared the performance and drawbacks of each method in localizing sources for a dual visual evoked response experiment, and we contrasted the results of adding fMRI information to simple EEG-only inversion methods. We found that adding fMRI information in a variety of ways gives superior results to classical minimum norm source estimation. Our findings lead us to favor a method which attempts to match EEG scalp dynamics along with voxel power obtained from ICA-processed blood oxygenation level dependent (BOLD) data; this method of joint inversion enables us to treat the two data sources as symmetrically as possible.

Introduction

The rapid growth of large-scale, high-precision neuroimaging technology has allowed the study of cognitive processes formerly only accessible via behavioral measures. A key question in cognitive neuroscience is a fundamental one: when and where do task-related activations in the brain occur? Unfortunately, no single noninvasive imaging technology is sufficient to fully capture spatiotemporal brain dynamics on psychologically relevant temporal and spatial scales. High density surface EEG has millisecond temporal resolution, but it is based on 100-200 spatially correlated measurements on the scalp.\textsuperscript{1} Complementary to EEG measurements, mapping with fMRI has millimeter spatial resolution, but the low-pass filtering nature of the hemodynamic response function (HRF) makes it difficult or impossible to resolve dynamic events separated by less than several seconds.

Recently there has been increasing interest in combining spatially resolved BOLD signal measurements from fMRI with temporally resolved EEG data in order to enhance resolution in both space and time when trying to understand the neural basis of cognitive processes. Several methods have been proposed including Bayesian statistical methods (Phillips et al., 2005) and those based on linear time-invariant system theory (Liu and He, 2008). The immediate challenge with such a joint approach is that the methods yield fundamentally different measurements in relation to underlying physiologic processes. A critical first step is to find a modeling framework that can represent both data types in a complementary way. In this paper, we do this by first modeling the two datasets with temporal and spatial ICA. From this a joint optimization that attempts to fit both EEG and fMRI data simultaneously can be developed.

Independent component analysis (ICA) is a method for so-called blind source separation, and was originally developed in the context of speech signals (Jutten and Herault, 1991). ICA has been increasingly employed in medical imaging (McKeown et al., 1998; Makeig et al., 1996; Eichele et al., 2009, 2008; Moosmann et al., 2008). These joint decompositions either used EEG information from only a single electrode (Calhoun et al., 2006) or used the joint decomposition to attempt to avoid the EEG inverse problem (Moosmann et al., 2008). Additionally, they required truly simulta-

\textsuperscript{1}Unless otherwise noted, “EEG” means “surface EEG” and not intracranial recordings.

Preprint submitted to Elsevier May 12, 2009
neous EEG and fMRI measurements, which will not be available to all investigators interested in combining the two imaging modalities.

We employ ICA as well, but emphasize that ICA alone does not solve the mapping resolution problem. ICA is a very useful data compression technique and in the context of neuroimaging experiments can be an invaluable “automatic” method for artifact removal (LeVan et al., 2006; Ting et al., 2006). We use ICA to clean and compress the data, but additional machinery is required to merge the two imaging modalities. We present a principled way of performing this merger, using joint cost functions that simultaneously match both EEG and fMRI data to obtain fine temporally and spatially resolved brain maps.

Our approach is to treat the determination of source dynamics in the brain as a large joint inverse problem; we build on and expand the work using synthetic data in Brookings et al. (2009). Sources of electrical activity in the cortex arising as a result of particular tasks give rise to both EEG signals, via volume conduction through brain, cerebrospinal fluid, skull, and scalp, and measured fMRI signals, via the BOLD response. Our methods generate a single set of dynamic brain sources that best reproduce the observed EEG and fMRI signals.

The EEG-only inverse problem is highly underdetermined (see below), and the way in which we should incorporate fMRI information is unclear. For this reason, we compare and contrast several different joint inverse problems that incorporate fMRI measurements in different ways. We consider both quadratic cost functions, which can be solved with straightforward linear-algebraic techniques, and higher-order cost functions that require iterative methods. We show promising results for several of these methods in a dual visual evoked response experiment with human participants. We argue that experimental design and study goals are important considerations when deciding how to fuse EEG and fMRI to obtain high-resolution human brain maps.

Materials and Methods

The data processing pipeline for joint inversion is illustrated schematically in Fig. 1. The steps in this pipeline are discussed in more detail in the following sections.

Experimental Design

Participants

Twenty volunteers (five female) participated in the present study involving EEG, fMRI, or both. All were right-handed by the Edinburgh Handedness Inventory (Oldfield, 1971) and had normal or corrected-to-normal visual acuity. None had any prior or current neurological or psychiatric impairment, as ascertained by a detailed anamnesis. Mean age of participants was 24 years (range 18-44 years). Here we focus on data from four volunteers of the group who were studied with both EEG and fMRI. Prior to participation, volunteers provided written informed consent that had been approved by the Ethical Committee of the University of California, Santa Barbara.

![Figure 1: Box diagram illustrating data pipeline steps leading to joint inversion.](image-url)
hand-on-object action. The same design was used in EEG and fMRI experiments. For the EEG experiment, inter-trial interval (ITI) varied in 100 ms random increments from 1000 to 2000 ms. For the fMRI experiment, ITI varied from 2000 to 6000 ms in 2000 ms random increments. The stimulus presentation used a discrete trial procedure so that the temporal gap between the first and the second image of a video-clip was very short (one refreshed screen) in order to simulate a continuous image sequence that created a percept of an action.

Participants received explicit instructions to observe carefully all the stimuli, and to try to decode why a familiar/meaningful action was performed. In order to avoid any saccadic movements, participants were asked to fixate on a central visual cross during the whole experiment. Before recording, every participant was familiarized with all actions for three minutes.

For the EEG study, visual stimuli were presented on a PC computer using Cogent 2000 (http://www.vislab.ucl.ac.uk/Cogent2000/index.html) running in Matlab 7.0.1 under Windows XP, which provides control of display durations and accurate recordings of reaction times. Participants were comfortably seated 150 cm away from a PC computer screen in which video clips were presented centrally. A total of five experimental blocks were presented throughout the whole experimental session. A total of 240 trials were administered which took up to a total of 40 minutes including breaks between each block.

During the fMRI recordings, visual stimuli were back-projected onto a screen at the head of the scanner bore by a standard LCD projector; participants viewed the screen using a mirror mounted on the head coil. Stimulus presentation was controlled with Psychtoolbox and Matlab 7.4.0 (R2007a) running on the Mac OS X platform. Functional images were collected in 4 discrete runs of 90 images each.

**EEG Data Collection and Preprocessing**

Continuous surface electroencephalogram (EEG) was recorded from 128 AgCl carbon-fiber coated electrodes using an Electric Geodesic Sensor Net (GSN300; Electrical Geodesic Inc., Oregon; http://www.egi.com/), where EEG electrodes are arrayed in a regular distribution across the head surface and the inter-sensor distance is approximately 3 cm. The EEG was digitized at 500 Hz (corresponding to a sample bin of 2 ms), band-width of 0.01-200 Hz, with the vertex electrode (Cz) serving as an on-line recording reference. Impedances were kept below 50k. Data logging were via NetStation Recorder. Data were collected in multiple sessions with brief intervening rest periods for the participant. Before ICA was performed, each electrode was linearly detrended, the multiple sessions were concatenated, and the data were band pass filtered between 1 and 30 Hz.

**Magnetic Resonance Imaging Recordings and Preprocessing**

Both functional and electrical neuroimaging was conducted at the UCSB Brain Imaging Center. fMRI recordings were conducted using a 3T TIM Trio Siemens Magnetom with a 12 channel phased-array head coil. Foam padding was used for head stabilization. For each functional run, an echo planar gradient-echo imaging sequence sensitive to BOLD contrast was used to acquire 33 slices per repetition time (TR) (3 mm thickness, 0.5 mm gap), with a TR of 2000 ms, echo time (TE) of 30 ms, flip angle of 90 degrees, field of view (FOV) of 192 mm, and 64 x 64 matrix. Before all the functional runs, a high-resolution T1-weighted mprage sequence image of the whole brain was acquired (TR = 2300 ms; TE = 2.98 ms; flip angle = 9 degrees, 3-D acquisition, FOV = 256 mm; slice thickness = 1.1 mm, matrix = 256 x 256).

Initial data processing was performed with SPM 5.0 (http://www.fil.ion.ucl.ac.uk/spm/). fMRI image volumes were slice time corrected, motion corrected, unwarped, spatially normalized to the Montreal 152 Average T1 atlas, and resliced to 3 mm X 3 mm X 3 mm voxel sizes. Each voxel in every imaging session was linearly detrended and then the sessions were concatenated; these steps occurred after all SPM preprocessing. The high resolution anatomical MRI was segmented into gray and white matter, warped to the 152 T1 atlas, and resliced to 3 mm X 3 mm X 3 mm voxel sizes in order to obtain a set of potential solution points.

**Independent Component Analysis**

Independent Component Analysis was performed on both the EEG data and the fMRI data, in each case using projection pursuit as implemented in FastICA (Hyvärinen and Oja, 1997, 2000). Our decompositions are similar in
spirit to those used by other investigators (Jung et al., 2001; McKeown et al., 1998) and we employ both temporal and spatial ICA (Makeig et al., 1996; Hyvärinen and Oja, 2000; McKeown et al., 1998; Stone, 2004). In the case of EEG data, the ICA decomposition is

$$E = As$$  \hspace{1cm} (1)

where each row of E is the time series of one electrode; we decomposed the full experimental session before doing any epoching or averaging. In our notation A is the mixing matrix and s is the matrix of statistically independent sources; there need not be as many sources as electrodes, but we attempted to extract as many sources as possible, up to the number of electrodes. This is temporal ICA; the statistically independent sources that are extracted are time series. The rows of s are the independent sources that mix to give the EEG data, and the columns of A give the so-called scalp maps, which represent the degree to which a source is present in each electrode.

We use spatial ICA on the fMRI data; we write the decomposition as

$$F^T = DS. \hspace{1cm} (2)$$

$F^T$ is a matrix with the same number of rows as there are fMRI image volumes, and it has as many columns as there are voxels. Here, the sources $S$ are independent spatial maps, and the corresponding column of $D$ gives temporal variation of a map. In this case the sources $S$ are statistically independent maps, not time series as in the EEG case (Stone, 2004).

Obtaining Task-Related Sources

Many of the sources returned by both temporal and spatial ICA have nothing to do with the task; hence we must sift through the components to discover those that have something to do with the task. In the EEG case, we selected components by hand, as illustrated in Fig. 3. For each source (row of $s$ in Eqn. 1) returned by temporal ICA, we constructed an epoched raster plot, seen in the top row of Fig. 3. We also considered the corresponding scalp map (column of $A$ in Eqn. 1), shown in the bottom row of Fig. 3. Task-relevant sources showed strong vertical banding in the raster plot and a generally smooth scalp map. In Figure 3, the source shown in panel A has these properties and that in panel B does not. All task-relevant sources — those that look like the source in panel A of Fig. 3 — were chosen when rebuilding the EEG data (see below). Sources like those in panel B of Fig. 3 were not used as temporal basis functions in joint inversion and were not included in stimulus-locked time-averaged EEG data. The process of source selection could be made more quantitative and automatic, and we are currently working on metrics for picking EEG sources that would make the process more similar to our process for picking fMRI sources.

The selection of fMRI components was more automatic. A stimulus function $f(t)$, having the value 1 if an image was presented that TR and zero otherwise, was constructed and its correlation with $D(t)$ (see Eqn. 2) was calculated for positive and negative lags between -20 and 20 seconds

$$C(\tau) = \langle D(t)f(t + \tau) \rangle_t. \hspace{1cm} (3)$$

The use of a lagged correlation in this case reflects the expected hemodynamic lag. The stimulus function was then randomly permuted 1000 times and lagged correlations recomputed; this gave a mean noise correlation and its standard deviation (denoted by the dotted blue lines in Figure 4). Any component that showed above-noise correlations at negative lags — corresponding to the response following the stimulus — was then selected as task-related. In all cases for both fMRI and EEG data, the number of task related sources was much smaller than the total number extracted (a few as opposed to hundreds), so ICA served...
The methods we then time averaged \( \tilde{v} \) voltage matrix \( V \) facts and task-unrelated activity removed. To obtain the matrix of basis functions \( \varepsilon \) used below (see the Inversion Algorithms section), we time averaged the components \( s_i \), locked to stimulus presentation, and then smoothed them. This process ensures that we will avoid overfitting the noise in the EEG data; a linear combination of the smoothed sources cannot perfectly fit the data in \( V \), whereas a direct ICA decomposition of \( V \) could reconstruct \( V \) with zero error. Our smoothing filter is a multitaper (Percival and Walden, 1993) one that is commonly used in geophysical research (Flagg et al., 1976; Limeburner et al., 1983); it attempts to balance frequency ringing from filters of finite (temporal) support with the reduction in peak amplitudes from infinite (temporal) support filters like a Gaussian. The filter width was 32 milliseconds.

Rebuilding the fMRI data proceeded in a manner very similar to that for EEG data; the task-related bold data \( \tilde{F} \) was built from task-related fMRI components as in Eqn. 4, using the appropriate rows and columns of the \( D \) and \( S \) matrices in Eqn. 2. Explicitly, if there were \( n \) task-related fMRI sources, labeled \( \{j_1, j_2, \ldots, j_n\} \) we computed

\[
\tilde{F}^T = [D_{j_1}^T \ D_{j_2}^T \ \ldots \ \ D_{j_n}^T] \begin{bmatrix} S_{j_1} \\ S_{j_2} \\ \vdots \\ S_{j_n} \end{bmatrix}.
\]

This data was then time averaged, again locked to stimulus, to obtain a matrix of BOLD data \( B \). The matrix \( B \) has as many rows as there are fMRI voxels and as many columns as there are time samples in the epoch chosen for averaging. However, in most cases we employed not \( B \) but a set of weights \( w \), sometimes arranged in a diagonal matrix \( W \) (with \( w \) on the diagonal). To construct \( w \) we found all the spatial maps carrying task-related information, and then formed \( w = \sum_{j_1, \ldots, j_n} S_j \). We subsequently normalized \( w \) so that its largest element equals unity.

There is a symmetry in the way the fMRI and EEG results are treated; in each case the mixing elements were used in selecting task-related components, but then that information was discarded, and only the sources were kept. In the case of fMRI, all task-related sources were collapsed into one vector \( w \), rather than maintaining them separately as in the EEG case. If the experiment were truly recorded simultaneously (EEG measurements inside an MRI scanner) we might be able to associate EEG sources with fMRI sources, but we have no confidence in our ability to do so when the imaging is performed in separate sessions, nor do we necessarily expect such a correspondence in this situation.
Lead Field Calculation

Electrode locations were measured using an infrared tracking system (Northern Digital), and put in registry with the participant’s structural MRI by using anatomical fiducials also measured with the TMS tracker (outer canthus of the eyes and the tip of the nose). These electrode positions were then warped into template coordinates using SPM 5, and the subsequent warped positions used for all subjects. Three-shell spherical lead fields were calculated using Berg’s method (Berg and Scherg, 1994), employing the BrainStorm toolbox for MATLAB (http://neuroimage.usc.edu/brainstorm/). Each shell was assumed to have a different isotropic, uniform conductance. Gain matrices obtained from lead field calculation were depth weighted (using electrode power) to discount superficial sources (Köhler et al., 1996).

A subset of gray matter voxels obtained from anatomical image segmentation were chosen to be solution points. Only voxels with gray matter intensity above a threshold of 75% of maximum were chosen, in order to restrict the size of the solution grid. The size of the solution grid was further decreased by selecting a desired number at random. Dipole orientations were chosen to be radially outward from the center of the head; hence all solution methods yield a scalar (rather than a three-component vector) at each solution point, and the lead field gain matrix reflected this choice of dipole orientation.

Inversion Algorithms

Inferring which sources in the brain produced measured EEG signals is a problem from classical electrodynamics; we wish to obtain the dynamic charge distribution in the interior of a set of roughly spherical conducting shells using only measurements on the surface of the outermost shell. Given that classical electrodynamics is a well-understood theory, this seems straightforward but is hardly so. One can represent the forward problem as

\[ V = GX, \]  

where \( V \) is the matrix of electrode voltages (all electrodes, all times), \( G \) is the linear Green’s function which propagates the interior sources to the scalp, taking account of geometry and material properties of cortex, cerebrospinal fluid, skull, and scalp, and \( X \) is the desired matrix of sources (Hallez et al., 2007).

Formulated as a least squares problem, the desired matrix of sources minimizes the following cost function:

\[ f_{\text{EEG}}(X) = \frac{1}{2} \text{Tr} \left\{ (V - GX)^T (V - GX) \right\}. \]  

(7)

\( G \) has a huge null space; we would like source information over the entire cortex — tens of thousands of locations, depending on desired resolution — at all times, but given typical EEG electrode montages we have at most around two hundred correlated spatial measurements to constrain the solution. While anatomical information (Phillips et al., 2002) or careful consideration of the types of bulk currents that produce EEG signals (de Peralta Menendez et al., 2000) can further constrain the problem, all EEG-only mapping methods must eventually impose regularization constraints to give a unique \( X \). These constraints are usually chosen to impose desirable properties on the solution, like that it be of minimum norm (Hämäläinen and Ilmoniemi, 1994) or maximum smoothness (Pascual-Marqui et al., 1994), or have a particular covariance structure (Sekihara et al., 2001). For a recent review of EEG-only source localization see Grech et al. (2008).

We incorporate fMRI information because it has complementary strengths and weaknesses to EEG; it has high spatial resolution but very poor temporal resolution, on the order of several seconds. The single term in Eqn. 7 fits only the observed EEG electrode voltages \( V \), and we will always represent the EEG portion of the cost function this way. In order to incorporate fMRI information we will modify the cost function in Eqn. 7 by adding terms constructed to fit fMRI data. There are many ways to do this, and comparing and contrasting some of them are the focus of the rest of this manuscript.

We summarize the joint cost function approach with the following “master” cost function, which we wish to minimize:

\[ C(X) = \frac{1}{2} \text{Tr} \left\{ (V - GX)^T \Sigma_V^{-1} (V - GX) \right\} + \frac{\mu_1}{2} f_{\text{fMRI}}(X) + \frac{\mu_2}{2} f_R(X). \]  

(8)

The first term represents the fit to the EEG data \( V \), and it is identical to Eqn. 7 except for the addition of \( \Sigma_V \), discussed below. The second term is the fit to fMRI data, and the third is a regularizing term. These final two terms are written less explicitly than the EEG term because they will have different representations in different joint inversion methods. However, whatever form \( f_R(X) \) and \( f_{\text{fMRI}}(X) \) take, both must be a function of the brain sources \( X \) (these are not the ICA sources) in order to relate the two imaging modalities and best constrain the parameters \( X \) using all available experimental measurements.

\( \Sigma_V \) is the electrode-electrode covariance matrix. Scaling by this matrix serves two purposes; the off-diagonal elements ensure that we do not overweight correlated data points, and the on-diagonal terms ensure the relative contributions to the cost function of terms with very different units is similar. The parameters \( \mu_1 \) and \( \mu_2 \) control the relative weights of the various terms; we include them for completeness, but we generally only consider equal weighting (\( \mu_1 = \mu_2 = 1 \)) in this manuscript. Detailed descriptions of each method follow, and a table summarizes them (see Table 1).

Minimum Norm (MN) Minimum norm is a well-known method for regularizing underdetermined least squares problems (Wunsch, 1996). It is a widely used
method to solve the EEG-only source localization problem (Hämäläinen and Ilmoniemi, 1994). Minimum norm simply seeks a solution in which the source matrix $X$ is as small as possible, in the sense of having minimum $L_2$ norm. The second term in Eqn. 8 is not present, and the regularization term becomes

$$f_R(X) = \text{Tr} \left\{ X^T X \right\},$$

which is the correct expression for the sum of the squares of the elements of the matrix $X$. Despite, or perhaps because of, its simplicity, some investigators continue to argue for the use of minimum norm estimation (Hauk, 2004).

ICA Only (ICAO) This is a modification of the minimum norm procedure. Instead of solving for the full matrix $X$, which has row dimension equal to the number of fMRI voxels and time dimension equal to the number of sampled EEG voltages, we posit that all source activities can be written as a linear combination of our previously described ICA-derived EEG basis functions, where

$$X = \alpha \varepsilon,$$

with $\varepsilon$ the averaged, smoothed, task-related EEG sources and $\alpha$ a matrix whose column dimension is much smaller than the number of times at which EEG data is collected. We refer to $\alpha$ as the loadings, as they are static spatial weights. The temporal dynamics of the solution are contained in $\varepsilon$. In any method employing this decomposition, all the functions in Eqn. 8 become functions of $\alpha$ instead of $X$.

The first term in Eqn. 8 may be written as

$$f_{\text{EEG}}(\alpha) = \frac{1}{2} \text{Tr} \left\{ (V - G\alpha)^T \Sigma_V^{-1} (V - G\alpha) \right\}.$$  

This equation is implemented here and in all subsequent methods which use ICA decomposition of the source activity. The regularization term in this case may be written as $f_R(\alpha) = \text{Tr} \left\{ \varepsilon^T \alpha^T \alpha \varepsilon \right\}$. We potentially gain something with this technique even though it is still EEG only; the use of ICA to generate a set of independent basis functions allows us to spatially separate the dynamics of independent cognitive processes, insofar as ICA is able to detect them in the EEG data.

$W$-scaled Gain Matrix (WGM) In this method, we column-scale the lead field by the diagonal matrix $W$ obtained from ICA of fMRI. This has the effect of suppressing solution intensity from solution points where no significant task-related fMRI activity was found. Mathematically, there is still no direct fMRI term, but a modified gain matrix $\hat{G} = GW$ is used for inversion. Replacing $G$ by $\hat{G}$ in Eqn. 11 gives the EEG fitting term for this method; no additional fMRI or regularization terms are present.

$W_2$-regularized ICA (W2ICA) This method still has no direct fMRI term; fMRI information is included in the regularization term. The form of the third term in Eqn. 8 is

$$f_R(\alpha) = \text{Tr} \left\{ \alpha^T W^{-1} \alpha \right\}.$$  

To understand the purpose of this term, note that it is similar to the minimum norm constraint in Eqn. 9 except that the loadings $\alpha$ (rather than the full source activity) are being constrained and $W^{-1}$ is present. Multiplication by $W^{-1}$ has the effect of increasing the penalty term for solution points which show little task-related activity (small $W_{ii}$, hence large $1/W_{ii}$). The less fMRI activity found at a solution point, the bigger a penalty this term introduces, and hence the more its activity will be suppressed in the inverse solution.

$W_1$-regularized ICA (W1ICA) This method is very similar to W2ICA except for the form of the regularization term

$$f_R(\alpha) = |W^{-1}\alpha|$$

which uses the $L_1$ norm rather than $L_2$. The $L_1$ norm has been used with great success in many regularized optimization problems, and in Bayesian statistical problems would correspond to a Laplacian rather than the more common Gaussian prior.

Operationally, the $L_1$ norm is often used for finding sparse solutions to inverse problems because the force on the parameters (derivative of the cost function) coming from the regularization term is constant rather than linear with $X$ as in the $L_2$ case. No matter how small a parameter becomes, the $L_1$ term continues to push it towards zero until it is identically zero. $L_2$ terms become less and less effective as the parameters shrink, so it is difficult to obtain truly sparse solutions with them. Solving this optimization problem requires an iterative (though deterministic) algorithm for a matrix of parameters based on that previously described for a vector of parameters (Alliney and Ruzinsky, 1994). At a significant memory cost this minimization problem can also be formulated as a quadratic program.

Power Constrained (PowR) This is the first of two methods which explicitly fit fMRI data, utilizing the second term in Eqn. 8. This fMRI fitting term takes the form

$$f_{\text{fMRI}}(\alpha) = \frac{1}{\sigma_P^2} (\lambda P - P(\alpha))^T (\lambda P - P(\alpha)).$$

The vector $P$ is the observed fMRI signal power at each solution point; its $i^{th}$ element is computed from the reconstructed BOLD data $B$ as

$$P_i = \sum_t B_{it}^2.$$
\( P(\alpha) \) is the calculated solution power, which for parameters \( \alpha \) is
\[
P_1(\alpha) = \begin{bmatrix} \alpha C \alpha^T \end{bmatrix}_{\alpha},
\]
where \( C = \varepsilon \varepsilon^T \). The matrix \( C \) would be diagonal if \( \varepsilon \) came directly from ICA of the averaged EEG data (due to independence of the components), but our full-session EEG ICA and smoothing process introduces off-diagonal elements. The number \( \sigma^2_{\varepsilon} \) is the variance of the distribution of voxel power, which gives an appropriate scaling for the fMRI term relative to the EEG fitting term. The scalar \( \lambda \) absorbs the unknown conversion between BOLD units and solution intensity (essentially local field potential); it can be computed by setting \( \partial C / \partial \lambda = 0 \) and does not require a separate minimization step.

We seek to match scalp dynamics via EEG, and dynamics in the bulk brain by comparing solution power to observed voxel power from fMRI measurements. This algorithm must be solved via an iterative method; we use Møller's scaled conjugate gradient algorithm (Møller, 1993). While the cost is no longer simply quadratic, it is not highly nonlinear - merely quartic. Also, an analytical gradient is available making derivative-based methods accurate and easy to implement.

**Model Reduced Joint Inverse (MRJI)** This method is described in more detail elsewhere (Brookings et al., 2009), but we summarize it here. The cost function contains both EEG and fMRI terms and is
\[
C(\alpha, \lambda, \kappa) = \frac{1}{2} \text{Tr}\left\{ (V - G\alpha^T)^T (V - G\alpha^T) \right\} + \frac{g}{2} \text{Tr}\left\{ (\lambda B + \kappa - \alpha^T \beta)^T (\lambda B + \kappa - \alpha^T \beta) \right\}.
\]
This is the entire cost function; it replaces the first two terms of Eqn. 8 and contains no third \( f_R(\alpha) \) term. The goal is to treat fMRI and EEG as symmetrically as possible; both types of data are fit to dynamical models. \( \beta = \beta(\varepsilon) \) is a matrix of fMRI sources, analogous to the EEG sources. These are obtained by feeding each EEG source (row of \( \varepsilon \)) into a balloon-type (Buxton et al., 1998) model for BOLD dynamics (Robinson et al., 2006). The \( \beta \) act as basis functions for attempting to fit the BOLD activity \( B \) in the way \( \varepsilon \) are used to fit the EEG dynamics. So in the same way that we assume the local source activity can, via propagation of the lead field, describe measured EEG voltage, we assume that fMRI activity can be written as a linear combination of fMRI basis functions, corresponding to computed BOLD responses to the sources \( \varepsilon \).

The details of the solution in this case and the performance of the method on synthetic data are contained in Brookings et al. (2009). This cost function is entirely quadratic in the parameters (\( \lambda, \alpha \) and \( \kappa \)) allowing us to find a (global) minimum using only linear algebra. \( g \) is a gain term that attempts to make the EEG and fMRI terms in the cost function of the same magnitude; we use the ratio of the average electrode variance to the average voxel variance. Note there is no covariance weighting in the function; see the Discussion section for further comments on this issue.

**Results**

In this section we compare results for the inversion methods described above. We have chosen the visual evoked response to test our methods because it generates a strong cortical signal and much is already known about its dynamics and localization. To wit, we expect up to three visual evoked potentials — an early peak and a late single or double peak — appearing in the occipital cortex due to a visual stimulus. The reader should keep this in mind when considering the results presented below.

Figures 3 and 4 show implementation of ICA on full-session human EEG (Fig. 3) and fMRI (Fig. 4) data and give examples of which components were included in mapping and which were not. Figure 3 shows two ICs from one participant’s EEG data; both panels show a raster plot of the trial-by-trial source activity and the components scalp map, displayed as a Lambert equal-azimuthal-area polar plot. In the raster images stimuli are marked with heavy black arrows. One sees a clear dipolar scalp map for the component in panel A, and the raster plot shows good synchronization to the stimuli across trials (note the strong vertical banding in the raster plot). This source, and others with similar characteristics, were used in generating \( V \) and \( \varepsilon \). Panel B shows an artifactual or noise component; judging by the raster plot it shows little synchronization with the experimental stimuli and has a nonsmooth scalp map dominated by loading onto only a few electrodes. This component, and others like it, were not used in joint inversion.

Figure 4 shows similar information for ICA of fMRI data; it shows both a representative slice from the head map obtained via ICA, along with a plot of the lagged stimulus correlation (described in Materials and Methods). In panel A we show a task-related map. The intensity in the head map is occipitally localized, as expected for a visual stimulus, and the correlation function shows a strong, significant peak near a negative lag of 6-8 seconds. This indicates that the activity of the head map is most strongly correlated with the stimulus 6-8 seconds after presentation, a reasonable value for the hemodynamic lag. This head map and others like it were used for generating the weight vector \( w \) and constructing \( B \). As in the EEG case, we also show an artifactual or noise component in panel B. The noise head map shows no clear domains of activity; there is instead a uniform speckling. Most importantly, the
lagged correlation function shows nothing of significance. Components of this type were not used in joint inversion.

Figures 5, 6, and 7 show the results of the inversion methods described in Methods for the evoked visual response data. Figure 5 shows a set of summary plots for each algorithm. The first panel indicates the ability of each inversion method to reproduce observed EEG voltages; hence rows of the matrix \( V \) are plotted simultaneously with rows of the matrix \( GX \) (for the MN method) or the matrix \( G_{\alpha \varepsilon} \) (for all other methods, see Eqns. 7,10). The second panel shows the agreement between observed fMRI activity, in the form of the weight vector \( w \) obtained from ICA of the BOLD data, and \( P(\alpha) \), the square root of the power computed at each solution voxel (see Eqn. 16). For the MN method, \( X \) is not decomposed as in Equation 10, so the solution power in this one case is computed exactly as in Eqn. 15, except with the matrix \( X \) in the place of \( B \). The final panel shows a representative slice for a single participant with solution power superimposed on subject anatomy. The size of each circle corresponds to solution power at that voxel, and voxels with very little solution power have been omitted for the sake of clarity. These anatomical plots indicate if significant solution power has been placed in physiologically reasonable locations.

There are several notable features of Figure 5. First, by scanning the first column one can see that all the inversion methods used fit the EEG data quite well. This is true for the entire head and not just representative electrodes. One can obtain a global goodness of EEG fit by looking at the EEG portion of the cost function at the best parameters. The MRJI fits almost perfectly due to lack of covariance weighting in the EEG fitting term; we discuss this below. W1ICA fits the EEG data slightly worse than the other methods, particularly at electrode T5. More striking is that similar EEG quality of fit is obtained for drastically different spatial distributions of solution power, as one can see by examining columns two and three. In the extreme, W1ICA was able to obtain an excellent fit to the EEG data by placing solution intensity at less than 10 of 4000 voxels!

One should also note that the WGM, W2ICA, and PowR methods all placed significant power in physiologically relevant regions. These results illustrate the under-determined nature of the EEG-only problem and suggest there may be no single correct way to increase mapping resolution. We argue below that one needs to think about the interplay between joint inversion, experimental design parameters, and robustness issues when combining multiple imaging modalities for improved resolution in whole-brain mapping.

Figure 6 shows the solution, rather than just power, for four methods we have discussed. Minimum norm has been
Figure 5: Solution diagnostics for the seven inversion methods described in the text, for one participant in the visual evoked response experiment. For each method, three panels are shown. The left column displays the quality of fit to the EEG data; the inverse solution is shown in red and the data in black, for four electrodes. They are, from top to bottom in the 10/20 naming system, P4, T5, O1, and F8. The voltage traces have been offset to allow for easier viewing. The second panel is a log-log plot of $w$ (weights obtained from ICA of fMRI as described in the text) on the $x$ axis and the square root of voxel power calculated from the inverse solution on the $y$ axis. Hence we are plotting the power in the (fMRI) data (the $x$ axis) versus the power in the model (the $y$ axis). The red dotted line has unit slope; the amount of point scatter about this line shows the degree of disagreement with measured fMRI voxel power. The third column shows a representative anatomical slice with dots with size proportional to solution power at that voxel. In these plots, voxels with power less than 25% of the most powerful voxel have been omitted for clarity. Whenever possible we have chosen the same slice for all methods, but in the case of the MRJI and W1ICA that was not possible, as very little if any solution power was present in the slice shown for the other methods. Method abbreviations are indicated at right, and a quick reference is contained in Table 1.
left out because we here show $\alpha$ and $\varepsilon$ separately and basic minimum norm admits no such decomposition as part of its solution process. The ICAO method is essentially pre-decomposed minimum norm and is shown. We also do not display the MRJI or W1ICA, as their results in Figure 5 did not look promising. For each method in Figure 6 we show the individual rows of $\varepsilon$ along with corresponding weights $\alpha$, plotted as a scatterplot on subject anatomy. Negative loadings are in cyan and positive loadings in magenta. The same slice is shown for each signal and all methods.

We now compare and contrast the various inversion schemes presented. First, it is clear that ICA of EEG data alone does very little to condition the inverse problem; while it is valuable to pre-split temporally independent neural events before inversion mapping, the regularization problem remains, and minimum norm solutions produce substantial power at most solution voxels. Among the remaining methods, W1ICA is an interesting and surprising case, in that it was able to accurately reproduce the paired visual stimulus into separate components for each image. However, for more complicated experimental paradigms in which we expect independent responses — for example visual evoked response followed by a motor response — ICA should be able to produce one component for each which, upon joint inversion, will show distinct, physiologically relevant localization.

Four participants who performed the visual evoked response experiment were selected, and Fig. 7 shows the temporal and spatial characteristics of their responses when computed using the PowR joint inversion method. First notice that the number of basis functions (rows of $\varepsilon$) comprising the solution varied from subject to subject. There is no guarantee that ICA will find the same number of task-related EEG components in different individuals, and we see that this is the case here. This statement is also true of spatial ICA of fMRI data; while all components for each participant have been collapsed into a single spatial map (not shown), the number of task-correlated components was variable among individuals. Figure 7 shows intersubject differences in both activity timing and localization. However, common to all the subjects are single and double peaks following stimulus presentation, with significant occipital loading. These results show that joint inversion could be a powerful tool for probing intersubject variability.

**Discussion**

We now compare and contrast the various inversion methods described in the text: ICAO, WGM, W2ICA, and PowR. The top row of the plot shows the rows of $\varepsilon$, the matrix of temporal basis functions; heavy black arrows indicate stimulus presentation times. In the anatomical plot negative loadings (elements of $\alpha$, the parameters obtained by inversion, see Eqn. 10) are shown in cyan and positive loadings in magenta. The same slice is shown for each signal and all methods. To obtain the full solution for all solution voxels and all times, one constructs $\alpha \varepsilon$ which essentially multiplies the signal in the first row of the figure by the head map below it and then sums over the figure row.

Notice that the loading maps are relatively smooth, despite having no explicit smoothness constraint in the cost functions. All the solution smoothness in WGM, W2ICA, and PowR comes directly from the smoothness of the activity map obtained from fMRI. This is a particularly nice feature, as the solution smoothness found is a side effect of fitting both modalities well, rather than a condition imposed by fiat. We also note that all three components used in the task-only EEG average (the top row of Figure 6) have similar but not identical dynamical structure and localization. This is expected; the dominant response to the two successive stimuli will be the evoked visual response, and we do not expect nor desire that ICA will oversplit the paired visual stimulus into separate components for each image. However, for more complicated experimental paradigms in which we expect independent responses — for example visual evoked response followed by a motor response — ICA should be able to produce one component for each which, upon joint inversion, will show distinct, physiologically relevant localization.

However, common to all the subjects are single and double peaks following stimulus presentation, with significant occipital loading. These results show that joint inversion could be a powerful tool for probing intersubject variability.

To obtain the full solution for all solution voxels and all times, one constructs $\alpha \varepsilon$ which essentially multiplies the signal in the first row of the figure by the head map below it and then sums over the figure row.

Notice that the loading maps are relatively smooth, despite having no explicit smoothness constraint in the cost functions. All the solution smoothness in WGM, W2ICA, and PowR comes directly from the smoothness of the activity map obtained from fMRI. This is a particularly nice feature, as the solution smoothness found is a side effect of fitting both modalities well, rather than a condition imposed by fiat. We also note that all three components used
Figure 7: Joint inverse solution using PowR for four subjects participating in the dual evoked visual response experiment. Each basis function along with its loadings in a representative slice are shown, as in Figure 6. Small loadings have again been omitted for clarity. Each row, labeled S1–S4, shows a different subject, and the columns pair a basis function with the corresponding loadings in a representative slice, as in Figure 6. The grouping is not arbitrary. Subject S1 had only two task-related components. His/her entire solution is shown in row S1, and for the other subjects, we show the two basis functions most correlated with S1’s directly below (grouped with a dotted box). Continuing with S2, in which we found three basis functions, we then vertically group this component with the component in the remaining subjects most similar as again judged by correlation. Heavy black arrows in row S4 indicate stimulus onsets (first frame, second frame); these onsets are the same for all basis function panels but suppressed to avoid clutter.
Also too extreme in our estimation is WGM; rather than trading quality of fit to EEG data with quality of fit to fMRI data, WGM essentially assumes fMRI maps report all relevant activity and imposes a crushing penalty on signals placed at voxels without strong task-related fMRI activity. This is reflected in the extremely tight correlation between measured fMRI power and resulting solution power (see the second column of Figure 5). W2ICA showed a similar phenomenon, although slight differences may be discerned; in general, inversion with W2ICA tends to place less power at each solution point than indicated by fMRI. However, it still shows an uncomfortable degree of confidence in the fMRI data and would not easily allow sources that one can detect more easily with EEG than fMRI, due to signal-to-noise issues.

MRJI was somewhat of a surprise. MRJI performed extremely well in tests on synthetic data (Brookings et al., 2009), but was far less successful with human subject data. This underperformance comes about for several reasons. For one, the individual rows of $\beta$ are really all identical in this case: the lowpass filtering nature of the hemodynamic response makes events separated by less than several seconds unresolvable when used to force a Balloon/Windkessel model. In essence, rather than using a sum of distinct basis functions to fit the data in $B$, we are really only fitting a rescaled hemodynamic response function (HRF) to the time-locked fMRI average data. If the time-averaged voxel dynamics deviate from this simple shape, the fMRI term in the MRJI will not model them well.

This brings up the crucial subject of the interaction of inversion method with experimental design; this fMRI experiment was a rapid-event related paradigm, in which the responses to successive stimuli “pile up” in the measured fMRI data. Upon averaging, we do not recover simple rescalings of the HRF in the data. However, in a slower event related design, where the hemodynamic signal is allowed to return to baseline after each stimulus presentation, this may not be true and the MRJI may approach its performance on simulated data (in which the assumed forward models generated the data, so a better fit was essentially ensured). It is therefore extremely important to consider experimental design when choosing how to improve mapping resolution using joint inversion; this is an issue which we are addressing with ongoing empirical studies.

Lack of similarity between the fMRI model employed and the observed data is not the only complication for the MRJI. Another comes when we try to properly weight the two terms in the cost function. Covariance weighting in least-squares cost functions performs two essential functions: it ensures we do not give ourselves more credit than we should when we fit correlated data, and it introduces a scaling term that can accommodate data in drastically different units, since the standard deviation should generally be of the same order of the mean and at least has the same units. While it is simple to include electrode-electrode covariance weighting in the MRJI, voxel-voxel covariance weighting poses a challenge. We cannot simply insert $\Sigma_B^{-1}$, the full voxel-voxel covariance matrix, into Equation 17 in the obvious place; this is because of the parameter $\lambda$. Note that as written, $\lambda$ rescales the data $B$ to best match the model, for any choice of $\alpha$. However, rescaling the data by $\lambda$ rescales $\Sigma_B^{-1}$ by $\lambda^{-2}$, and doing so destroys our ability to use the linear-algebraic methods described previously (Brookings et al., 2009). For this reason, one would like to leave $B$ alone and rather rescale $\alpha \beta$, which would adjust the units on the model to best fit the experimental units in the data. This again destroys our ability to use linear algebraic methods to solve the problem, as there are now terms quartic in the parameters present in the cost function $(\lambda^2 \alpha^T \alpha)$. We have used iterative methods to attempt to solve the problem in this case, but they prove particularly inefficient, partly owing to the fact that $\alpha \equiv 0$ is not a valid initial starting solution — $\lambda$ becomes undefined in this case. While the use of $g$ in Eqn. 17 tries to remedy the units problem, it does nothing to ensure that we do not overfit correlated fMRI data.

We should also mention that while placement of $\lambda$ can ensure a quadratic cost, there are several additional parameters involved in the generation of $\beta$. These additional parameters are present regardless of whether one uses a fully nonlinear model for $\beta$ or a linear time-invariant one. One would like to estimate these parameters or integrate them out as Bayesian nuisance parameters (Sivia and Skilling, 2006), but this would come at a significant increase in computational cost. We continue to work on variants of the MRJI that address these difficulties, but for the type of rapid event-related experiment considered here we do not recommend its use.

We therefore recommend for this application the PowR method. It requires iteration, but the all-zero starting guess gives a very efficient estimation with scaled conjugate gradient. The cost function is quartic rather than quadratic, but this is a relatively mild nonlinearity. Figure 5 shows that the PowR method yielded solution power that is well correlated with fMRI activity, particularly when considering the most active voxels, but which is not absolutely coincident with fMRI measurements. This is due to the explicit tradeoff created by simultaneously fitting the temporal dynamics of EEG with the spatial structure of fMRI. We also suggest that PowR attempts to take advantage of the best features of each data modality; it fits the temporal dynamics of EEG and the spatial dynamics of fMRI to obtain one high-resolution map.

In going beyond the present study, issues of experimental design and data resolution are key. Slow versus fast event-related designs may require different joint inversion methods and we are currently conducting further empirical studies to address this issue. Also important are issues of data resolution; we have used a simple three sphere head model in this study, but detailed anatomical head models using cortical and scalp surface extraction from structural MRIs exist (Spinelli et al., 2000; Ermer et al.,
How much descriptive power is gained by using an anatomically detailed head model, and how does this choice interact with the signal-to-noise ratio in the EEG data? The EEG electrodes are also highly spatially correlated, suggesting the use of a lower-density EEG montage would result in the same quality of inverse solution. In addition, the EEG data could be temporally downsampled and the fMRI data spatially smoothed and/or resliced at lower resolution. All of these choices affect both the size of the subsequent joint inverse problem (and hence solution efficiency) as well as the nature of the solution obtained. The effect of choices in the data processing pipeline is a complex and active area of study, even for single imaging modality data (Strother et al., 2004; Strother, 2006). We would like to know what solution features are robust to pipeline choices in the context of EEG/fMRI joint inversion and are currently studying this issue.

The next phase of cognitive neuroscience is to go beyond studying local brain regions and to begin to learn about the global, distributed networks in the brain underlying cognitive activity (Bullmore and Sporns, 2009). In order to understand the dynamics within and between such networks, measurements with both high temporal and spatial resolution are essential. In addition, as researchers move towards more complicated, naturalistic stimuli (Malinen et al., 2007) in an effort to understand complex decision making, ever more sophisticated techniques for extracting relevant information from neuroimaging datasets will be required. Pushing EEG/fMRI fusion techniques to the limit will optimize our ability to extract information from these noninvasive, proxy measurements of the neuronal basis of cognition.

Acknowledgements

K. S. B. thanks T. Brookings for useful discussions, and A. Johnson for helping in the preparation of Figure 7. This work was supported by the David and Lucile Packard Foundation and the Institute for Collaborative Biotechnologies through contract no. W911NF-09-D-0001 from the U. S. Army Research Office.

References

McKeown, M. J., Maleig, S., Brown, G. G., Jung, T.-P., Kinder-


