Using ICA and realistic BOLD models to obtain joint EEG/fMRI solutions to the problem of source localization

Ted Brookings[1], Stephanie Ortigue[2], Scott Grafton[2*], Jean Carlson[3]

[1] *Department of Biology, Brandeis University,*
Waltham, MA 02454, USA
telephone: 781 736 3141; email: brookings@brandeis.edu
[2] *Department of Psychology, University of California,*
Santa Barbara, 93106, USA
telephone: 805 893 3045; email: grafton@psych.ucsb.edu
[*] *corresponding author*
[3] *Department of Physics, University of California,*
Santa Barbara, CA 93106, USA
telephone: 805 893 8345; email: carlson@physics.ucsb.edu
(Dated: February 29, 2008)

We develop two techniques to solve for the spatio-temporal neural activity patterns using Electroencephalogram (EEG) and Functional Magnetic Resonance Imaging (fMRI) data. EEG-only source localization is an inherently underconstrained problem, whereas fMRI by itself suffers from poor temporal resolution. Combining the two modalities transforms source localization into an overconstrained problem, and produces a solution with the high temporal resolution of EEG and the high spatial resolution of fMRI. Our first method uses fMRI to regularize the EEG solution, while our second method uses Independent Components Analysis (ICA) and realistic models of Blood Oxygen-Level Dependent (BOLD) signal to relate the EEG and fMRI data. The second method allows us to treat the fMRI and EEG data on equal footing by fitting simultaneously a solution to both data types. Both techniques avoid the need for ad hoc assumptions about the distribution of neural activity, although ultimately the second method provides more accurate inverse solutions.

I. INTRODUCTION

Electroencephalogram (EEG) and Functional Magnetic Resonance Imaging (fMRI) are two commonly used modalities for investigating human brain states in cognitive neuroscience experiments. Both are noninvasive, but in other respects they are complimentary. EEG measures voltage changes in roughly one hundred electrodes placed on the scalp. EEG has millisecond time sensitivity, but spatial information must be inferred through an inversion process, and has at most as many independent spatial measurements as there are electrodes (there may be fewer due to correlations between nearby electrodes). fMRI measures changes in blood oxygen level [1, 2] (called the BOLD signal) throughout the brain. It produces a 3D image with a spatial resolution of roughly a few millimeters, but temporal resolution is on the order of a few seconds. Furthermore the BOLD signal is a complicated convolution of brain activity because the blood oxygen level takes several seconds to rise and even longer to fall in response to an impulse of activity. Thus EEG provides an excellent measure of temporal dynamics but a poor measure of spatial locations; and fMRI provides an excellent measure of spatial locations but a poor measure of temporal dynamics.

In this paper we develop two novel methods for source localization using both EEG and fMRI data. By combining the two modalities, the high temporal resolution of EEG can be augmented with the high spatial resolution of fMRI. Existing literature has established the potential gains from combining EEG and Positron Emission Tomography [3] (another imaging modality with high spatial resolution), as well as EEG and fMRI [4]. However, in past studies the difficulties inherent in combining such dissimilar modalities have led to a reduced scope of the analysis: inclusion of data from only one EEG lead [5], or reduction of the size of a unified space for joint inversion of EEG and fMRI [6]. Here we present two techniques for working with full EEG and fMRI data sets and solving to obtain neural activities throughout the cortex at high spatial and temporal resolution.

Our first method uses standard techniques to invert EEG data, but employs fMRI data to constrain the solution instead of more ad hoc regularization schemes frequently employed. Throughout this paper, we refer to this technique as our "fMRI regularized inverse", and we describe it in Section III. The second method uses model reduction algorithms (Principle Component Analysis, or PCA; and Independent Component Analysis, or ICA) to decrease the size of the inverse problem, and a detailed model of the BOLD signal (discussed in Section IV) to relate EEG and fMRI data. This enables us to simultaneously fit the EEG and fMRI data. We refer to this method as our "model-reduced joint inverse", and it is described in Section IV. The model-reduced joint inverse has the additional advantage of treating the EEG and fMRI data on equal footing, instead of using the fMRI merely as a constraint. In Section V, we evaluate and contrast the effectiveness of these techniques on synthetically generated data to demonstrate the potential

effectiveness of using this methodology to analyze data recorded from human subjects.

II. EEG SOURCE LOCALIZATION

We begin with a brief, general description of EEG source localization [7]. This background provides the starting point for our first method of combining EEG and fMRI (Section III) which begins with EEG, but makes use of the fact that the basic problem of EEG source localization is underconstrained. When EEG is considered alone, different methods employ different techniques for *regularization*, —selecting a particular solution out of an infinite family of possibilities, which would otherwise share an equal quality of fit. In Section III, we introduce a specific method of regularization which uses complementary fMRI data. Our solution optimizes agreement with the fMRI data without altering the quality of fit to EEG data.

To perform source localization and thus obtain spatial information from EEG data, one must model the effect that currents within the brain produce on voltages measured at the scalp. Sources in different regions of the brain contribute to electrode voltages by summing linearly. Thus, given a set of source currents, the measured EEG signal can be calculated with a linear Green's function matrix (or *lead field*).

This matrix is typically calculated by modeling the head as a series of concentric regions with different conductivities, and solving Maxwell's equations for dipole current sources. Often the head model is quite simple, although it is possible to incorporate 3D images from a structural MRI (as in Figure 1) of the subject's head. More complicated head models may produce a more accurate lead field matrix, but the resulting matrices are used the same way.

Once the lead field L has been calculated, if the neural sources $\vec{S}$ are known, then the expected EEG signals $\vec{E}$ can be computed simply through the linear *forward model*

$$\vec{E} = L\vec{S}. \tag{1}$$

Source localization is then a matter of inverting the linear equation relating the unknown source currents to the known EEG signal via the Green's function. Therefore for some generalized inverse L^-, we have

$$\vec{S} = L^-\vec{E}. \tag{2}$$

This system is highly underdetermined. If the brain is divided up into regions similar in size to fMRI resolution, there are tens of thousands of voxels, but there are only of order one hundred EEG signals to constrain the solution. With no additional constraints, it is possible to add to any solution a set of sources in the null space
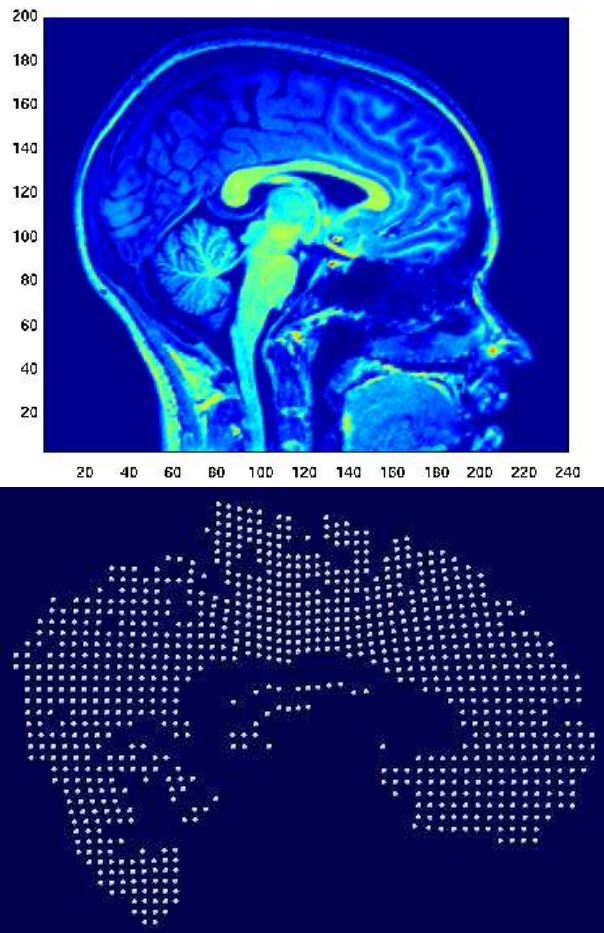

FIG. 1: **Lead Field from Structural MRI.** A structural MRI (top) captures a high resolution 3D image of the brain. This image can be used to identify regions of grey matter within the cortex, which are likely locations for EEG sources (bottom).

of the Green's function and obtain another valid solution. Although it is possible to decrease the extra degrees of freedom by appealing to physiology[8] or careful treatment of the electric field (as in ELECTRA[9] and LAURA[10]), ultimately the EEG data cannot produce a unique solution by itself.

To produce a unique solution, the problem must be regularized. This is typically accomplished through a simplifying hypothesis: the solution is of minimum norm [11], or constrained by a prescribed tradeoff between quality of fit and smoothness (e.g. LORETA[12]), or by imposing covariance constraints on the solution (as in Beamformer[13] techniques), etc. Alternatively, Bayesian methods [14–17] incorporate hypotheses through the introduction of explicit *priors*. These priors are used as a starting probability distribution for source activities, and this distribution is updated by EEG data in a manner consistant with Bayesian statistics. However, none

of these hypotheses are guaranteed to be correct. Thus, aside from convenience, there is little reason to suppose that they are more "correct" than the infinitely many other solutions that equally agree with data (including those produced by different regularization schemes). More problematically, these regularization techniques actually exclude reasonable source distributions from possible solutions when those source distributions do not conform to the regularizing hypothesis. Because of this, for *any* given EEG-only regularization technique, there are distributions of sources that can *never* be found as a solution, no matter what the experiment is, or what EEG data is produced [18, 19]. If the true source distribution is one of those unobservable distributions, the solution may bear little resemblance to the true sources.

III. FMRI REGULARIZED INVERSE

One way to improve upon regularization schemes is to use an independent source of data (in our case, fMRI) to chose between the infinitely many solutions allowed by the EEG data. Our regularization strategy is to start with a minimum norm inverse of EEG data, and then alter it in ways that improve agreement with fMRI data, without altering the quality of the fit to EEG data. We accomplish this by adding vectors in the null space of the lead field matrix. We still arrive at a unique solution, but we do so with more principled assumptions.

To construct our fMRI regularized inverse, we begin by considering solutions of the form

$$\vec{S} = L^- \vec{E} + N_L \vec{\alpha}, \qquad (3)$$

where L^- is the Moore-Penrose generalized inverse [20], and N_L is a matrix whose columns are vectors in the null space of L. With this form, the unknown vector $\vec{\alpha}$ controls the projection of the solution into the null space. Thus, our solution is the minimum-norm solution, plus some unknown vector in the null space of L.

Here we briefly justify this choice for the solution form. We choose our sources to minimize disagreement between the actual EEG data and our modeled EEG data. We obtain the modeled EEG data E_{solution} with the lead field matrix and the forward model (Equation 1). Thus, we define the solution error as

$$\sigma_{\vec{E}} = \| \vec{E}_{\text{solution}} - \vec{E} \| = \| L\vec{S} - \vec{E} \| . \qquad (4)$$

Here $\| M \|$ denotes the sum of the squares of all the elements of matrix M. Substituting the form of our solution from Equation 1, we obtain

$$\sigma_{\vec{E}} = \| L(L^- \vec{E} + N_L \vec{\alpha}) - \vec{E} \| . \qquad (5)$$

Then since N_L is a matrix of columns in the null space of L, $LN_L = 0$, and therefore

$$\sigma_{\vec{E}} = \| LL^- \vec{E} - \vec{E} \| . \qquad (6)$$

Thus the error is independent of the value of $\vec{\alpha}$ —adding the null vector to the solution does not affect the quality of the EEG fit. Regularization is then a process whereby we choose a specific $\vec{\alpha}$ (for each time step) and arrive at a unique solution. Our goal is to choose an $\vec{\alpha}$ that causes our solution to be maximally consistent with accompanying fMRI data. Since the fMRI data will vary with different experiments, the constraints placed on the solution will vary as well. Thus, this technique dramatically decreases the number of solutions that *in principle* cannot be found, and thereby avoids eliminating from consideration potentially valid solutions.

Any attempt at combining information from EEG and fMRI invariably confronts the problem that EEG and fMRI data are measuring fundamentally different quantities. fMRI measures blood oxygen level, which is affected by blood volume, flow rate, and of course the oxygen content of the blood. The various factors contributing to the BOLD signal are directly or indirectly related to the metabolism of brain cells, and thus neural activity. On the other hand, the currents responsible for EEG are ionic currents produced when neuronal activity causes alterations in the flow of ions into and out of neurons. Since this activity varies spatially, currents are produced in the brain fluid, roughly synchronous with and proportional to activity. Additionally, current is fundamentally a vector quantity, although it is possible to describe the EEG sources as scalars by fixing the direction of the currents (e.g. by anatomical constraints), or by working with current source density, or local field potential (voltage as measured by an electrode inserted into the brain).

Several studies by different researchers have shown that BOLD signal and neural activity are correlated —in fact, if the frequency of the neural activity is fixed, the BOLD signal is roughly linearly proportional to the neural activity [21–24]. BOLD signal has also been shown to be proportional to local field potential [23, 24]. Since the current sources should be proportional to neural activity (and local field potential) it is reasonable to expect the sources of the EEG signal to be proportional to the BOLD signal. This is the basis of our regularization scheme: modes from the null space of the lead field are added to ensure that the solution is proportional to the BOLD signal $\vec{B}$.

To regularize, we minimize the disagreement between the solution and linearly-scaled fMRI at each time. We introduce the unknown constants λ (which describes the proportionality between fMRI signal and neural activity) and κ (which is an offset).

$$\sigma_{\vec{B}} = \min_{\{\lambda, \kappa, \vec{\alpha}\}} \| |L^- \vec{E} + N_L \vec{\alpha}| - \lambda \vec{B} - \kappa \|, \qquad (7)$$

with $|M|$ used to denote the matrix whose elements are the absolute values of the elements of M. λ and κ are unknown a priori because they vary from machine to machine (and even between trials on a single machine). Thus

λ, κ, and $\vec{\alpha}$ are determined by our minimization procedure. Since we assume the proportionality constants are not changing over the course of a brief period of time, we obtain a single set of values for λ and κ that characterize the relationship between BOLD and neural activity for the entire experiment.

An additional issue is that the BOLD signal is temporally very different from the EEG signal. We will discuss this in detail in Section V, but it can be thought of (as a simple approximation) as following neural activity, but delayed by roughly four seconds. Also, fMRI data is sampled at a much lower rate than EEG data (a typical fMRI machine may record a new image every two seconds, whereas EEG is often sampled every two milliseconds), and therefore most EEG data will not have accompanying BOLD signal at *precisely* the same time.

Typically analysis involves creating a poststimulus time histogram (PSTH) collecting all images taken after a trial of a certain type: for example, a subject may look at pictures of spiders or puppies, in which case all "spider" images would be grouped together and categorized by the relative time since the most recent spider picture was displayed (EEG data is averaged through a similar process). If there is a variable inter-trial interval that is not synchronous with the image acquisition, then the BOLD signal is effectively sampled several times per second. Because the actual signal is changing much more slowly than this (over the course of a few seconds), it can be interpolated or shifted in time with a Discrete Fourier Transform or a Fourier Series without missing features. Thus we shift the transformed signal to account for the temporal delay. We group vectors together so that for each time t, $\vec{\alpha}(t)$ for that time is a column in a matrix $\underline{\alpha}$ (and similarly $\vec{B}$ is replaced by $\underline{B}$ and $\vec{E}$ is replaced by $\underline{E}$). Then Equation 7 becomes a single equation to simultaneously solve for λ, κ, and the matrix $\underline{\alpha}$ (i.e. we solve for every vector $\vec{\alpha}(t)$ simultaneously):

$$\sigma_B = \min_{\{\lambda, \kappa, \underline{\alpha}\}} \| \, |L^- \underline{E} + N_L \underline{\alpha}| - \lambda \underline{B} - \kappa \, \| . \qquad (8)$$

Equation 8 is not a linear equation. The EEG source currents are inherently signed —large positive and large negative currents both indicate large neuronal activity, but they contribute oppositely to measured EEG signal. Thus it is the absolute value of the EEG currents that is proportional to the BOLD signal. However, this nonlinearity is mild enough that Equation 8 can be solved iteratively through the following program:

$\underline{S_0} = L^- \underline{E}$
Repeat (increasing n)
   $\min_{\{\lambda_n, \kappa_n\}} \| \, |\underline{S_{n-1}}| - \lambda_n \underline{B} - \kappa_n \, \|$
   define $\underline{F_n}$ s.t. $|\underline{F_n}| = |\lambda_n \underline{B} + \kappa_n|$ and
      $\text{sign}(\underline{F_n}) = \text{sign}(\underline{S_{n-1}})$
   $\underline{\alpha_n} = \arg \min_{\underline{\alpha}} \| \, \underline{S_0} + \overline{N_L \underline{\alpha}} - \underline{F_n} \, \|$
   $\underline{S_n} = \underline{S_0} + N_L \underline{\alpha_n}$
Stop when $\| \, \underline{S_n} - \underline{S_{n-1}} \, \| <$ Tolerance

Once the difference between consecutive iterations of this procedure falls below the prescribed tolerance, $\underline{S} \equiv \underline{S_n}$ is a solution to the nonlinear minimization problem. Each of the minimization steps is linear, and reasonably fast. The first, minimizing over λ and κ, is merely a best-fit line (albeit one with many data points). The second is solved to obtain $\underline{\alpha_n}$ as:

$$\underline{\alpha_n} = N_L^- (\underline{F_n} - \underline{S_0}), \qquad (9)$$

and N_L^- need only be calculated once in advance. We continue the iteration until the change in $\underline{S_n}$ from one iteration to the next is below our specified tolerance. The result of this algorithm, $\underline{S_n}$ is the fMRI regularized inverse solution $\underline{S}$ that we sought. We present the results from tests of the fMRI regularized inverse in Section V.

IV. MODEL-REDUCED JOINT INVERSE

To better incorporate fMRI data into source localization, we next incorporate a model that provides detail beyond merely noting that the BOLD signal is proportional to neural activity. As mentioned previously, the BOLD signal response is delayed (by roughly four seconds for a brief impulse of neural activity). Our fMRI regularized inverse takes advantage of the correlation between BOLD signal and neural activity, but using the correlation alone is not ideal because the BOLD signal does not simply mirror lagged neural activity. The signal is also temporally broadened, and therefore simply shifting BOLD to eliminate the time lag does not result in a signal that matches neural activity. By incorporating a detailed model of BOLD into an inversion algorithm, we obtain closer agreement between our model and data, and thus a more accurate inverse.

Detailed models of the BOLD response exist —we use a variant of the Balloon/Windkessel model [25]. The model we use (developed by Robinson, *et al.* [26]) is a set of coupled nonlinear differential equations that start with neural activity as an input, and model physical changes in the blood vessels, changes in the blood oxygen content, and the measured response of these changes in an fMRI machine. The result is shown in Figure 2. The differential equations are nonlinear, but over a range of input activities the amplitude of the BOLD response varies approximately linearly with the amplitude of the neural activity. This *must* be the case, because we know that BOLD is proportional to neural activity.

Our fMRI regularized inverse cannot make use of the realistic model of BOLD response because it does not have an estimate of how the neural activity varies with time until it produces its solution. In order to incorporate a modeled BOLD response into our inversion procedure, we take an alternate approach that starts by identifying candidate neural activity signals, and then fits them
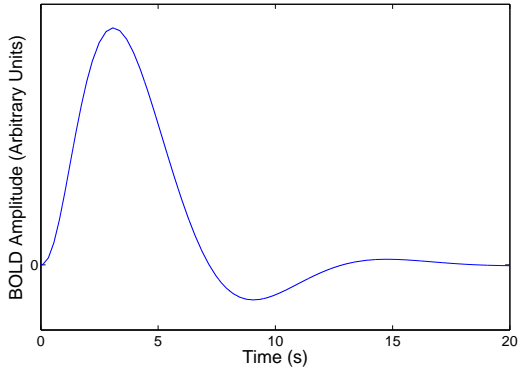
FIG. 2: **BOLD Response.** Here we show the modeled BOLD response after an impulse of activity at $t = 0$.

to data. Principle Component Analysis (PCA) and Independent Component Analysis (ICA) are ideally suited to this job. PCA extracts orthogonal signals that explain the majority of variance in the data (while reducing noise) and ICA forms linear combinations of those orthogonal vectors to produce a basis of signals that are maximally independent of each other. Figure 3 shows the result of this process.
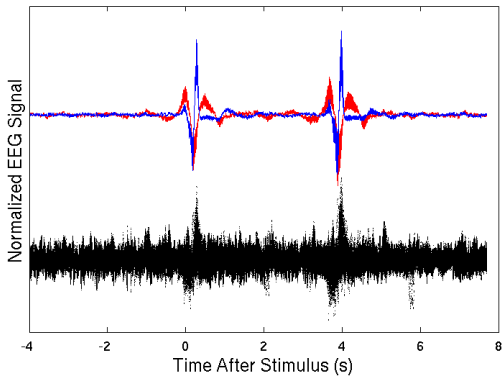

FIG. 3: **ICA Basis Signals.** Here we show an example of how ICA extracts the critical features of EEG measurements with a real data set. Human volunteers watched a series of movies with different pairs of everyday objects. Either a human would be present in the scene and look at one object, or one object would be highlighted. Here we show EEG data averaged over the condition where during two sequential movies the same object is looked at by a human, but from one movie to the next the object has switched sides (e.g. from left to right). The first two EEG signals determined by ICA are plotted in blue and red, respectively. The 128 original EEG signals are plotted in black.

The procedure to compute our model-reduced joint inverse begins by using PCA followed by ICA on EEG data to produce candidate EEG signals $\underline{\varepsilon}$ (organized so that

signals make up the columns of the matrix). Next we use $\underline{\varepsilon}$ and the BOLD model to produce candidate BOLD signals $\underline{\beta}$. Then we simultaneously fit linear combinations of these signals to EEG and fMRI data by solving the minimization problem:

$$\min_{\{\underline{\alpha}, \lambda, \kappa\}} \sigma = \| (\underline{E} - L\underline{\alpha}\,\underline{\varepsilon})\sigma_E \| + \| (\lambda\underline{B} + \kappa - \underline{\alpha}\,\underline{\beta})\sigma_B \|, \tag{10}$$

where σ_E and σ_B are diagonal matrices whose entries are the measurement uncertainty for EEG data and BOLD data respectively. Although superficially the minimization problem appears to be complicated, it is linear and can be solved efficiently. Once $\underline{\alpha}$ has been determined, the solution is

$$\underline{S} = \underline{\alpha}\,\underline{\varepsilon}. \tag{11}$$

We derive the solution to this minimization problem in Appendix C.

Aside from a more accurate treatment of BOLD, this technique has several additional advantages.

- Linearity guarantees a fast, exact solution.

- Since the number of temporal signals is far less than the number of sampled times, the size of the inverse problem is dramatically reduced.

- BOLD does not have to be interpolated or shifted in time.

- This procedure allows the EEG and fMRI data to be inverted simultaneously, and each fMRI data point and each EEG data point are weighted equally in the inversion. Thus we treat fMRI and EEG on equal footing, instead of using fMRI as a constraint on an EEG solution.

The basic technique is also very flexible and can be altered to include additional temporal signals that the experimenter may expect to be present. For example, the experimenter could insert a cardiac signal as a column of $\underline{\alpha}$ to provide an alternative to preprocessing data. Additionally, it is possible to group solution points together that the experimenter may expect to have correlated activity (either due to neurophysiological, functional, or anatomical priors, or from PCA/ICA used on fMRI data).

V. TESTING THE ALGORITHMS

We tested each source localization technique by generating a data set with known activity patterns and then comparing the actual activities to the solutions produced by the inversion technique. Inversion of multiple sources was a critical test, as many existing EEG inversion techniques (such as those based on ad hoc regularization schemes) have particular difficulty with multiple

sources. We tested the techniques against sources that were distributed over moderate-sized regions of the brain (roughly ten percent of its width). We chose distributed sources (instead of sources with a single active point) because many regularization schemes have difficulty with distributed sources. Of course, the presence of multiple, distributed sources is also to be expected in any experiment involving human subjects.

To produce the synthetic data, we began by calculating a lead field and an associated set of solution points from an anatomical MRI image, using the SMAC [27] toolkit. We then randomly selected one or more of those solution points to have significant activity. For the majority of our trials, activity consisted of two randomly chosen locations, with a cluster of nearby sites all having the same activity pattern. To produce a realistic time course for such a signal [26] we used the form

$$z(t) = ae^{-\beta t} sin(\beta \delta t). \tag{12}$$

All active sites (those in any active cluster) had an activity $z(t)$ (with parameters a, β, and δ randomly generated), and all other points had an activity $z(t)$ equal to zero. We grouped the activities for all solution points together into the vector $\vec{z}(t)$. Synthetic EEG data was then generated by multiplying the sources $\vec{z}(t)$ by the lead field matrix L. The BOLD signal was generated via the differential equations, acting directly on $z(t)$ for each solution point. Gaussian noise (uncorrelated in space and time) was added to all the values for these synthetically produced EEG and BOLD signals. The noise level was scaled so that the standard deviation was a pre-chosen fraction of the maximal signal value. For example, if the noise fraction was 0.1, then the Gaussian noise added to the EEG signal would have a standard deviation of 0.1 times the maximum EEG signal amplitude (maximum across all channels and the whole trial), and the Gaussian noise added to the fMRI signal would have a standard deviation of 0.1 times the maximal fMRI signal. After noise was added, the signals were then preprocessed (fMRI data was whitened, and the EEG signal was passed through a notch filter to remove signal components above 50 Hz). After performing the inversion procedure, the solution sources were compared to the true sources, z. The most critical comparisons were true source location versus solution source location, and true signal time courses versus solution time courses.

For our model reduction procedure, PCA and ICA were performed by the FastICA algorithm [28], using a "skew" non-linearity. We found that FastICA's "symmetric approach" produces better quality solutions, but was more likely to fail to converge than the "deflation approach"; thus our procedure made up to two attempts with the symmetric approach, and if they both failed, used the deflation option as a fall-back. Over the course of thousands of test trials with synthetic data sets, this strategy was sufficient to produce a solution in each case.

Since it was impossible to manually oversee the generation of so many solutions, invariably some low quality ICA results were obtained. Refinements to the ICA algorithm (or finding a more robust alternative) could further improve the quality of the model-reduced joint inverse.

As expected, a simple minimum norm inverse produces results of limited usefulness. While some high activity areas overlap the true sources, there are high activity areas found far away as well (Figure 4), and also there are true sources that are not not found. The time courses of correctly identified points are of decreased amplitude in low noise data (Figure 5) and are lost altogether in high noise data (Figure 6). Regularizing the inversion with fMRI produces only slightly better results. The fMRI regularized inverse solution contains false positives and negatives, and still finds a greatly decreased signal amplitude. By contrast the model-reduced joint inverse method correctly identifies the true locations of the sources, and the time courses it finds are closer to the correct amplitude. When the synthetic data has very little noise added to it, the model-reduced joint inverse has slightly less noise in its solution than do the other two techniques —this is because the model-reduced joint inverse produces a solution that is a linear combination of the (presumably smooth) signals found by ICA. However, the model-reduced joint inverse also responds more robustly to increased noise, producing useful solutions even after the minimum norm and fMRI regularized inverses produce solutions dominated by noise. Figure 7 shows plots of the *receiver operating characteristic* (ROC) as well as signal error at increasing values of noise.

VI. CONCLUSIONS

The fMRI regularized inverse solution provides results that are only a slight improvement over a simple minimum norm solution. Any regularized solution involves selection of a single point near or within the (very large) null space of the lead field matrix. The main advantage of our fMRI regularized inverse is that it arrives at a unique solution in a data-driven manner, and therefore represents an improvement largely in conceptual framework rather than solution quality. We suspect that any solution that takes the form of an EEG-only solution plus a null vector will tend to be biased in the same ways that the minimum norm solution is. Specifically, which solution points are active will be due to influences from EEG data as well as distortions introduced because the resolution matrix $R = L^- L$ is far from the identity. Put another way, the EEG-only part of the solution is distorted because the lead field matrix can only be inverted approximately, and in a manner that biases the result towards certain solution locations. Thus starting with the EEG-only solution and then adjusting it through some regularization process necessarily entails a poor starting
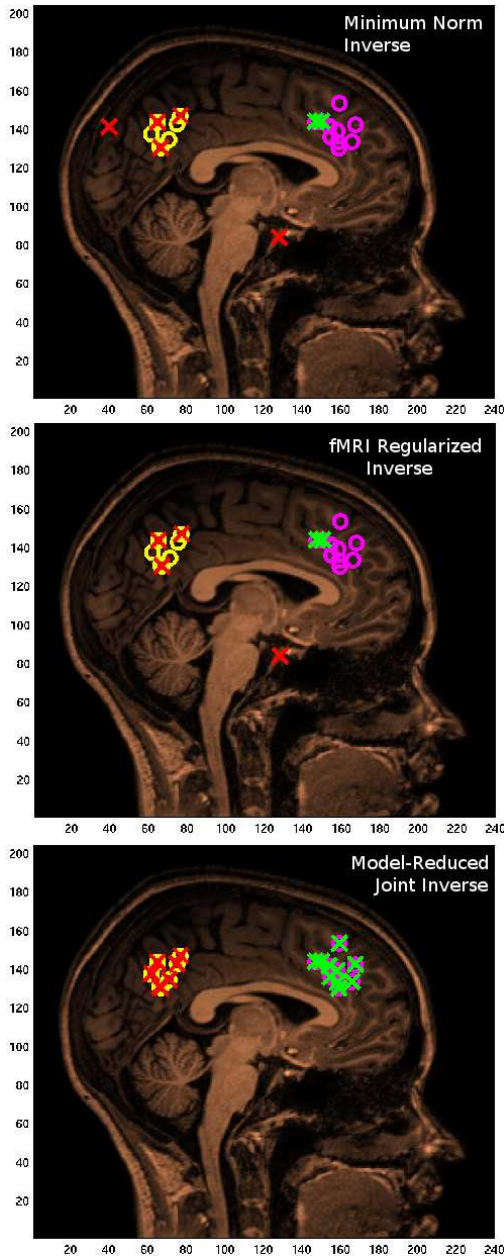
FIG. 4: **Typical Location Accuracy of Source Inversion.** A view from the side of solutions found by different inverse algorithms applied to the same synthetic data set: the minimum norm inverse (top), fMRI regularized inverse (middle), and model-reduced joint inverse (bottom). The different true sources are circles colored magenta and yellow, with X's colored green and red marking the locations found by the inversion procedure (green corresponds to magenta, and red to yellow). As was typical in our trials with low noise data, the model-reduced joint inverse correctly identifies the location of each solution point, while the other two techniques have false positives and false negatives. The fMRI regularized inverse performed marginally better in this example, which was also typical in our low noise tests.
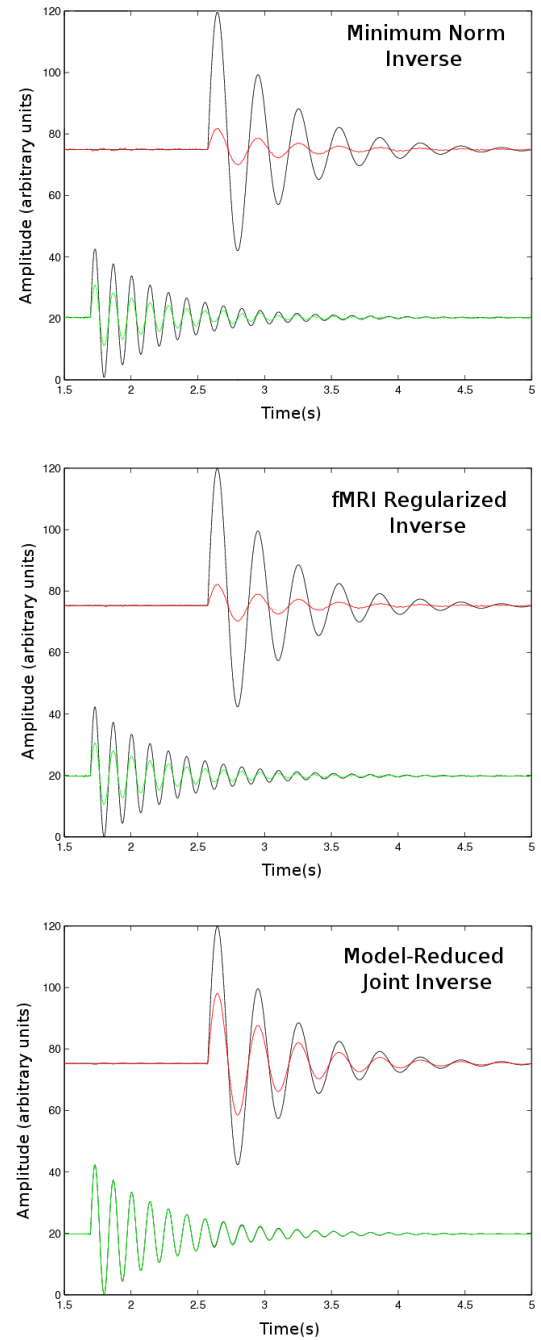
FIG. 5: **Typical Time Course Accuracy of Source Inversion.** In this test with low noise data, the signals recovered from minimum norm inverse (top) and fMRI regularized inverse (middle) are significantly reduced in amplitude compared to signals from the model-reduced joint inverse (bottom). The true synthetic signals are in black with the inverted signals in red and green (different colors indicate that the algorithm found two distinct sources). In the bottom signal for the model-reduced joint inverse, the inverted signal overlays the true signal.
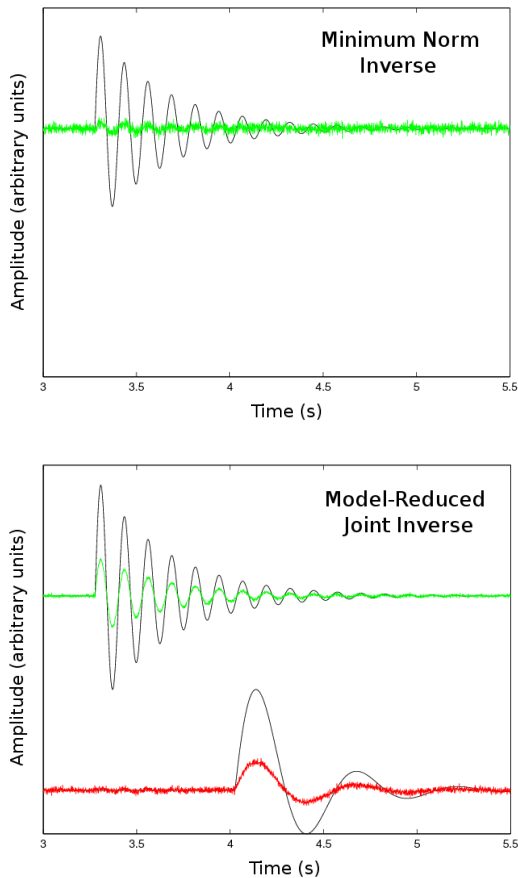
FIG. 6: **Time Courses in a High Noise Test.** Here are the time courses found by the minimum norm inverse (top) and the model-reduced joint inverse (bottom) in a test with increased noise. The minimum norm solution finds only one source, which is dominated by noise. The model-reduced joint inverse has produced a good solution (albeit one with a decreased amplitude). The fMRI regularized inverse (not shown here) also produces a solution that is dominated by noise.
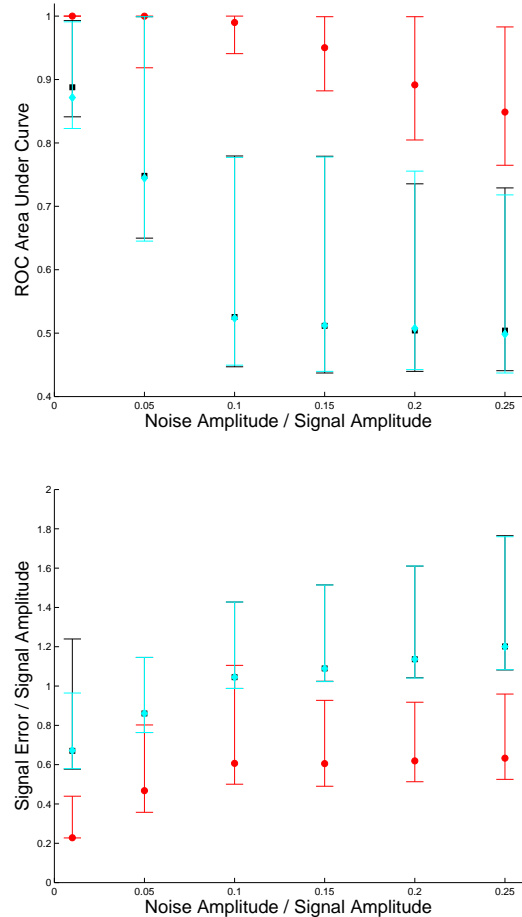

FIG. 7: **Accuracy at Progressively Higher Noise Levels.** Cumulative results for 1000 trials at each of six noise levels. Gaussian noise (uncorrelated in space and time) was added to the synthetic EEG and fMRI data, with the noise amplitude scaling as a fraction of the maximal EEG and fMRI signal strength respectively. The top graph shows area under the ROC curve, and the bottom graph shows signal error normalized to signal magnitude (taken in the frequency domain). The results are color-coded by inversion algorithm, with minimum norm inverse as cyan, fMRI regularized inverse as black, and model-reduced joint inverse as red. The bars indicate standard deviation. Note that an ROC area of 0.5 indicates that a test has provided no information, while an ROC area of 1.0 indicates perfect discrimination.

point. The result of regularized EEG inverse solutions is therefore likely to include false positives and false negatives.

By contrast, the model-reduced joint inverse algorithm outperforms minimum norm consistently, and over every range of test parameters we tried. The model-reduced inverse has several advantages, including a general reduction in noise, inherent uniqueness of the solution, and more symmetric treatment of EEG and fMRI data via equally-weighted simultaneous fitting. It is perhaps unsurprising that the more detailed treatment of BOLD results in better determination of source location, but using ICA to extract source signals results in superior characterization of signal time courses as well.

One potential source for further improvement in our algorithm is is in the ICA process itself. ICA is an iter-

ative technique, that occasionally fails to converge, and occasionally converges to a sub-optimal solution. Our procedure (skew nonlinearity, two tries with symmetric approach followed by deflation) was able to overcome this difficulty and produce unsupervised solutions to thousands of synthetic trials. However, further improvements are likely possible through additional adjustment of the ICA parameters, human supervision of the process (e.g. experimenter judgment of whether the ICA procedure

has been successful), or replacement of ICA with another algorithm that may be superior for this particular purpose.

The principle remaining difficulty with combined EEG-fMRI techniques is accurate registration of the EEG solution points to the underlying structural MRI image —a problem we avoided by dealing with purely synthetic data. However, this subject is an active area of research[27, 29] with solutions already in existence. We expect that the advantages of the model-reduced joint inverse as applied to synthetic data will carry over in application to human data.

Appendix: The Minimization Problem for the Model-Reduced Joint Inverse

In this Appendix we show the steps necessary to solve the minimization problem presented in Section V. Although the problem and derivation are novel, the numerical algorithms required to obtain the solution are all widely available. The minimization problem is:

$$\min_{\{\underline{\alpha}, \lambda, \kappa\}} \sigma = \| (\underline{E} - L\underline{\alpha}\,\underline{\varepsilon})\sigma_E \| + \| (\lambda\underline{B} + \kappa - \underline{\alpha}\,\underline{\beta})\sigma_B \|, \tag{13}$$

Henceforth we drop the underline notation —all the vectors have been converted to matrices. To begin, we first compute the variation of σ with respect to α, and set it to zero:

$$L^T E \sigma_E^2 \varepsilon - L^T L\alpha\varepsilon\sigma_E^2\varepsilon^T + (\lambda B + \kappa)\sigma_B^2\beta^T - \alpha\beta\sigma_B^2\beta^T = 0. \tag{14}$$

Right-multiplying this equation by $(\varepsilon\sigma_E^2\varepsilon^T)^-$ puts this equation into the form of a Sylvester equation,

$$L^T L\alpha + \alpha\beta\sigma_B^2\beta^T(\varepsilon\sigma_E^2\varepsilon^T)^- + (\lambda B + \kappa)\sigma_B^2\beta^T(\varepsilon\sigma_E^2\varepsilon^T)^- \tag{15}$$

$$-L^T E\sigma_E^2\varepsilon(\varepsilon\sigma_E^2\varepsilon^T)^- = 0$$

Sylvester equations (equations of the form $A\alpha + \alpha B + C = 0$) are solvable numerically, and have a nice linear property. For matrices $C1$, $C2$ and scalars u, v:

$$\alpha(A, B, uC1 + vC2) = u\alpha(A, B, C1) + v\alpha(A, B, C2). \tag{16}$$

Thus we can write

$$\alpha = \alpha_0 + \lambda\alpha_\lambda + \kappa\alpha_\kappa, \tag{17}$$

where α_0, α_λ and α_κ are the solutions to the Sylvester equation considering only the constant parts, the parts with coefficient λ, and the parts with coefficient κ respectively.

Now we substitute this form into Eq. 13 to obtain

$$\min_{\{\lambda, \kappa\}} \sigma = \| (E - L(\alpha_0 + \lambda\alpha_\lambda + \kappa\alpha_\kappa)\varepsilon)\sigma_E \| \tag{18}$$

$$+ \| (\lambda B + \kappa - (\alpha_0 + \lambda\alpha_\lambda + \kappa\alpha_\kappa)\beta)\sigma_B \|,$$

In this seemingly complicated expression, we are only interested in solving for λ and κ. Since λ and κ are linear inside the norms, this is just a best-fit plane. The data points are organized into two separate matrices, but that is irrelevant. We simply treat every entry in the matrices as a separate data point, and thereby reduce the problem to one of the form

$$\min_{a,b} \| a\vec{x} + b\vec{y} + \vec{z} \| . \tag{19}$$

After fitting for λ and κ, we can now substitute their values into Eq. 17 and thereby obtain the solution,

$$S = \alpha\varepsilon. \tag{20}$$

[1] Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W., Proceedings of the National Academy of Sciences of the United States of America **87**, 9868 (1990).

[2] H. B. K.-D. M. W. H. Jens Frahm, Journal of Magnetic Resonance Imaging **2**, 501 (1992).

[3] H. J. Heinze et al., Nature **372**, 543 (1994).

[4] G. S. Kevin Whittingstall and M. Schmidt, Human Brain Mapping **28**, 134 (2007).

[5] G. P. V.D. Calhoun, T. Adali and K. Kiehl, NeuroImage **30**, 544 (2006).

[6] H. N. K. H. Matthias Moosmann, Tom Eichele and V. D. Calhoun, International Journal of Psychophysiology (2007), (In Press, Corrected Proof).

[7] C. M. Michel et al., Clinical Neurophysiology **115**, 2195 (2004).

[8] C. Phillips, M. D. Rugg, and K. J. Friston, NeuroImage **16**, 678 (2002).

[9] R. G. de Peralta Menendez, S. G. Andino, S. Morand, C. Michel, and T. Landis, Human Brain Mapping **9**, 1 (2000).

[10] R. G. de Peralta Menendez and S. G. Andino, Inter. J. of Bioelectromagnetism **4** (2002).

[11] M. S. Hmlinen and R. J. Ilmoniemi, Medical and Biological Engineering and Computing **32**, 35 (1994).

[12] C. M. M. R. D. Pascual-Marqui and D. Lehmann.

[13] D. P. A. M. Kensuke Sekihara, Srikantan S. Nagarajan and Y. Miyashita, IEEE Transactions on Biomedical Engineering **48**, 760 (2001).

[14] K. Friston et al., NeuroImage **39**, 1104 (2008).

[15] S. J. Kiebel, J. Daunizeau, C. Phillips, and K. J. Friston, NeuroImage **39**, 728 (2008).

[16] R. H. C. P. J. M. Karl Friston, Human Brain Mapping **27**, 722 (2006).

[17] C. Phillips, J. Mattout, M. D. Rugg, P. Maquet, and K. J. Friston, NeuroImage **24**, 997 (2005).

[18] R. Grave de Peralta Menendez and S. L. Gonzalez Andino, IEEE Trans Biomed Eng **45**, 440 (1998).

[19] R. Grave de Peralta Menendez and S. L. Gonzalez Andino, Brain Topography **13**, 97 (2000), 10.1023/A:1026603017734.

[20] W. Menke.

[21] D. J. Heeger, A. C. Huk, W. S. Geisler, and D. G. Albrecht, Nat Neurosci **3**, 631 (2000), 10.1038/76572.

[22] D. J. Heeger and D. Ress, Nat Rev Neurosci **3**, 142 (2002), 10.1038/nrn730.

[23] A. Shmuel, M. Augath, A. Oeltermann, and N. K. Logothetis, Nat Neurosci **9**, 569 (2006), 10.1038/nn1675.

[24] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, Nature **412**, 150 (2001), 10.1038/35084005.

[25] R. Buxton, E. Wong, and L. Frank, Magn Reson Med **39**, 855 (1998).

[26] P. Robinsona et al., NeuroImage **31**, 585 (2006).

[27] L. Spinelli, S. G. Andino, G. Lantz, M. Seeck, and C. M. Michel, Brain Topography **13**, 115 (2000).

[28] A. Hyvärinen, Neural Computing Surveys **2**, 94 (1999).

[29] S. B. John J Ermer, John C Mosher and R. M. Leahy, Physics in Medicine and Biology **46**, 1265 (2001).