EEG-fMRI Fusion of Paradigm-Free Activity Using Kalman Filtering

Thomas Deneux
thomas.deneux@weizmann.ac.il
Odyssée Team, Ecole Normale Supérieure, Département d’Informatique, 75005 Paris, France; and Weizmann Institute of Science, Department of Neurobiology, 76100 Rehovot, Israel

Olivier Faugeras
olivier.faugeras@sophia.inria.fr
NeuroMathComp Team, Institut National de Recherche en Informatique et Automatique, France; and Ecole Normale Supérieure, Département d’Informatique, 75005 Paris, France

We address here the use of EEG and fMRI, and their combination, in order to estimate the full spatiotemporal patterns of activity on the cortical surface in the absence of any particular assumptions on this activity such as stimulation times. For handling such a high-dimension inverse problem, we propose the use of (1) a global forward model of how these measures are functions of the “neural activity” of a large number of sources distributed on the cortical surface, formalized as a dynamical system, and (2) adaptive filters, as a natural solution to solve this inverse problem iteratively along the temporal dimension. This estimation framework relies on realistic physiological models, uses EEG and fMRI in a symmetric manner, and takes into account both their temporal and spatial information.

We use the Kalman filter and smoother to perform such an estimation on realistic artificial data and demonstrate that the algorithm can handle the high dimensionality of these data and that it succeeds in solving this inverse problem, combining efficiently the information provided by the two modalities (this information being naturally predominantly temporal for EEG and spatial for fMRI). It performs particularly well in reconstructing a random temporally and spatially smooth activity spread over the cortex.

The Kalman filter and smoother show some limitations, however, which call for the development of more specific adaptive filters. First, they do not cope well with the strong nonlinearity in the model that is necessary for an adequate description of the relation between cortical electric activities and the metabolic demand responsible for fMRI signals. Second, they fail to estimate a sparse activity (i.e., presenting...
sharp peaks at specific locations and times). Finally their computational cost remains high. We use schematic examples to explain these limitations and propose further developments of our method to overcome them.

1 Introduction

Noninvasive human brain imaging techniques provide impressive measurements of the brain’s functional activity. Moreover, since different techniques measure different facets of this activity with different spatiotemporal characteristics, it is important to combine these techniques through simultaneous acquisitions as well as by elaborating new analysis methodologies.

Here we focus on the combination of functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), although our methods could apply to other modalities as well. The combination (or fusion) of fMRI and EEG data has always been stressed as an important issue because of their very different spatial and temporal resolutions. Indeed, the spatial resolution of fMRI has reached submillimetric scale for whole-brain acquisitions (Moon, Fukuda, Park, & Kim, 2007), whereas its temporal resolution is limited by the acquisition rate (\(\sim 1\) Hz), and, more intrinsically, the slow nature of the hemodynamic response to neural activity (several seconds). EEG signals have a temporal resolution in the millisecond range, whereas the spatial localization obtained by solving the inverse problem of the propagation of electric currents through tissue and bone has a precision on the scale of the centimeter only (Baillet et al., 2001).

Several strategies have been employed to build a framework for fMRI-EEG fusion, and several criteria can be used to distinguish between them (Deneux, 2006): Which kind of information do EEG and fMRI share in common: information on the localization of neural activity, on its time course, or both spatial and temporal information? Does it require simultaneous EEG-fMRI acquisition—or can acquisitions be separated? Is the method asymmetric or symmetric (does it use a first analysis of either EEG or fMRI data to later constrain the analysis of the second set of data, or does it use simultaneously the two data sets in order to estimate the underlying neural activity)? What is the geometry of the neural sources: Isolated dipoles, distributed activity, parcellation of the cortex, or something else? What is the underlying forward model—that is, which predicts EEG and fMRI measures from neural activity?

Initial attempts at EEG-fMRI fusion were mostly asymmetric methods where the spatial information provided by an fMRI study was used to constrain the localization of EEG/MEG (Dale & Sereno, 1993; Baillet & Garnero, 1997; Liu, Belliveau, & Dale, 1998; George et al., 1995; George, Schmidt, Rector, & Wood, 2001; Ahlfors & Simpson, 2004; Babiloni et al., 2004, 2005). Daunizeau et al. (2007) proposed Bayesian frameworks in conjunction with a parcellation of the cortical surface in order to
localize the neuronal activity common to EEG and fMRI in a symmetric manner, but where the electrical and hemodynamic temporal patterns were supposed to be independent from each other.

More recently, the possibility of acquiring EEG signals inside the fMRI scanner led to the opposite asymmetric methods, where temporal information of the EEG at the single-trial level is used to construct the regressors for fMRI analysis (Riera et al., 2005; Debener, Ullsperger, Siegel, & Engel, 2006). This has applications, for example, in the localization of slow oscillation activity (Goldman, Stern, Engel, & Cohen, 2002; Laufs et al., 2003) and, more important, of epileptic activity (Lemieux, Krakow, & Fish, 2001; Bénar et al., 2003; Gotman, Bénar, & Dubreucq, 2004; Bagshaw et al., 2005; Waite et al., 2005). Also, symmetric methods have been developed in order to extract temporal patterns of activity from EEG (in the frequency domain) and fMRI time courses (Martinez-Montes, Valdes-Sosa, Miwakeichi, Goldman, & Cohen, 2004).

However, the most complete strategies should use the fact that EEG and fMRI have both spatial and temporal information in common, though at different scales. Once a forward model has been proposed to describe EEG and fMRI signals as functions of neural activity, the problem of estimating this neural activity becomes that of solving a very large inverse problem, where the number of unknowns is the number of neural sources times the number of time instants. Trujillo-Barreto, Martinez-Montes, Melie-García, and Valdés-Sosa (2001) addressed this problem using a Bayesian framework and showed results on simulated as well as experimental data from a somatosensory MEG/fMRI experiment. Halchenko, Pearlmutter, Hanson, and Zaimi (2004) proposed a linear programming algorithm that allowed them to introduce nonlinearities in their forward model and obtained results on a reduced synthetic data set. Other authors tried to simplify the problem; for example, Kiebel and Friston (2004) reduced it to a pixel-by-pixel estimation. However, such a simplification ignores the intrinsic covariance between sources, which states, for example, that several sources of configurations can account for the same EEG measure.

Our fusion algorithm belongs to the latter group and is inspired by previous work that used dynamical systems to model the hemodynamic response in BOLD fMRI analysis (Friston, Mechelli, Turner, & Price, 2000; Trujillo-Barreto et al., 2001; Friston, 2002; Riera et al., 2004). The most challenging problem is the high computational cost of the source activity estimation. Our method is also intended to take into account the development of new paradigms based on ongoing activity and detection of single events (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004) rather than on activity evoked by the repetition of the same stimulation or task. For such paradigms, it becomes highly valuable to be able to estimate activity for a long period of time rather than on short evoked signals.

In order to simplify the problem, one would like to divide it in smaller pieces. Unfortunately, the activities of different sources cannot be estimated
separately since the EEG forward model introduces correlations between
them, and activities of the cortex at different times cannot be estimated
separately since the slow hemodynamic response introduces correlations
between them. However, it is interesting to notice that two time instants
separated by more than twice the hemodynamic response can be consid-
ered independent. As a consequence, filtering techniques—which estimate
the activity at successive time instants while keeping in memory the in-
formation provided at previous time instants—appear well suited for this
kind of estimation. Moreover, if the forward model of the measurements is
linear with respect to neural activity, then once the parameters of the filter
have been calculated, the processing of even long signals is fast.

Hence, we present here an estimation method for neural activity from
EEG and fMRI measurements based on the Kalman filter and the Kalman
smoother. We show results on realistic-size synthetic data. We later discuss
our method, using schematic examples, to explain our understanding of
the Kalman filter and its characteristics and give ideas on how its specific
limits could be overcome by the use of other adaptive filters.

Our method has these key features:

- It is based on simple physiological models.
- It allows combining both spatial and temporal data at very different
  space and timescales.
- It is able to handle high-dimensional data.
- It allows EEG-alone and fMRI-alone estimations as well (such estima-
tions are constrained by the temporal and spatial correlations entailed
by the physiological models). When both modalities are used together,
our method is effective in the sense that it succeeds in accumulating
the information provided by both.

2 Methods

2.1 Dynamic System Formulation of the Forward Model. Let \( u \) denote
the neural activity of sources distributed on the cortical surface, \( y_{\text{eeg}} \) the
signals measured by the EEG electrodes, and \( y_{\text{fMRI}} \) the BOLD fMRI signals
at the location of the sources (we suppose that the BOLD signals in the 3D
MRI volume have been interpolated properly in each source location). A
quite general model formulation would be:

\[
\begin{align*}
  y_{\text{eeg}} &= \Psi_{\text{eeg}}(u) + \eta_{\text{eeg}} \\
  y_{\text{fMRI}} &= \Psi_{\text{fMRI}}(u) + \eta_{\text{fMRI}}
\end{align*}
\]

or

\[
y = \Psi(u) + \eta,
\]

where \( \Psi_{\text{eeg}} \) and \( \Psi_{\text{fMRI}} \) describe the forward model and \( \eta_{\text{eeg}} \) and \( \eta_{\text{fMRI}} \) are
measurement noises.
The goal is to estimate $u$ knowing $y_{eeg}$ and $y_{fmri}$ and the forward model described by $\Psi_{eeg}$ and $\Psi_{fmri}$. Some parameters of the model and the characteristics of the noise may also be unknown and need to be estimated. However, in our case, the dimensionality of $u$ is so high that it is necessary to have a more specific formulation to make the estimation feasible. Such a formulation is provided by dynamic systems theory, which is well adapted in this case and leads to filtering estimation techniques. Such a dynamic system includes an evolution equation, which describes how the cortical state evolves in time, and two measurement equations, which describe for both modalities how the measures at specific times are related to this cortical state. By “cortical state,” noted $x$, we mean the neural activity $u$ together with other variables of interest whose evolution in time depends on the neural activity and that are used to predict the measures (typically hemodynamic variables like blood flow and blood volume):

$$
\begin{align*}
    x(0) &= x^0 + \xi^0 \\
    \dot{x}(t) &= F(x(t)) + \xi(t) \\
    y_{eeg}(t_k) &= G_{eeg}(x(t_k)) + \eta_{eeg}(t_k) \\
    y_{fmri}(t_l) &= G_{fmri}(x(t_l)) + \eta_{fmri}(t_l)
\end{align*}
$$

2.2 Physiological Models. Our forward model is quite simple, but the methods described in this letter still apply with different choices of model. However, it is important for algorithm efficiency that the system be linear or that it can be linearized without major harm (this issue is discussed later). Figure 1 depicts the main features of our model. Here we give its details; the issues pertaining to the critical choices we made are addressed in section 4 as well. Table 1 summarizes the parameters of the model and displays the values used in our simulation.

The first question to ask is about the nature of the so-called neural activity $u$ underlying the EEG and fMRI measurements. Since we would like to estimate the activity of a few thousand sources distributed on the cortex, each source covering an area of a few tenths of squared millimeters, the activity of each source should be an average taken over a large number of neurons. It should also be an average of synaptic, spiking, and other activities. However, we will assume this activity to be the average of dendritic current intensities over the entire area covered by the source. Indeed, EEG signals originate from parallel dendritic currents in large assemblies of
neurons with a synchronous activity, and these dendritic currents are the direct consequence of synaptic activity. Moreover, some authors (Creutzfeld, 1975; Nair, 2005) claim that most of the energy consumption in the brain that leads to a hemodynamic response is caused by synaptic activity. Hence, the literature does not contradict the hypothesis that the neural origin of the EEG and BOLD signals is roughly the same. However, we have to keep in mind that some activity can be seen by only one of the two modalities (e.g., depending on whether the population activity is synchronous).

The relation between this synaptic current activity and EEG measurements is described by the EEG forward model matrix $B$, constructed according to the Maxwell equations for the propagation of current through the different tissues and through the skull (Hamalainen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). This leads to the linear forward model:

$$y_{eeg}(t) = Bu(t).$$

We consider that there is one dipole at each source location and a sufficient number of sources to have a good geometric representation of the cortex.

Figure 1: Model of neural activity, hemodynamic hidden states, and EEG and fMRI measurements. The neural activity is supposed to obey an autoregressive evolution that can involve correlation between the different sources. An EEG measurement by a given sensor at a given time depends on all the sources' activities at the same time. The evolution of the hemodynamic state variables according to the balloon model is independent at each spatial location. Finally, the fMRI measurement at a given location and time depends on only the venous blood volume and deoxygenation at the same location and time (note also that the sampling frequency of fMRI is much lower than for the EEG).
Table 1: Typical Values of the Physiological and Noise Parameters Used in the Forward and Inverse Problem.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neural innovation process</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_s$ (200 Hz)</td>
<td></td>
<td>Sampling frequency of cortical activity and EEG</td>
</tr>
<tr>
<td>$\sigma_u^2$ (2 s$^{-1}$)</td>
<td></td>
<td>Variance scaling parameter for the innovation noise (the resulting average variance of the innovative noise of individual dipoles in the 35,000 vertices mesh is 0.084 s$^{-1}$, and the resulting average variance of the sources when 2000 patches are used is 0.04 s$^{-1}$. Note that the units in s$^{-1}$ are due to the fact that one needs to use a variance equal to $\sigma_u^2/f_s$ in the discretized evolution equation, 2.10)</td>
</tr>
<tr>
<td>$\phi$ (.01)</td>
<td></td>
<td>Spatial smoothness parameter for the innovation noise (the resulting spatial correlations between sources are shown in Figure S1 in the supplemental materials)</td>
</tr>
<tr>
<td>$\lambda$ (5 s$^{-1}$)</td>
<td></td>
<td>Decay time constant</td>
</tr>
<tr>
<td><strong>EEG measure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B$ (= $\mathbb{B}$)</td>
<td></td>
<td>EEG forward problem matrix for the sources (the average square sum of its rows is 14.7, such that an average variance of 1 in the sources results in an average variance of 14.7 at the EEG sensors)</td>
</tr>
<tr>
<td>$\sigma_{eeg}^2$ (0.001 s$^{-1}$)</td>
<td></td>
<td>Variance of EEG measure noise at individual sensors (as for $\sigma_u^2$, the actual variance is defined as $\sigma_{eeg}^2/f_s$, such that the power of this measure noise in a given temporal frequency domain does not depend on the sampling frequency)</td>
</tr>
<tr>
<td><strong>Hemodynamic evolution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\varepsilon$ (1)</td>
<td></td>
<td>Electric-metabolic scaling factor</td>
</tr>
<tr>
<td>$\kappa_s$ (.65 s$^{-1}$)</td>
<td></td>
<td>Flow decay rate</td>
</tr>
<tr>
<td>$\kappa_f$ (.4 s$^{-1}$)</td>
<td></td>
<td>Flow feedback rate</td>
</tr>
<tr>
<td>$\tau$ (1 s)</td>
<td></td>
<td>Transit time through the venous balloon</td>
</tr>
<tr>
<td>$\alpha$ (.4)</td>
<td></td>
<td>Steady-state flow-volume relation $v = f\alpha$</td>
</tr>
<tr>
<td>$E_0$ (.4)</td>
<td></td>
<td>Baseline O$_2$ extraction function</td>
</tr>
<tr>
<td><strong>fMRI measure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RT$ (1 s)</td>
<td></td>
<td>fMRI acquisition repetition time</td>
</tr>
<tr>
<td>$V_0$ (1)</td>
<td></td>
<td>Baseline blood volume</td>
</tr>
<tr>
<td>$a_1$ (1)</td>
<td></td>
<td>Weight for blood volume change</td>
</tr>
<tr>
<td>$a_2$ (-4.8)</td>
<td></td>
<td>Weight for deoxyhemoglobin change</td>
</tr>
<tr>
<td>$\sigma_{fmri}^2$ (.25)</td>
<td></td>
<td>Variance of fMRI measure noise at the location of a single dipole (once fMRI signals have been averaged over the multiple dipoles of the patch constituting individual sources, the average variance of fMRI measures becomes .015 when using 2000 patches)</td>
</tr>
</tbody>
</table>

(30,000–50,000 sources). We explain in appendix A.1 how to decrease the number of sources by constructing patches of dipoles with equal activity.

The relation between this activity and the energy consumption that causes the hemodynamic response is less understood. For simplicity, we choose a linear dependence (the metabolic demand is equal to $\varepsilon_\text{metabolism}$, where $\varepsilon$ is a fixed parameter). A more complex, nonlinear modeling will be proposed in section 4.2.2.

We then adopt the damped oscillator model proposed by Friston et al. (2000) to relate this activity to the venous blood flow evolution $f$ and the
balloon model introduced by Buxton and colleagues (Buxton, Wong, & Frank, 1998; Buxton, Uluda˘g, Dubowitz, & Liu, 2004) to relate the venous blood flow to the venous blood volume \( v \), the venous blood deoxygenation \( q \), and finally the BOLD signal \( y_{fmri} \). All of these quantities are considered at a single spatial location. This leads to the nonlinear forward model:

\[
\begin{align*}
\ddot{f} &= \varepsilon u - \kappa_s \dot{f} - \kappa_f (f - 1) \\
\dot{v} &= \frac{1}{\tau} (f - v^{1/\alpha}) \\
\dot{q} &= \frac{1}{\tau} \left( f \frac{1 - (1 - E_0)^{1/f}}{E_0} - v^{1/\alpha} \frac{q}{v} \right) \\
y_{fmri} &= V_0 (a_1 (v - 1) + a_2 (q - 1))
\end{align*}
\]  

(2.5)

where \( \varepsilon, \kappa_s, \kappa_f, \tau, \alpha, E_0, V_0, a_1, \) and \( a_2 \) are physiological parameters.

Putting all this together, we define \( x \), the state variable, as the concatenation of the source activity and the above hemodynamic variables \( x = (u, f, \dot{f}, v, q) \). All evolution and measurement functions in equation 2.3 are now defined except the evolution equation for the neural activity itself.

A first, naive, model would be a simple white noise innovation, or Brownian, process:

\[
\dot{u}(t) = \xi_u(t).
\]  

(2.6)

However, such a process eventually diverges, so a relaxation term is needed to pull back the activity to zero, leading to the Ornstein-Uhlenbeck process:

\[
\dot{u}(t) = -\lambda u(t) + \xi_u(t),
\]  

(2.7)

where \( \lambda \) is a positive parameter that controls the temporal autocorrelation of sources’ time courses \( \langle u(t)u(t + \Delta t) \rangle / \langle u(t)u(t) \rangle = \exp(-\lambda \Delta t) \).

Additionally, one might want to assume some spatial smoothness of the sources in the model. Indeed, the literature reports smooth waves of neural activity during ongoing activity (Ferezou, Bolea, & Petersen, 2006; Benucci, Frazor, & Carandini, 2007), and this additional physiologically motivated constraint on the estimation helps to deal with the high degree of underdetermination of the system (i.e., the fact that there are many more unknown variables than measures). This is achieved by an adequate choice of the correlation matrix \( Q_u \) of the innovative noise \( \xi_u(t) \). We set

\[
Q_u = \sigma_u^2 (\phi I_{n_u} + \text{diag}(\text{sum}(C)) + C)^{-1},
\]  

(2.8)

such that \( p(\xi_u(t) = x) = \exp(-\frac{1}{2} x^T Q_u^{-1} x) = \exp(-\frac{1}{2\sigma_u^2} (\phi \sum_i x_i^2 + \sum_{(i,j) \in C} (x_i - x_j)^2)) \), where \( I_{n_u} \) is the identity matrix of the number of sources, \( C \) the connectivity matrix of the mesh connecting the sources, and \( \sigma_u^2 \) and \( \phi \) two
parameters controlling, respectively, the total variance of the noise and its smoothness (the smaller $\phi$, the higher the spatial smoothness; we observed that for $0 < \phi << 1$, changing $\phi$ has little effect on the average variance of the sources).

The state at rest is $x^0 = (u^0, f^0, \dot{f}^0, v^0, q^0) = (0, 1, 0, 1, 1)$ (in the equations above, the variables are normalized by their values at rest), and we choose for the initial state distribution the one to which the system converges after evolving for a long time according to its evolution equation, equation 2.3b (indeed, this is the probability distribution for the cortical state after a long period of rest; see also section 2.4). (Here and in the rest of this letter, we refer to specific parts of a multiline equation using lowercase letters: $a$ refers to the first line of the equation, $b$ to the second line, and so on.)

The innovation and measurement noises are chosen to be zero mean gaussian. In our simulations, we did not consider innovation noises in the flow, volume, and deoxygenation evolutions. We did not introduce any spatial coloration in the EEG and fMRI measurement noises. As a result, all the noise in the system is characterized by four parameters: $\sigma_{u}, \phi, \sigma_{\text{EEG}}$, and $\sigma_{\text{fMRI}}$, the last two being the variance of the EEG measure at each sensor and the fMRI signal at each spatial location, respectively.

2.3 Simplifications. A first simplification consists of reducing the total number of sources by creating patches of dipoles with equal activity. The EEG forward problem matrix and the innovation and measure noise parameters must be changed accordingly. This is explained in appendix A.1.

The Kalman filter (Kalman, 1960; Welling, n.d.; Welch & Bishop, 2006) is a recursive algorithm that computes the a posteriori distribution of the hidden state $x(t)$ given all the measures before time $t$, $p(x(t) | y_{\text{EEG}}(t_k), t_k < t; y_{\text{fMRI}}(t_l), t_l < t)$. It assumes that all the parameters of the evolution and measure functions and the variances of the different noises are known. If system 2.3 is linear, then the computation is exact and the a posteriori distribution gaussian (i.e., fully described by its mean and variance).

If the system is not linear (which is our case), the procedure performed by the extended Kalman filter is to linearize the system at each iteration around the current a posteriori mean value. The computed a posteriori distribution is then a gaussian approximation of the true distribution, and its accuracy depends on how nonlinear the system is. In the case of the balloon model, our simulations on reduced data sets (not shown) showed that the departure from nonlinearity is small enough to allow the following approximation, which consists of linearizing the evolution and measurement equations once around the rest state $x^0$:

$$\dot{x}(t) = F(x(t)) \approx F(x_0) + \left( \frac{dF}{dx} \right)_{x_0} (x(t) - x_0) = \left( \frac{dF}{dx} \right)_{x_0} (x(t) - x_0)$$

$$y(t) = G(x(t)) \approx G(x_0) + \left( \frac{dG}{dx} \right)_{x_0} (x(t) - x_0)$$

(2.9)
This is very advantageous in term of computational cost, since it allows us to (pre)compute the parameters of the filter independent of the data and then to apply the filter efficiently on large data sets (see below).

The way the algorithm proceeds is to start with the a priori distribution of $x$ at time 0, then to propagate this information for increasing values of $t$ according to the evolution equation, and at each measurement time to update the current distribution in order to incorporate the information provided by this measurement. More specifically, in our case, it means that the distribution is updated more frequently by the EEG measurements than by the fMRI measurements. In such a case, each time the distribution is updated by an fMRI measure, its variance will drop significantly, whereas the same variance will increase continuously during each in-between period when there are only EEG measure updates. In order to reach the convergence of this a posteriori distribution more quickly and not need to store all the possible distribution values at different positions between two successive fMRI measure updates, we preferred to assume that there are as many fMRI measure points as EEG: we interpolate the fMRI data to the timescale of EEG data in order to perform the EEG and fMRI updates simultaneously. In order to prevent fMRI data from having more weight in the estimation than it should, it is necessary to multiply the measurement variance $\sigma^2_{fmri}$ used by the filter by the number of times the fMRI information has been duplicated. This interpolation does not add any significant additional computational cost or storage as it takes place on the spot (interpolated value at a given time $t$ is calculated only when it is needed by the iterative algorithm when it reaches time $t$).

After these simplifications have been applied, we rewrite the equations as a discrete linear system, with one EEG and fMRI measurement at each time instant $k = 1 \ldots, n$, represented as a single vector $y$:

$$
\begin{align*}
    x_1 &= x^0 + \xi^0, \quad \xi^0 \sim N(0, Q^0) \\
    x_{k+1} &= A x_k + \xi_k, \quad \xi_k \sim N(0, Q) \\
    y_k &= D x_k + \eta_k, \quad \eta_k \sim N(0, R)
\end{align*}
$$

(2.10)

In order to obtain equation 2.10 from 2.9, it is necessary to redefine the state and measures $x$ and $y$, as $(x - x^0)$ and $(y - G(x^0))$. In our case, the matrix $R$ has nonzero terms on the diagonal only; $Q^0$ and $Q$ are full matrices since the neuronal innovation noise is colored spatially.

### 2.4 Kalman Filter and Kalman Smoother

The Kalman filter is a standard iterative algorithm to estimate the a posteriori distribution $p(x_k \mid y_1, \ldots, y_k)$ of the state at time $k$ given the past measurements, described by their mean and covariance matrix. It is necessary, however, to apply a second filter, the Kalman smoother, in order to include information
from the future in the estimation \( p(x_k | y_1, \ldots, y_n) \), \( n \) being the total number of time instants. Here we give the details of these filters (Welling, n.d.).

The distribution of the state at time \( k \) given measurements until time \( l \) is described by its mean and covariance matrix:

\[
\hat{x}_k^l = E(x_k | y_1, \ldots, y_l) \\
P_k^l = V(x_k | y_1, \ldots, y_l).
\]

The Kalman filter start with the a priori distribution of \( x_1 \):

\[
\hat{x}_0^1 = x^0 \\
P_0^1 = Q^0.
\]  

(2.11)

Then, starting with \( k = 1 \) and increasing it iteratively, they repeat the measurement update,

\[
K = P_{k-1}^k D^T (D P_{k-1}^k D^T + R)^{-1} \\
\hat{x}_k^k = \hat{x}_{k-1}^k + K(y_k - D \hat{x}_{k-1}^k), \\
P_k^k = (I - KD) P_{k-1}^k
\]

(2.12)

and the time update,

\[
\hat{x}_{k+1}^k = A \hat{x}_k^k \\
P_{k+1}^k = AP_k^k A^T + Q.
\]

(2.13)

After reaching the last time instant \( n \), the Kalman smoother starts with \( k = n - 1 \) and decreases \( k \) iteratively:

\[
J = P_k^k A^T (P_{k+1}^k)^{-1} \\
\hat{x}_n^n = \hat{x}_k^k + J(\hat{x}_{k+1}^n - \hat{x}_k^k) \\
P_n^n = P_k^k + J(P_{k+1}^n - P_{k+1}^k) J^T
\]

(2.14)

As pointed out in Welling (n.d.), the computation of the covariance matrices \( P_k^l \) does not depend on the measurements \( y_k \). As a result, they can be precomputed and then applied to any set of data. This might be surprising at first, but it is not once we realize that since all the equations are linear and since the energy minimized during the Kalman filter estimation is quadratic, the estimate itself is a linear function of the measures; it can be summarized by a (huge) matrix multiplication, where the matrix is an appropriate pseudo-inverse of the matrix of the associated forward problem, and therefore its parameters do not depend on the measurement values.

Moreover, Welling (n.d.) also points out that whatever the initial matrix \( P_1^0 \), after successive applications of equations 2.12a, 2.12c, and 2.13b, \( P_{k-1}^{k-1} \)
and $P_k^k$ converge to some limits $\bar{P}$ and $\hat{P}$ (which naturally obey the relations $\dot{\bar{P}} = (I - KD)\bar{P}$ and $\dot{P} = A\hat{P}A^T + Q$). Thus, if we first compute these limits and then choose the initial distribution of the state for our estimation to be $Q^0 = \bar{P}$, then all the $P_{k-1}^k$ are equal to $\bar{P}$ and all the $P_k^k$ are equal to $\hat{P}$. This choice for $Q^0$ is equivalent to saying that at time 0, the system has already been running for a long time according to its natural evolution equations and that we have information on its state from previous measures. Since in fact such previous measures do not exist, this a priori variance is too small, which could have the consequence that the beginning of the estimated sequence underfits the measures; however, we did not observe important estimation errors at the beginning of our simulations. On the other hand, this choice is particularly convenient in terms of computational and memory cost.

The EEG-fMRI fusion algorithm can then be defined by the following steps:

1. Apply iteratively, equations 2.12a, 2.12c, and 2.13b until $P_{k-1}^k$ and $P_k^k$ converge to their limits $\bar{P}$ and $\hat{P}$, starting with $P_0^0 = Q$ and stopping after a number of iterations determined heuristically. We found on reduced data sets that after 250 iterations, all the coefficients of the variance matrices are within $+/-1\%$ of their value after 1000 iterations, and the quality of the reconstruction does not improve when increasing the number of iterations; therefore, we set the number of iterations for any data set to 300. Since this step is by far the one with a higher computational cost, we also used a specific rule to increase the convergence speed (see appendix A.2).

2. Compute $K = \bar{P}D^T(D\bar{P}D^T + R)^{-1}$ and $J = \hat{P}A\hat{P}^{-1}$

3. Apply the Kalman filter (equations 2.11a, 2.12b, and 2.13a) and the Kalman smoother (equation 2.14b) to the data to compute the estimation of the hidden states (among them neural activity) time courses, using for all $k$, $P_{k-1}^k = \bar{P}$ and $P_k^k = \hat{P}$.

In the schematic examples which we will show in section 4, we also use other filters. First, the extended Kalman filter and smoother (Arulampalam & Maskell, 2002; Welch & Bishop, 2006) used for nonlinear dynamical systems consist of linearizing the evolution and measure equations around the current estimate. Second, for a time-reversed Kalman filter and smoother, the first filtering action is performed backward in time (from the last until the first time instant), while the smoothing action is performed forward in time. The equations of this last filter are derived in appendix A.3.

3 Results

3.1 Artificial Data. We applied our EEG-fMRI fusion algorithm to synthetic data. We used a realistic head model (35,000 vertices) obtained from
Figure 2: Simulated data, spatial and temporal visualization. The cortical surfaces display the data at different instants, and the graphs show example time courses at one source location or EEG sensor. Sources: Activity of the 2000 sources on the cortex generated according to the neuronal model. Notice the correlations between instants separated by 100 ms. EEG: EEG measures at the same instants. fMRI: fMRI activity at the sources locations is also displayed on the cortical surface; it is notably much smoother in time.

a segmentation of a T1 MRI image and a significant number of sources (2000 patches of 15–20 dipoles) laid out on the cortex (see appendix A.1).

Figure 2 shows the maps of the cortical activity and the measures at several time instants, as well as typical time courses. An epoch of 60 s random activity was generated according to model 2.7 and sampled at 200 Hz. We can observe some temporal and spatial correlations in the activity. Figure S1 in the supplementary materials shows a scatter plot of the pair-wise correlations between sources’ activities as a function of the geodesic distance between these sources (the geodesic distance is the distance from one point to another while remaining on the surface of the cortex) and displays the average over all sources of their time courses’ autocorrelation.¹ Sources that

¹Supplementary materials referred to throughout this letter are available online at http://www.mitpressjournals.org/doi/suppl/10.1162/neco.2009.05-08-793.
are connected have an average distance of 6.4 mm and an average correlation of 0.65; there is an average correlation of 0.5 between sources that are distant from each other by 18 mm, and an average autocorrelation of 0.5 at a time lag of 150 ms.

EEG measurements at 64 electrodes were generated using a realistic geometry of the head and the sensors (they have the same temporal resolution of 200 Hz). The different head compartments were segmented from a T1 MRI using BrainSuite software (http://brainsuite.usc.edu/) for the bone and the skin, and software developed in our laboratory (Odyssee Team, INRIA) for the segmentation of white and gray matter, as well as for specific surface corrections, such as ensuring a minimal distance between the surfaces (Adde, 2005; Ségonne, Pons, Grimson, & Fischl, 2005). The matrix $B$ of the forward model was constructed using a boundary element method algorithm (Adde, Clerc, & Keriven, 2005; Kybic et al., 2005) implemented in the OpenMEEG software (http://www-sop.inria.fr/odyssee/software/OpenMEEG/).

fMRI time courses were generated at each source location using the balloon model in equation 2.5. The physiological parameters used are those from Friston et al. (2000). We sampled the signals at 1 Hz. Table 1 shows the values used for the parameters. The variances of the EEG and fMRI measure noises, $\sigma_{\text{eeg}}$ and $\sigma_{\text{fmri}}$, were chosen equal to the variances of the noise-free measures (such that the resulting SNR in the measures is 50%).

3.2 Estimation of Source Activity. Estimation of source activity was then performed from these data. We performed three estimations: one using both EEG and fMRI data and the other two based on only one of the two data sets.

Figure 3 shows the estimation results. The usual consideration about the differences in temporal and spatial resolutions of EEG and fMRI appears clearly: the EEG reconstruction is smooth in space but varies fast in time, whereas the fMRI reconstruction shows more spatial details but varies much slower in time. The fusion estimate seems to combine efficiently the information provided by each modality and hence to be the most accurate.

3.3 Quantifications of Estimation Efficiency. We start to quantitatively measure how good these three estimations are by quantifying the fit between all estimated signals ($\hat{u}$) and true signals ($u$). The most common measure is correlation $<u, \hat{u}>/||u||_2||\hat{u}||_2$, but since correlation can be 1 when $u$ and $\hat{u}$ are not equal but only proportional, we also define the percentage of variance explained by the estimate as $1 - ||u - \hat{u}||_2^2/||u||_2^2$. Table 2 displays these values (part 1, rows 1–3, column 1), as well as fit values between the EEG and fMRI time courses predicted by the estimates and the original time courses (columns 2 and 3). The fourth row shows values for the true sources’ signals. Obviously its fit to itself is 100%, but we can verify here that the noise in EEG and fMRI measures is about 50% of the total variance.
Figure 3: Estimation results. Comparison of the three reconstructions using the Kalman filter and smoother on EEG data only, fMRI data only, and both simultaneously: activity maps at 5 instants separated by 100 ms (the same color map is used for the four displays), and example time courses of one source (at the location shown with the white arrow). On the displays of time courses, the five black stars show the instants corresponding to the five activity maps. On the activity maps, the EEG reconstructions are smooth spatially, but they change fast, whereas the fMRI reconstructions show more spatial details but change more slowly. On the time courses also, we notice that EEG can capture fast variations of the activity, whereas fMRI estimates accurately only the low-frequency part of the activity. The fusion estimate seems to combine efficiently the information provided by each modality: indeed it puts together the fast-changing, spatially extended components and slow, spatially detailed components to produce the activity maps and time courses that clearly, by eye, are the closest to those of the true activity.

(since the source predicts only about 50% of the variance of the measures). We notice that the Kalman estimations explain the measures slightly better, which indicates a small overfit. The variance of the true source explained by the fusion estimate is almost equal to the sum of the variances explained by each modality alone.
Table 2: Quantification of Estimation Accuracies in the Case of Three Different Simulations.

<table>
<thead>
<tr>
<th>% Correlation (and % of Variance Explained)</th>
<th>Original Source</th>
<th>EEG</th>
<th>fMRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Activity generated according to the dynamical system</td>
<td>Fusion estimate</td>
<td>43.1 (18.6)</td>
<td>75.5 (56.9)</td>
</tr>
<tr>
<td></td>
<td>EEG estimate</td>
<td>29.7 (8.8)</td>
<td>75.5 (56.9)</td>
</tr>
<tr>
<td></td>
<td>fMRI estimate</td>
<td>33.9 (11.5)</td>
<td>28.2 (7.9)</td>
</tr>
<tr>
<td></td>
<td>Source</td>
<td>100 (100)</td>
<td>74.2 (55.1)</td>
</tr>
<tr>
<td>2. Sparse activity</td>
<td>Fusion estimate</td>
<td>11.0 (1.1)</td>
<td>16.2 (2.2)</td>
</tr>
<tr>
<td></td>
<td>EEG estimate</td>
<td>7.2 (−14.3)</td>
<td>31.1 (8.9)</td>
</tr>
<tr>
<td></td>
<td>fMRI estimate</td>
<td>7.7 (0.4)</td>
<td>1.5 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Source</td>
<td>100 (100)</td>
<td>17.2 (3.0)</td>
</tr>
<tr>
<td>3. Smooth activity</td>
<td>Fusion estimate</td>
<td>74.1 (54.6)</td>
<td>93.5 (87.4)</td>
</tr>
<tr>
<td></td>
<td>EEG estimate</td>
<td>54.5 (27.6)</td>
<td>93.5 (87.4)</td>
</tr>
<tr>
<td></td>
<td>fMRI estimate</td>
<td>63.6 (40.4)</td>
<td>63.2 (39.3)</td>
</tr>
<tr>
<td></td>
<td>Source</td>
<td>100 (100)</td>
<td>93.1 (86.7)</td>
</tr>
</tbody>
</table>

Notes: We use two different measurements of how the estimate \( \hat{u} \) fits the real signal \( u \), both expressed in percentages. The first is the correlation \( \langle u, \hat{u} \rangle / \|u\|_2 \|\hat{u}\|_2 \); it does not inform whether signals were estimated with the correct amplitude (the correlation is 1 if the signals are only proportionals). Thus, we also use the percentage of variance of source \( u \) explained by estimate \( \hat{u} \), defined as \( 1 - \|u - \hat{u}\|_2^2 / \|u\|_2^2 \). Note that it can be negative even if the correlation is positive, in the case where subtracting \( \hat{u} \) to \( u \) does not decrease the variance of \( u \) (i.e., when the part of the variance of \( \hat{u} \) that really accounts for some variance present in \( u \) is less than the part of pure estimation error).

We would also like to quantify what we observe in Figure 3 about the spatial and temporal characteristics of each estimate: How well does the algorithm estimate each temporal and spatial frequency component of the hidden activity? To this aim, we developed a simple method to smooth spatially any activity on the cortex in a way that respects its geometry. First, the geodesic distance between each pair of sources is calculated iteratively, forming the matrix \( D = (D_{ij} = \|v_i - v_j\|_{mesh}) \). Each (true or estimated) activity \( u \) can be smoothed with a distance parameter \( \sigma \) into an activity \( \tilde{u} \) defined for each source \( i \)

\[
\tilde{u}_i(t) = \frac{\sum_j \exp \left( - \frac{D_{ij}^2}{2\sigma^2} \right) u_j(t)}{\sum_j \exp \left( - \frac{D_{ij}^2}{2\sigma^2} \right)}.
\]  

(3.1)

This allowed us to isolate specific temporal and spatial frequency components of the original source and of its reconstructions and thus evaluate (using the ratio of explained variance defined above as a criterion) the Kalman filter and smoother efficiency in different frequency regions. The result is displayed in Figure 4.
Figure 4: Quantification of the efficiency of the reconstructions at different temporal and spatial scales. True cortical activity and different reconstructions were filtered temporally and spatially in order to measure the efficiency of the reconstructions at different temporal and spatial scales. (A) Efficiencies for the fusion estimate. (B, C) Efficiencies for the EEG/fMRI estimates. The efficiencies for the fusion estimate are superimposed in light gray. We can visualize here that at least in our simulations, EEG has a spatial resolution of more than 1 cm, and fMRI has a temporal resolution of more than 2 s. Thus, the information they bring on neural activity overlap in the domain of more than 1 cm and 2 s. (D) Efficiencies for the fusion estimate again. The sum of EEG and fMRI efficiencies is superimposed in light gray: both are very similar, except in the region of more than 1 cm and 2 s, where the sum is higher because EEG and fMRI cooperate, but also share some redundant information. (Contrary to the fMRI and fusion efficiencies calculated based on the estimation data presented in Figure 3, the EEG efficiency, which was too noisy when calculated only from these data, is an average of efficiencies calculated from 40 simulations of a 5 min period of activity.)

We observe that EEG reconstructions start to recover some activity at spatial resolutions above 1 to 2 cm, while fMRI reconstructions start at temporal resolutions above 2 s. The EEG-fMRI reconstructions efficiently sum up the information provided by the two measures since the variance explained by the EEG-fMRI reconstruction is similar to the sum of variances explained by EEG and fMRI independently. However, it is inferior in the
region of more than 1 cm and 2 s, where both modalities are effective. In this region, the two modalities bring complementary information, but also part of the information is redundant (this is obvious when the sum is more than 1).

This does not mean that the fusion estimate is simply a sum of the EEG and fMRI estimates. Indeed such a sum would not correctly explain the EEG and fMRI measures. However, it is right to consider the fusion estimate as a nontrivial combination of the two independent estimates, which more precisely takes into account the temporal low-frequency, spatial high-frequency information provided by fMRI data; the temporal high-frequency, spatial high-frequency information provided by the EEG data; and the complementary or redundant temporal and spatial low-frequency information provided by the two data sets.

3.4 Estimation of Other Specific Forms of Activity. The model that was used to generate the artificial activity (see equation 2.7) is quite simple and does not necessarily reflect the characteristics of either spontaneous or evoked activities in the brain. We investigated to what extent the fusion algorithm remains robust when estimating activities different from what it expects.

We first created a very sparse activity, where a strong burst of activity occurred during a brief period of time (50 ms) at a specific location. Second, we created a smooth activity by filtering the normal activity generated by our physiological model, equation 2.7, temporal smoothing by a gaussian with a standard deviation of 0.5 s, and spatial smoothing as in (equation 3.1) with $\sigma = 1$ cm. Figures 5 and 6 show these simulated data and the estimation results (once again performed on EEG alone, fMRI alone, or both together). Figure S2 in the supplementary materials displays the spatial and temporal correlations of the smoothed activity: there is an average correlation of 0.5 between sources distant by 25 mm and an average autocorrelation of 0.5 at a time lag of 500 ms. Table 2 shows the quantification of estimation efficiencies.

The result is as expected from the considerations made previously about frequency components: although in both cases EEG and fMRI succeed in collaborating to produce a fusion estimation better than they do alone, the estimation of the sparse activity remains quite poor. On the contrary, the estimation on the smooth data set is very satisfying: more than 50% of the variance of the activity of the true sources is explained.

4 Discussion

We first discuss the new approach that we propose for EEG and fMRI analysis and fusion through the use of dynamic system modeling and of adaptive filters. Then we examine the specific choice of the Kalman filter and smoother—its advantages and its limitations.
Figure 5: Simulated cortical activity and estimation results in the case of a much localized activity. This activity consists of the strong activation over 50 ms of a single source. (A) Activity maps. The color maps are not the same in the four displays, as can be seen in the time courses. (B) Original and estimated time courses of the activated source. (C) Original and estimated time courses of the source at the center of the activation found by the EEG reconstruction. We observe that the EEG-only estimation results in a large activated area (which in addition is not even centered on the true activated source), whereas the fMRI-only estimation results in an activity, correctly localized, that lasts for several seconds. The fusion estimate shows improvements compared to previous estimates but does not succeed in reconstructing a sparse activity.

4.1 New Approaches to EEG-fMRI Fusion and Proof of Feasibility.

4.1.1 Spatiotemporal Inverse Problem. Beyond the specific choices made in this work, we advocate an approach to fMRI-EEG fusion that consists of estimating the whole spatiotemporal profile of cortical activity during an arbitrary period of time, as illustrated by the schematic example in Figure 7. Note that because of the use of a gray-levels color map, we already notice at a local level a dual-mode aspect of the fusion estimation, which will be discussed later. (Figure S3 in the supplementary materials displays the same data using a blue-red color map, which allows a better appreciation of the estimation quality at the global level.) This is unlike standard approaches, which typically target the localization of a specific activity (in
Figure 6: Simulated cortical activity and estimation results in the case of a very smooth activity. This activity consists of the output of the modeling dynamic system, later smoothed temporally and spatially (see Figure S2 in the supplementary materials for a measure of the correlations of this activity). We can make similar remarks as in the case of the first simulation (see Figure 4—that the EEG estimate is smoother spatially than the fMRI estimate, the fMRI estimate is smoother temporally, and the fusion estimate realizes a combination of both that better fits the true activity). Here however, the final estimation is much more accurate.

general, evoked by a specific stimulation or cognitive task) or the estimation of temporal patterns at a specific source locations. However, some other works have used approaches similar to ours (Trujillo-Baretto et al., 2001; Halchenko et al., 2004; Kiebel & Friston, 2004).

This approach will be especially well suited for analyzing simultaneously acquired EEG and fMRI data in a paradigm-free context. There are no repetitions of specific stimulations or tasks—for example, the study of ongoing activity at rest, epileptic activity, or such realistic situations as watching a movie (Hasson et al., 2004). Despite its ambition, this approach has the advantage of being very simple. We try to estimate an unknown cortical activity by taking into account all the information contained in EEG and fMRI data sets, given some physiological models of the EEG and BOLD signals. Since this inverse problem is intensively underdetermined, it is necessary to add some constraints on this cortical activity, which we intend
Figure 7: Schematic illustration of the EEG-fMRI framework. EEG and fMRI are measurements of a hidden cortical activity (represented by a matrix of 100 sources × 100 time samples). In the case of EEG, some spatial information is lost, whereas in the case of fMRI, some temporal information is lost. Reconstruction can be performed using either or both of these two measures. The estimate should explain the measure and at the same time minimize a specific energy—here the mean square (L2) norm of the reconstructed activity. See Figure S3 in the supplementary materials for a color display of the same data.

to recover. This is done through the introduction of a simple model of the cortical activity itself.

As a consequence, our algorithms are fully symmetric in the sense that they do not use only a part of the information provided by EEG and fMRI data sets (e.g., use only the spatial information of fMRI) or consist of a two-step procedure where the information given by the first data set is used to constrain the analysis of the second one.

4.1.2 Dynamic System Formulation and Adaptive Filters. Of course, such a problem cannot be handled directly because of the very high dimensionality of the unknown activity, and this leads to the second novelty of our approach: the use of adaptive filters. Indeed, if we go beyond the common concept that EEG has a high temporal resolution but a low spatial resolution
and fMRI has a high spatial resolution but a low temporal resolution, we notice some differences between time and space. For a given experiment, the “space” is given—it is the set of all cortical sources’ places—and it is hardly possible to subdivide this set into smaller subsets where one could estimate the activity independently since the signal in one EEG electrode potentially depends on activity in all cortical locations. On the contrary, the “time” is not so well defined in advance, it can be arbitrarily long, there can be several trials, and it is obvious that although BOLD signals result from temporal smoothing of several seconds or even tens of seconds, the activities at time instants separated by more than 1 minute can be considered independent.

Hence the idea of using iterative filters, which take into account the entire space at each step and proceed sequentially, while keeping in memory information from the previously visited time instants (this memory being limited in range; nothing, for example, is remembered about time instants distant by more than a minute). For a better sense of what such a filter should do, let us imagine that at a given instant, the filter sees a high-amplitude EEG measurement. It should remember that there was some uncertainty about its exact location, so that when it later considers fMRI measurements a few seconds away, which brings in more spatial information about this activity, it will be able to reduce this uncertainty.

4.1.3 Proof of Feasibility on Realistic Data. Attempting to estimate the whole spatiotemporal activity of the cortex raises two questions: Is it possible practically on data sets of realistic size and at a reasonable computational cost? and How much of this activity can be recovered effectively?

We proposed an algorithm based on the Kalman filter and smoother that allowed us to estimate the activity of 2000 sources on the cortical surface sampled at 200 Hz. This answers positively the first question, but the computational cost is quite high.

The convergence of the variance matrices (300 iterations) took 24 hours on an AMD Opteron 285, 2.6 GHz and occupied 5 Gb RAM. The time complexity of this convergence is \( O(n^2 f_s) \), where \( n \) is the number of cortical sources and \( f_s \) the sampling frequency (indeed, the limiting factor is the multiplication of matrices of size \( O(n) \), repeated \( O(f_s) \) time to reach the convergence of the variance matrices). The space complexity is \( O(n^2) \).

Once the convergence of the variance matrices is reached and the matrices needed in equations 2.12b and 2.14b are computed, the estimation itself is faster (it took 3 hours to estimate 1 minute of activation—12,000 iterations for a 200 Hz sampling). The complexity here is \( O(n^2 f_s T) \), where \( T \) is the total time duration of the experiment.

It is not surprising that the computation took so long to estimate an unknown activity of size \( n f_s T \), equal to \( 2000 \times 200 \times 60 \) in the case of the presented simulations. However, it seems to us that the complexity could be significantly decreased if the variance matrices were handled carefully as they seem to have some structure. Possibilities are using an approximated
sparse representation, some other compact representation (e.g., PCA, tensor product), or a multilevel approach where a coarse and a fine representation of the brain coexist.

The several simulations we ran and the frequency analysis we proposed answered the second question: the Kalman estimation performs particularly well for estimating a slow, spatially smooth activity, and in the case of EEG-fMRI fusion, the information brought by the two measurements is effectively summing up together. In the rest of this discussion, we go into details on this Kalman filter and smoother.

4.2 Kalman Filter and Smoother and Whether It Is Adequate to Physiological Models.

4.2.1 Kalman Filter and Smoother: The Simplest Filtering Technique. As explained in section 2, the Kalman filter is an iterative filter that estimates at increasing times $t$ the (gaussian) a posteriori distribution of the hidden state $x_t$ given measures $y_{t'}$ for all $t' < t$, whereas the Kalman smoother, which operates backward (from the last time instant to the first) on the results of the Kalman filter, estimates the a posteriori distribution of states $x_t$ given all measures $y_t$. Figure 8 (left) illustrates the principle of these two filters in the case of a very simple dynamical system. The hidden state is a one-dimensional Brownian motion, and the measure is a noisy version of this hidden state (see Figure 8A). Figures 8B and 8C show the mean and variance of the a posteriori distribution for the Kalman filter and smoother, respectively. One immediately notices the smoother aspect of the Kalman smoother. Figure 8D, which compares the two estimates, shows that the filter detects changes (increases or decreases) with a small delay, which disappears in the smoother estimation.

In this schematic example, the difference is still small, but using the Kalman smoother becomes essential in the case of fMRI measurements, which provide information on neural activity with a delay of several seconds. Figure S4 in the supplementary materials shows the estimated time courses of all the hidden states (neural activity and hemodynamic states) for a large number of sources after applying only the Kalman filter to the data, and after applying successively the Kalman filter and smoother (data from a reduced data set—1000 sources, 20 Hz—where the activity to be estimated is sparse, as our second simulation in Figure 5). One observes that the information about the activation is somehow stored in the hidden state estimates after the Kalman filter has been applied, but not in an accurate way, and only once the Kalman smoother has been applied is this information “put back where it belongs.”

As noted in section 2, the main advantage of the Kalman filter is its linearity—the fact that the estimate is a linear function of the measures. Hence, it is enough to compute only once the parameters of this linear function (see the matrices in equations 2.12–2.14). The estimation is then
Figure 8: Simple illustration of the Kalman filter, Kalmer smoother, and extended Kalman filter and smoother. (A) Hidden state of a linear dynamic system and measure points. (B) Display of the Kalman filter a posteriori distribution (light gray; bold line = mean, dashed lines = mean $\pm$ SD) superimposed on the true hidden state (black). (C) Kalman smoother a posteriori distribution (dark gray). (D) Comparison between filter and smoother. We can observe the smooth-er (!) aspect of the smoother estimate, and the correction of delays of the filter estimate in the detection of increases or decreases. (E) Hidden state of a nonlinear dynamic system and two sets of measures (light gray: linear measure; dark gray: nonlinear absolute value measure—no information about the sign of the hidden state). The estimation is valid (F). When more weight (smaller variance) is given to the nonlinear measurement (G), the estimation can diverge and become strongly erroneous (H).
relatively fast on any data acquired on the same subject. Also, the variance matrices converge to some limits, which allows computing and storing only these limits (this convergence can be observed, for example, in Figures 8B and 8C: the width of the error band initially decreases but later remains constant).

4.2.2 Inadequacy of the Linearity Hypothesis to the Electric-Metabolic Relation. Since the EEG forward model is linear and the balloon model that describes the relation between metabolic activity and the BOLD signal is well approximated by its linearization (Deneux & Faugeras, 2006), the physiological models for EEG and fMRI models fit the Kalman requirements well.

This is not the case for the relation between electrical and metabolic activity. Here, we supposed a linear relation (through the scaling factor $\varepsilon$) between the “neural activity” $u$ representing the dendritic currents responsible for EEG measures and the metabolic demand responsible for the hemodynamic response and the fMRI BOLD signals. Several studies are in favor of other kinds of relation, such as the power of electric signals in a specific frequency domain (Martinez-Montes et al., 2004; Kilner, Mattou, Henson, & Friston, 2005). A complete model taking into account both the electric-metabolic interaction at the neuronal level and the vascular response can be found in Sotero and Trujillo-Barreto (2008).

Before going into the details of any particular model, we can notice that since electrical currents at a given source location pointing to opposing directions can generate the same BOLD signal, a good model for the electric-metabolic relation must be strongly nonlinear. And although we mentioned an adaptation of the Kalman filter to nonlinear systems, called the extended Kalman filter (Arulampalam & Maskell, 2002; Welch & Bishop, 2006), which performs local linearizations, it cannot perform well in the case of such strong nonlinearities. As an illustration, we slightly changed the model of Figure 8A by adding a second measurement, which this time is a noisy version of $|x|$, the absolute value of the hidden state (see Figure 8E). In a first case where both measurement errors have same variance, the (extended) Kalman filter and smoother succeed in estimating the hidden state, with an error reduced compared to the case of only one measurement (see Figure 8F). However, when more weight is given to the second measurement by decreasing its noise variance and increasing the noise variance of the first (see Figure 8G), the estimation fails: the sign of the estimation is sometimes erroneous, and error bars become invalid (see Figure 8H). This is because when there are small errors in the estimate, they can result in larger errors in the local linearizations, which can cause the filter to update the estimation in the wrong direction (reach local minimums) and diverge.

It thus appears that before it can be applied to real data, our method should include some additional nonlinear modeling of the electric-metabolic coupling. But such strong nonlinearities may be difficult to handle for the filtering algorithm. We show below on a simplified example that a
careful and well-balanced choice of the nonlinear model and the filtering algorithm can overcome this difficulty.

First, we suppose that the BOLD response depends mainly on the power of the neural currents in a specific frequency domain. It is convenient then to describe the neural currents seen by both the EEG and fMRI measures as

\[ u(t) = \chi(t) \cos(\varphi(t)), \tag{4.1} \]

where \( \chi(t) \) is the envelope of the signal and \( \varphi(t) \) a phase shift. \( \chi(t) \) is proportional to the instantaneous power of the signal, and we use it in this new model to drive the metabolic response. The evolution equation for the flow (see equation 2.5a) becomes

\[ \dot{f} = \varepsilon \chi - \kappa_s \dot{f} - \kappa_f (f - 1). \]

In brief, the fMRI measure depends on \( \chi(t) \), according to the balloon model equations 2.5, whereas the EEG measure depends on \( u(t) \) according to the EEG forward problem (equation 2.4). The fusion algorithm will consist in estimating \( \chi \) and \( \varphi \) given the measures. The main advantage in this formulation is that the algorithm will use the fMRI measures to estimate \( \chi \) but not \( \varphi \), and thus avoid the problems of local minima and divergences shown in the counterexample of Figure 8H that would occur if the fMRI measure was used to estimate \( u \) directly due to the rapid changes in the sign of \( u \). On the contrary, \( \chi \) can vary at a much slower rate than \( u \)—more in accordance with the weak temporal resolution of fMRI (see Figure 9). Note, however, that in the Kalman formulation we present below, the estimation of \( \chi \) cannot be completely dissociated from the estimation of \( \varphi \) since covariance matrices introduce correlations between the two quantities.

We summarize the evolution and measure equations of this new model (here, each hidden states \( \chi, \varphi, f, v, q \) are vectors with as many elements as cortical locations):

\[
\begin{align*}
\dot{\chi} &= -\lambda_\chi \chi + \xi_\chi, \quad \xi_\chi \sim N(0, Q_\chi) \\
\dot{\varphi} &= 2\pi f_c - \lambda_\varphi (\varphi - \tilde{\varphi}) + \xi_\varphi, \quad \xi_\varphi \sim N(0, Q_\varphi) \\
\dot{\chi} &= \varepsilon \chi - \kappa_s \dot{\chi} - \kappa_f (f - 1) \\
\dot{v} &= \frac{1}{\tau} (f - v^{1/\alpha}) \\
\dot{q} &= \frac{1}{\tau} \left( f \frac{1 - (1 - E_0)^{1/\alpha}}{E_0} - v^{1/\alpha} \frac{\dot{q}}{v} \right) \\
y_{\text{eeg}} &= B(\chi \cos(\varphi)) + \eta_{\text{eeg}}, \quad \eta_{\text{eeg}} \sim N(0, \sigma_{\text{eeg}}^2 I) \\
y_{\text{fmri}} &= V_0 (a_1 (v - 1) + a_2 (q - 1)) + \eta_{\text{fmri}}, \quad \eta_{\text{fmri}} \sim N(0, \sigma_{\text{fmri}}^2 I)
\end{align*}
\]

\( \chi \) is driven by the same Ornstein-Uhlenbeck process as in our earlier model, equation 2.7, whereas \( \varphi \) is driven by another innovative process with a mean corresponding to oscillations at the frequency \( f_c \), the center of the frequency
Figure 9: Hidden states of the nonlinear model for neural activity. The signal $u$ is defined by two hidden states $\chi$ and $\varphi$, which can vary much more slowly than $u$ itself (concerning the phase, it changes more slowly only after subtraction of the average trend $2\pi f_c$).

domain considered, and a term that pulls the phase toward its local average ($\tilde{\varphi}^{\sigma_{\varphi}}$ is defined as in equation 3.1 and depends on the smoothing parameter $\sigma_{\varphi}$), thus ensuring some spatial smoothness to the phase. The variance matrices $Q_\chi$ and $Q_\varphi$ can include some correlation between different cortical locations as in equation 2.8. In fact, we introduced correlation only in $Q_\chi$ since the spatial smoothness of the phase is already controlled by $\lambda_{\varphi}$ and $\sigma_{\varphi}$. Additionally, we did not allow $\chi$ to become negative (at each iteration of the generation of the simulated data, whenever it would become negative, it was reset to zero).

The simplest estimation procedure for this model would be to apply the extended Kalman filter, which performs successive local linearizations at the current estimate. Such linearizations will be accurate only if the current estimation of $\varphi(t)$ is good enough (an error in the linearization would first appear during the EEG measure update but later propagate to all the covariance matrices). It is critical, then, that during the first sweep of the algorithm, the estimate be as good as possible. But this is not the case when performing the Kalman filter, which gathers the information provided by fMRI on activity at time $t$ only when considering fMRI measurements some seconds after $t$. We mentioned in section 4.2.1 that in the linear case, the Kalman smoother allows this information to be used efficiently, but it might not be true in the nonlinear case if the first sweep makes too many errors. Therefore, we designed a modification of the Kalman filter and smoother that first performs a backward sweep (estimate activities at time $t$ given all
Table 3: Quantification of Estimation Accuracies for the Nonlinear Model in the Case of Two Different Simulations Using 10 Sources.

<table>
<thead>
<tr>
<th>% Correlation (and % of Variance Explained)</th>
<th>Original Source</th>
<th>Envelope of Source</th>
<th>EEG</th>
<th>fMRI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Activity generated according to the dynamical system</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusion estimate</td>
<td>29.0 (−14.0)</td>
<td>34.1 (−2.0)</td>
<td>83.7 (67.6)</td>
<td>73.7 (54.0)</td>
</tr>
<tr>
<td>EEG estimate</td>
<td>24.7 (−17.8)</td>
<td>17.6 (−33.4)</td>
<td>82.2 (64.9)</td>
<td>29.3 (−15.3)</td>
</tr>
<tr>
<td>fMRI estimate</td>
<td></td>
<td>50.7 (25.5)</td>
<td></td>
<td>88.5 (72.9)</td>
</tr>
<tr>
<td>Source</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>82.5 (68.0)</td>
<td>84.2 (70.9)</td>
</tr>
<tr>
<td><strong>Sparse activity (pulse)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusion estimate</td>
<td>42.2 (−35.1)</td>
<td>68.5 (44.0)</td>
<td>66.4 (40.0)</td>
<td>63.8 (40.7)</td>
</tr>
<tr>
<td>EEG estimate</td>
<td>23.2 (−86.4)</td>
<td>36.2 (−6.8)</td>
<td>54.2 (28.8)</td>
<td>28.9 (2.1)</td>
</tr>
<tr>
<td>fMRI estimate</td>
<td></td>
<td>49.6 (18.1)</td>
<td></td>
<td>57.8 (30.3)</td>
</tr>
<tr>
<td>Source</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>37.7 (14.2)</td>
<td>77.1 (59.4)</td>
</tr>
</tbody>
</table>

Notes: The measurements are the same as in Table 2, i.e., correlation and percentage of variance explained. For every estimation, we quantify the fit between the estimated sources and the original simulated sources ($u = \chi \cos \varphi$), as well as between the estimated and original envelopes (hidden state $\chi$), and the predictions of EEG and fMRI measures. Note that in the case of fMRI-only estimation, there is no information on $\varphi$, so no number is given for the fit to the original source and for the fit to EEG. As in Table 2, negative values for the percentage of variance explained can be obtained, although the correlation is positive and indicate only that the part of estimation that is erroneous is larger than the part that is correct.

We have run this algorithm on a highly reduced data set where the cortical surface was downsampled to 10 cortical patches and only three EEG electrodes were considered (so as to preserve the indetermination of the EEG backward problem). We generated 1 min long patterns of activity, sampled at 200 Hz, with an average frequency of the oscillations $f_c = 10$ Hz. First, we generated a cortical activity and measures according to the model, equation 4.2. Figure 10 shows the simulated data and Figure 11 and Table 3 the estimation results. We compare estimations obtained using a simple multiplication of EEG signals by the pseudo-inverse of $B$, and those obtained by applying the filtering technique to either the EEG alone; fMRI alone, or both together (in the case of fMRI alone, since there is no information on the phase $\varphi$, only the estimated envelope is displayed). We see that the envelope of the signals is best estimated using fMRI alone, whereas the EEG-fMRI fusion estimates the signals better than EEG only. We also generated a 1 sec width pulse of activity and noisy fMRI and EEG measures (see Figure 12), and the estimation results (see Figure 13 and Table 3) are even more significant. The fusion estimate clearly provides the best estimate of the signals and of their envelopes. It is also interesting to note that applying the filter to EEG measures results in a more focused estimation (less activated sources) than the pseudo-inverse multiplication.
Figure 10: Schematic cortical activity and EEG and fMRI measures generated according to a nonlinear model. The model generates cortical oscillations in a frequency domain centered on \( f_c = 10 \) Hz, EEG signals are obtained by multiplication with the forward problem matrix, and fMRI signals are obtained through the balloon model, using the envelope of the cortical signals as input. This simulation used 10 sources and 3 electrodes and generated data for 60 s at 200 Hz (and 1 Hz for the fMRI).

We also ran the algorithm on a reduced data set with 100 sources and the 64 EEG electrodes. Results are shown in Table 4 and are similar, except that in the case of the simulation according to the model equations, the fMRI-EEG estimation does not show improvement to estimate the source compared to the EEG-alone estimation (although it still shows a clear improvement in estimating the envelope). This can be explained by the fact that some activity is seen by the fMRI measure but not by the EEG measure, so this activity is finally estimated with a correct amplitude, but an erroneous phase, leading to a worse fit than if it was not estimated at all. On the contrary, when applied on a sparse pulse activity, such ambiguity on the phase due to overlapping sources does not exist, and the fusion estimate shows improvements on all aspects compared to other estimates.

In order to apply such algorithm to real-size data, some additional steps should be taken, in particular to reduce the computation cost of the method. Indeed, in this nonlinear modification of the Kalman filter, it is not possible
Figure 11: Estimations using a Kalman-based nonlinear filter. The nonlinear filter has been applied to EEG data alone, fMRI data, or both together, and a simpler estimation based on EEG data and using the pseudo-inverse of the matrix of the forward problem is shown for comparison (blue traces). The global views allow visualizing that the envelope of the signals is better estimated using EEG+fMRI compared to EEG alone (thanks to information provided by fMRI), whereas the insets allow a comparison of the estimations of the signals and also detect some errors in the estimation of the phase.

to precompute the covariance matrices independent of the data, since they depend on the data and, in particular, the current estimate of the phase. Therefore, it would be necessary that the covariance information on the hidden states be stored in a more compact way than full matrices, both for memory and computational cost purposes. Additionally, special care should be given to the issue raised above about indetermination in the phase due to overlapping sources of activity. In fact, it is likely that more specialized nonlinear filters, such as an unscented Kalman filter (Julier & Uhlmann, 1997) or particle filter (Arulampalam & Maskell, 2002), will perform better than the extended Kalman filter or its modification presented here.

4.2.3 Characteristics of the Estimated Signals. We now return to the initial linear model to discuss how much of the original source signal can be recovered by the estimation. Let us start with a counting remark: if we want to
estimate the cortical activity on a mesh of $n$ sources, sampled at frequency $f_s$, using EEG signals from $n_e \ll n$ sensors and fMRI signals sampled at frequency $f_i \ll f_s$, it is obvious that the ratio $(n f_s)/(n_e f_s + n f_i)$ between the number of unknowns and the number of measures is disproportionate. Thus (especially in the case of linear systems), a tremendous number of activities are invisible to both EEG and fMRI, or, in other words, any estimated activity will belong to a specific, reduced family of functions sharing some characteristics. These characteristics depend on both the constraints encoded in the modeling of cortical activity and the algorithm chosen for the estimation.

In fact, we observed in our results that the estimations were particularly smooth; more precisely, they seem to be the sum of an activity that is smooth in time, presumably mostly due to the fMRI measures, and of an activity that is smooth in space, presumably mostly due to the EEG measures. This means that large, smooth activities or waves of activity, spanning the whole brain as observed in epilepsy or cortical areas as observed, for example, in the visual cortex (Ferezou et al., 2006), can be estimated accurately, whereas the estimation of a focused activity remains problematic.

Figure 12: Simulation of a sparse oscillating activity (1 s width pulse in one source) and noisy EEG and fMRI measures.
Figure 13: Estimation result for a pulse of oscillating activity. The fusion estimation clearly takes advantage of the temporal information provided by EEG and spatial information provided by fMRI. Note that the EEG-only estimate using the filter is already much more focused spatially than the pseudo-inverse reconstruction. Note also that there is no error in the estimation of the phase here since there is no confusion between overlapping sources.

Table 4: Quantification of Estimation Accuracies for the Nonlinear Model in the Case of Two Different Simulations Using 100 Sources.

<table>
<thead>
<tr>
<th>Activity generated according to the dynamical system</th>
<th>% Correlation (and % of Variance Explained)</th>
<th>% Correlation (and % of Variance Explained)</th>
<th>EEG</th>
<th>fMRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusion estimate</td>
<td>10.9 (−31.0)</td>
<td>40.9 (13.2)</td>
<td>71.4 (49.2)</td>
<td>73.6 (50.8)</td>
</tr>
<tr>
<td>EEG estimate</td>
<td>11.3 (−11.2)</td>
<td>23.0 (−0.3)</td>
<td>69.3 (45.2)</td>
<td>40.5 (15.8)</td>
</tr>
<tr>
<td>fMRI estimate</td>
<td></td>
<td>42.7 (18.2)</td>
<td>78.3 (55.9)</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>70.2 (49.2)</td>
<td>71.8 (51.5)</td>
</tr>
</tbody>
</table>

Sparse activity (pulse)

| Fusion estimate                                     | 58.2 (32.6)                               | 64.4 (41.2)                               | 68.4 (45.6) | 38.3 (14.3) |
| EEG estimate                                        | 31.2 (−1.6)                               | 35.7 (7.8)                                | 67.0 (44.1) | 19.2 (0.6)  |
| fMRI estimate                                       |                                            | 53.6 (26.1)                               | 43.1 (18.1) |      |
| Source                                              | 100 (100)                                 | 100 (100)                                 | 66.3 (43.9) | 53.9 (29.0) |
We show that this feature is a characteristic of the Kalman filter and that the use of alternative filtering techniques might produce different results. More precisely, the Kalman filter and smoother, based on the assumption of gaussian distributions, minimize the L2 norm of the measurement errors and the hidden state innovation process. In Figure 14, columns 2 and 3, we show example estimations similar to the one in Figure 7, where we minimized either the L2 norm or the L1 norm of the source (while maintaining
that the measures are predicted without error). Figure S5 in the supplementary materials displays the same data using a blue-red color map, which allows a better appreciation of the quality of some estimations.

We can see that a Dirac is perfectly estimated by the L1-norm algorithm and very poorly by the L2-norm algorithm, which prefers the sum of two components—one spread in time, the other spread in space. In general, the L1 norm will result in sparse estimates and the L2 norm in smooth estimates (with this particularity, when the measures have very different time-space resolutions, one can distinguish two components—one smooth in time, the other smooth in space). This is actually a well-known fact in the image processing domain (Aubert & Kornprobst, 2006; Morel & Solimini, 1995).

We also wanted to compare, at least schematically, our fusion algorithm with others found in the literature. Since the most popular techniques rely on constraining the EEG reconstruction by fMRI results, we designed a similar schematic reconstruction. First, the variance of each source is computed

Figure 14: Comparison of activity reconstruction using the minimization of an L2 or L1 norm or a method equivalent to the fMRI-constrained EEG reconstruction. (A) Each row is a different simulation: from top to bottom, first a Dirac activity, representing a unique spatially and temporally localized activity (since it is linear, any L2 reconstruction will be a convolution of the activity of the L2 estimate of a Dirac); then a sparse activity made of few events at different instants and location and with various widths; then a “shot noise” activity, where each instant and each space location is independent; then a “correlated noise” activity, smooth in both time and space; finally, the sum of a similar smooth background activity and smooth activities at a few spatial locations with variance 100 times higher (see the text for this last simulation). The first column shows the activity, and the next columns the estimation results obtained by the three methods on the simulated EEG and fMRI measures of this activity (see Figure 9). In the case of the schematic fMRI-constrained EEG reconstruction, the variances of the source used as an a priori are shown on the left with a grayscale color map. The L2 norm method produces smooth estimations (more precisely, the sum of a component smooth in time and sharper in space and of a second component sharp in time and smooth in space), whereas the L1 norm method produces sparse estimations. fMRI-constrained EEG reconstruction requires that only a small number of sources are activated during the time period, in which case it performs the best; otherwise, it performs more poorly. (B) Estimations are shown in the case of discrepancy between the activities seen by EEG and fMRI. Data are similar to the second example in Figure 13, except that one source is not seen by the EEG and another one is not seen by the fMRI. The L2 reconstruction still sees these activities, but only at the resolution specific to the modality that captured them, and with a decreased amplitude; the L1 reconstruction is less predictable. See Figure S5 in the supplementary materials for a color display of the same data.
from the activity of the original sources (rather than from the “fMRI” time courses, in order to ensure more accuracy). Second, the activity is reconstructed based on the EEG, using these variances as an a priori in the EEG inverse problem as in Liu et al. (1998).

The results appear in the right-most column of Figure 14A. It immediately appears that in the cases of shot noise or smooth activity, where the whole space is activated, the a priori on the total variances of the sources provides very little information, and the estimation result is quite similar to an estimation based on EEG alone. Besides, the case of a set of local activities is particularly interesting to better understand our method. In this case, the activity is quite sparse, but the active sources are not the same throughout time. In such a case, the fMRI-constrained EEG reconstruction, which requires a static constraint on which sources are active for the whole experiment, makes more mistakes in the estimation than the L1 and L2 algorithms, which allow variations in which sources are active across time.

In the case where only one source is activated (Dirac), the fMRI-constrained EEG reconstruction performs the best. We also produced (see the last row of Figure 14A) one example where only a small number of sources is activated during the entire experiment. In such a case also, it performs better than L1 and L2 reconstructions. From these simulations, it appears that our algorithm and any other algorithm that would use the same framework is particularly suited for an activity that can span large areas, even the entire brain, and does not occur at the same position at different moments. Therefore, it could be particularly useful for the analysis of experiments investigating ongoing activity, or any other complex activity, rather than experiments investigating the activity evoked by a simple paradigm.

Finally, we address what would happen in the case that there are discrepancies between the EEG and fMRI signals—for example, if some activity is seen by only one of the two modalities (Nunez & Silberstein, 2004). Since we noticed that the Kalman fusion estimate is a complex combination of EEG- and fMRI-alone estimates, such an activity will still appear in the fusion estimate, as it appears in the estimate from the modality that can see it, but with a reduced amplitude. Simulations shown in Figure 14B confirm this statement, whereas in the case of the L1 reconstruction, the shape of these sources is quite distorted after the estimation, which suggests that special attention should be paid to this issue in the design of L1 fusion algorithms. Note also that our method estimates only a cortical activity as sources are laid on the cortical surface. This makes sense in the case of the EEG, which detects subcortical activities; however, it means that the potential information of fMRI concerning deep activities is lost.

In conclusion, the Kalman filter-smoother algorithm we propose is best suited for estimating activities that are smooth in both time and space and where a large portion of the cortical surface is activated during the time of the experiment. Estimating localized activities requires the development of new algorithms based on the L1 norm.
4.2.4 Additional Remarks. We stress that our method is not only about EEG-fMRI fusion. It is also interesting to estimate the cortical activity when applied to EEG or fMRI data alone. In the first case, it performs inverse problem solving for EEG, which can enforce specific temporal and spatial smoothness of the sources (see also Baillet & Garnero, 1997; Daunizeau et al., 2006). In the second case, it performs a deconvolution of fMRI data which can enforce spatial smoothness of the sources, whereas existing methods (e.g., Glover, 1999) operate only on a pixel-by-pixel basis.

We have discussed above the oversimplicity of our electrical-metabolic coupling. The modeling of the EEG and fMRI does not pose such a specific problem. Note that the specific choices we made for the fMRI part could still be discussed and other models might fit better the physiology (Zheng et al., 2005), but this appears to be a minor issue, since the balloon model that we used can account for the BOLD signal (Buxton et al., 2004; Deneux & Faugeras, 2006). On the other hand, our modeling of the measurement noise might be simplistic, since we used temporally uncorrelated noise and thus ignored possible physiological noises (heart pulsation, breathing, eye blinking in the EEG) or physical autocorrelated noises (50 Hz). One could choose to remove these noises prior to the analysis, but it is also possible to account for noises with specific temporal autocorrelation in the model by adding the appropriate hidden states and autoregressive equations.

Finally, the choice of the distribution of sources on the cortical surface was quite arbitrary (see appendix A.1), and could be done according to prior anatomical knowledge such as subdivisions of Broca areas (Daunizeau et al., 2007). Also, if the sources are grouped according to such functional areas, one could add stronger continuity constraints inside such areas, which would greatly improve the problem of underdetermination of the system and at the same time allow more discontinuities between the areas.

5 Conclusion

We summarize here the novelties of the EEG-fMRI fusion method we propose and sketch some of the ways to improve it in order to be able to apply it efficiently on real data.

First, our method is based on a number of interesting and important concepts for combining the data from two modalities with very distinct characteristics in order to estimate a hidden cortical activity:

- It relies on physiological models, using dynamical system formalism.
- It estimates a whole spatiotemporal pattern, and both modalities provide some partial spatial and temporal information.
- We advocate the use of adaptive filters (i.e., estimate the cortical activity iteratively with respect to time and globally with respect to space).
The Bayesian formalism of the Kalman filter allows derivations of a posteriori distributions in a natural and symmetric way.

We have also proved that:

- When supposing a linear coupling between electrical and metabolic activities, the method can be applied on real-size data at a reasonable computational cost. We have shown simulation where the cortical surface was divided into 2000 patches of activity and electrical activity sampled at 200 Hz.

- It is possible to apply the same methodology in the case of nonlinear couplings at a greater computational cost. We have shown it on reduced examples with 10 and 100 sources.

Further improvements should focus on:

- The computational (and memory) cost of the algorithm. In particular, in order to apply our Kalman-based filter to nonlinear coupling, it is necessary to store the covariance information in a more efficient way than full matrices.

- The developments of alternative filtering techniques. It might be interesting to use more specialized nonlinear filtering methods from the literature. In particular, the extended Kalman filter estimations are systematically smooth in time and space. Because it is minimizing an L2 norm, alternative methods could be investigated in order to produce sharper estimates.

- The preprocessing of real data should be examined carefully. For example, one has to decide if it is better to first remove artifacts or to include them in the model.

Appendix

A.1 Mesh Downsampling. In our simulation, each source was defined as a small area on the cortical surface spanning a number of dipoles, all having the same activity (these dipoles themselves are placed on the vertices of an initial mesh with approximately 35,000 vertices). Thus, defining the sources consists of clustering the vertices of the initial mesh in a uniform manner, such that all sources will have approximately equal area measures. We do it in the following way:

1. Choose randomly the first vertex that we call a “seed.”
2. Compute the distance of all vertices to this seed (this is done by “propagating” the distance from one vertex to its neighbors, starting from the seed).
3. Choose a second seed as the vertex that is the most distant to the first seed.
4. Compute the distance of all vertices to its closest seed (done with a similar propagating routine).
5. Choose a new seed as the vertex that is the most distant to any seed.
6. Repeat steps 4 and 5 until the number of seeds reaches the desired number of sources. Then vertices are clustered according to their closest seed.

Figure S6 in the supplementary materials shows the 2000 patches defined on our initial 35,000 vertices mesh. The average distance between 2 sources is 6.4 mm.

Once this clustering is done—say, \( n_u \) patches have been defined out of \( N \) vertices—we can define an \( N \times n_u \) transformation matrix \( \mathcal{F} \) that interpolates the currents at the \( N \) dipole locations as a function of the \( n_u \) patches activities. This matrix \( \mathcal{F} \) serves to redefine several parameters used by the algorithm:

- The \( n_e \times n_u \) matrix \( B \) in equation 2.4 is defined from the \( n_e \times N \) EEG forward model matrix \( \mathcal{B} \) computed on the initial mesh as \( B = \mathcal{B} \mathcal{F} \).
- The \( n_u \times n_u \) connectivity matrix \( C \) used to compute the geodesic distances for equation 3.1 is defined from the \( N \times N \) connectivity matrix of the initial mesh \( \mathcal{C} \) as \( C = (\mathcal{F}^T \mathcal{C} \mathcal{F} > 0) \).
- The \( n_u \times n_u \) innovative noise correlation matrix \( Q_u \) is computed by first applying equation 2.8 to the original mesh, thus obtaining an \( N \times N \) correlation matrix \( \mathcal{Q} \) between all dipoles, and then computing \( Q_u = \mathcal{L} \mathcal{Q} \mathcal{L}^T \), where \( \mathcal{L} \) is the pseudo-inverse of \( \mathcal{F} \), \( \mathcal{L} = (\mathcal{F}^T \mathcal{F})^{-1} \mathcal{F}^T \).

A.2 Convergence of Variance Matrices. We mentioned in the discussion the computational cost of the convergence of the variance matrices \( \hat{P} \) and \( \check{P} \) that we use a specific heuristic rule to increase the speed of this convergence. This rule is:

- Start with \( \hat{P}^{(0)} = 0 \).
- The first step \( \hat{P}^{(0)} - > P^{(1)} - > \check{P}^{(1)} \) of the convergence is normal.
- After each next step \( \hat{P}^{(k)} - > \hat{P}^{(k+1)} \), we compute the correlation between the last two steps \( c = corr(\hat{P}^{(k+1)} - \hat{P}^{(k)}) \) and \( \hat{P}^{(k+1)} - \hat{P}^{(k)} \):
  - If this correlation is negative, nothing is done.
  - If it is positive, we replace \( \hat{P}^{(k+1)} \) by \( \hat{P}^{(k+1)} + K_c (\hat{P}^{(k+1)} - \hat{P}^{(k)}) \), where \( K \) is a fixed parameter.
- We use \( K = 2 \) for steps 2 to 10, \( K = 3 \) for steps 11 to 200, \( K = 6 \) for steps 201 to 295, and \( K = 0 \) for the last five steps.

A.3 Modification of the Extended Kalman Filter and Smoother. Based on derivations of the Kalman filter and smoother in Welling (n.d.) and on the linearization principle of the extended Kalman filter (Arulampalam & Maskell, 2002; Welch & Bishop, 2006), we designed a modified filter that first performs backward and then forward. More precisely, we first estimate
\( p(x_k | y_k, \ldots, y_n) \) for decreasing \( k \) and then \( p(x_k | y_1, \ldots, y_n) \) for increasing \( k \). This has the advantage that \( p(x_k | y_k, \ldots, y_n) \) is already more accurate than \( p(x_k | y_1, \ldots, y_k) \) calculated by the classic extended Kalman filter since it takes into account the information on the state provided by the delayed fMRI response.

We introduce new notation:

\[
\begin{align*}
\hat{x}_k^l &= E(x_k | y_l, \ldots, y_n) \\
\hat{P}_k^l &= V(x_k | y_l, \ldots, y_n).
\end{align*}
\]

First, it is important to point out that in the absence of any measurement, all states \( x_k \) have the same a priori distribution \( N(x^0, P^0) \), where \( P^0 \) can be obtained by repeating the Kalman filter time update step until convergence:

\[
P^{(0)} = 0, \quad P^{(k+1)} = AP^{(k)}A^T + Q, \quad P^0 = \lim_{k \to \infty} P^{(k)}
\]

(note that \( A \) is the derivative of the evolution equation 4.2a to 4.2c at the resting state: \( F(x) \approx F(x^0) + A(x - x_0) = x^0 + A(x - x_0) \); so this a priori distribution is only an approximation, all the more so since it does not take into consideration that we do not allow the state \( \chi \) to become negative. A more accurate way to compute this a priori distribution could improve the estimation results, but at least in our case, this did not represent a serious limitation).

We start with the a priori distribution of \( x_n \):

\[
\begin{align*}
\hat{x}_n^{n+1} &= x^0 \\
\hat{P}_n^{n+1} &= P^0.
\end{align*}
\]

Then starting with \( k = n \) and decreasing it iteratively, we repeat the “measurement update”:

\[
\begin{align*}
K &= \hat{P}_k^{k+1}D^T(D\hat{P}_k^{k+1}D^T + R)^{-1} \\
\hat{x}_k^k &= \hat{x}_k^{k+1} + K(y_k - G(\hat{x}_k^{k+1})) \\
\hat{P}_k^k &= (I - KD)\hat{P}_k^{k+1}
\end{align*}
\]

where \( D \) is the derivative of the (strongly nonlinear) measure function at \( \hat{x}_k^{k+1} \). \( G(x) \approx G(\hat{x}_k^{k+1}) + D(x - \hat{x}_k^{k+1}) \), and the “backward time update”:

\[
\begin{align*}
J &= P^0A^T(AP^0A^T + Q)^{-1} \\
\hat{x}_k^{k+1} &= x^0 + J(\hat{x}_k^{k+1} - x^0) \\
\hat{P}_k^{k+1} &= J\hat{P}_k^{k+1}J^T + (I - JA)P^0
\end{align*}
\]
where \( A \) is the derivative of the evolution function at \( \tilde{x}_{k+1} \): \( F(x) \approx F(\tilde{x}_{k+1}^t) + A(x - \tilde{x}_{k+1}^t) \). We derived this step first by proving, in a similar manner as for the measure update (Welling, n.d.), that \( p(x_k \mid x_{k+1}) = N(x^0 + J(x_{k+1} - x^0), (I - JA)P^0) \), and then by computing \( p(x_k \mid y_{k+1}, \ldots, y_n) = \int_{x_{k+1}} p(x_k \mid x_{k+1}) p(x_{k+1} \mid y_{k+1}, \ldots, y_n) dx_{k+1} \).

After reaching the first time instant, the second, smoothing, step starts with \( k = 1 \), and increases \( k \) iteratively:

\[
L = (Q^{-1} + (P_{k+1}^{-1})^{-1})^{-1} Q^{-1}
\]

\[
\tilde{x}_{k+1}^1 = \tilde{x}_{k+1}^t + L(F(\tilde{x}_k^t) - \tilde{x}_{k+1}^t)
\]

\[
\tilde{P}_{k+1}^1 = (Q^{-1} + (P_{k+1}^{-1})^{-1})^{-1} + LA\tilde{P}_{k}^1 (LA)^T.
\]

We derived this step first by proving, again in a similar manner as for the measure update, that \( p(x_{k+1} \mid x_k, y_1, \ldots, y_n) = N(\tilde{x}_{k+1}^t + L(F(\tilde{x}_k^t) - \tilde{x}_{k+1}^t), (Q^{-1} + (P_{k+1}^{-1})^{-1})^{-1} \), and second by computing \( p(x_{k+1} \mid y_1, \ldots, y_n) = \int_{x_{k+1}} p(x_{k+1} \mid x_k, y_1, \ldots, y_n) p(x_k \mid y_1, \ldots, y_n) dx_{k+1} \).

Since we observed some divergences when applying this second smoothing step due to accumulating errors in the estimation of the phase, we applied it only on the estimation of the envelope (and kept the result from the first step only for the phase estimate).

Acknowledgments

We thank Anne-Lise Paradis (LPPA Collège de France) for providing us with anatomic MRI data and Geoffray Adde (Odyssée Team, INRIA) for help with constructing the EEG forward problem matrix.

References


Received May 28, 2008; accepted August 4, 2009.