

Hierarchical Bayesian estimation for MEG inverse problem

Masa-aki Sato,^{a,b,*} Taku Yoshioka,^a Shigeki Kajihara,^c Keisuke Toyama,^c Naokazu Goda,^d
Kenji Doya,^{a,b} and Mitsuo Kawato^a

^aATR Computational Neuroscience Laboratories, Seika, Soraku, Kyoto 619-0288, Japan

^bCREST, Japan Science and Technology Corporation, Japan

^cTechnology Research Laboratory, Shimadzu Co., Seika, Soraku, Kyoto 619-0237, Japan

^dNational Institute for Physiological Sciences, Myodaiji, Okazaki 444-8585, Japan

Received 24 November 2003; revised 1 March 2004; accepted 22 June 2004

Source current estimation from MEG measurement is an ill-posed problem that requires prior assumptions about brain activity and an efficient estimation algorithm. In this article, we propose a new hierarchical Bayesian method introducing a hierarchical prior that can effectively incorporate both structural and functional MRI data. In our method, the variance of the source current at each source location is considered an unknown parameter and estimated from the observed MEG data and prior information by using the Variational Bayesian method. The fMRI information can be imposed as prior information on the variance distribution rather than the variance itself so that it gives a soft constraint on the variance. A spatial smoothness constraint, that the neural activity within a few millimeter radius tends to be similar due to the neural connections, can also be implemented as a hierarchical prior. The proposed method provides a unified theory to deal with the following three situations: (1) MEG with no other data, (2) MEG with structural MRI data on cortical surfaces, and (3) MEG with both structural MRI and fMRI data. We investigated the performance of our method and conventional linear inverse methods under these three conditions. Simulation results indicate that our method has better accuracy and spatial resolution than the conventional linear inverse methods under all three conditions. It is also shown that accuracy of our method improves as MRI and fMRI information becomes available. Simulation results demonstrate that our method appropriately resolves the inverse problem even if fMRI data convey inaccurate information, while the Wiener filter method is seriously deteriorated by inaccurate fMRI information.

© 2004 Elsevier Inc. All rights reserved.

Keywords: MEG/EEG inverse problem; Hierarchical Bayesian method; Variational Bayesian method; Variance estimation; MRI and fMRI information; Spatial smoothness constraint

Introduction

Functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) are the major recording means of brain activity. fMRI records brain activity with millimeter-order spatial resolution, but its temporal resolution is on the order of several seconds due to slow hemodynamic responses to neural activity (Bandettini, 2000; Belliveau et al., 1991; Churchland and Sejnowski, 1988; Logothetis et al., 2001; Ogawa et al., 1990). Conversely, MEG measures brain activity with millisecond-order temporal resolution, but its spatial resolution is poor due to the ill-posed nature of the inverse problem for estimating source currents from the electromagnetic measurement (Nunez, 1981). In addition, the number of MEG sensors is generally insufficient to provide a precise reconstruction of the source current (Grave de Peralta Menendez and Gonzalez Andino, 1998). Therefore, prior information on the source currents is essential to solve the inverse problem. Inverse procedures are commonly classified as dipole and distributed source methods. The dipole method (Aine et al., 2000; Hari, 1991; Mosher et al., 1992) approximates brain activity by a small number of current dipoles. Although this method gives good estimate when the number of active areas is small, it is difficult to determine the appropriate number of dipole sources for complicated spatio-temporal activity. In addition, neural current distribution over the cortical surface cannot be estimated by the dipole method. The distributed source method assumes distributed currents in the brain (Hamalainen et al., 1993). In a linear approach to resolve the inverse problem, several prior assumptions have been used such as the (weighted) minimum norm method (Hamalainen and Ilmoniemi, 1994; Hamalainen et al., 1993; Wang et al., 1992), the maximum smoothness method (Pascual-Marqui, 1999; Pascual-Marqui et al., 1994), the minimum L1-norm method (Uutela et al., 1999), and others (Huang et al., 1997; Toyama et al., 1999). Bayesian methods have also been proposed to incorporate nonlinear smoothness constraints (Baillet and Garnero, 1997). Unfortunately, the prior assumptions are still insufficient to fully resolve

* Corresponding author. ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika, Soraku, Kyoto 619-0288, Japan. Fax: +81 774 95 1259.

E-mail address: masa-aki@atr.jp (M. Sato).

Available online on ScienceDirect (www.sciencedirect.com.)

the ill-posed nature of the inverse problem, and thus the spatial resolution of these methods is still rather low. Attempts have been made to overcome these limitations by supplementing the information of other imaging means (Ahlfors et al., 1999; Dale and Sereno, 1993; Dale et al., 2000; Fujimaki et al., 2002; Heinze et al., 1994; Kajihara et al., 2004; Liu et al., 1998; Phillips et al., 2002; Schmidt et al., 1999). The MRI image gives information on the position and orientation of the cortical dipoles, while fMRI provides topographical information on active dipoles. Although fMRI has high spatial resolution, it has been pointed out that the hemodynamic signals of fMRI may not precisely correspond to neural activity due to various factors such as the effects of noise and artifacts. In particular, the fMRI activity corresponds to an average of several thousands of MEG time series data because of the time resolution difference between MEG and fMRI. When higher brain functions are examined, several different processes may occur within several seconds, and these activities are averaged out in the fMRI data. Consequently, the fMRI active areas may include inactive current areas in the MEG time slice data, and the fMRI activity for the active current may be smeared by the temporal averaging. The recent approaches, such as the Wiener filter or the Bayesian method (Dale et al., 2000; Kajihara et al., 2004; Phillips et al., 2002; Schmidt et al., 1999), use the fMRI data as prior information on the source current variance by assuming that the current variances for fMRI active dipoles are large compared with those for fMRI inactive dipoles. Source current estimation based on these methods may fail if the fMRI data contain incorrect information.

In this article, we propose a new hierarchical Bayesian method introducing a hierarchical prior that can effectively incorporate both structural and functional MRI data. In our method, the variance of the source current at each source location is considered an unknown parameter and estimated from the observed MEG data and prior information. The fMRI information can be imposed as prior information on the variance distribution rather than the variance itself so that it gives a soft constraint on the variance. Therefore, our method is capable of appropriately estimating the source current variance from the MEG data supplemented with the fMRI data, even if fMRI data convey inaccurate information. Accordingly, our method is robust against inaccurate fMRI information. The spatial smoothness constraint that the neural activity within a few millimeter radius tends to be similar due to the neural connections can also be implemented as a hierarchical prior. Because of the hierarchical prior, the estimation problem becomes nonlinear and cannot be solved analytically. Therefore, the approximate posterior distribution is calculated by using the Variational Bayesian (VB) method (Attias, 1999; Sato, 2001). The resulting algorithm is an iterative procedure that converges quickly because the VB algorithm is a type of natural gradient method (Amari, 1998) that has an optimal local convergence property. The VB method also provides a model selection criterion and is used to select the most probable surface on which the source current is located when there is no prior information on the source position.

This paper presents principles and performance of the new hierarchical Bayesian method in comparison with the conventional linear inverse methods for the three different situations: (1) MEG with no other data, (2) MEG with structural MRI data on cortical surfaces, and (3) MEG with both structural MRI and fMRI data. We also examined the performance of our method and the Wiener filter method for incorrect fMRI information.

Methods

Bayesian and linear filter approaches for MEG inverse problem

When neural current activity occurs in the brain, it produces a magnetic field observed by MEG. The relationship between the magnetic field $\mathbf{B} = \{B_m | m = 1:M\}^1$ measured by M sensors and the primary source current $\mathbf{J} = \{J_n | n = 1:N\}$ in the brain is given by

$$\mathbf{B} = \mathbf{G} \cdot \mathbf{J}, \quad (1)$$

where $\mathbf{G} = \{G_{m,n} | m = 1:M, n = 1:N\}$ is the lead field matrix. The lead field $G_{m,n}$ represents the magnetic field B_m produced by the n th unit dipole current. The MEG inverse problem is to estimate the source current \mathbf{J} from the observed magnetic field data \mathbf{B} .

Conventional linear inverse filters can be derived by minimizing an error function defined by

$$E(\mathbf{J}) = \beta \|\mathbf{B} - \mathbf{G} \cdot \mathbf{J}\|^2 + \mathbf{J}' \cdot \boldsymbol{\Sigma}_0 \cdot \mathbf{J}. \quad (2)$$

The first term in the right hand side (r.h.s) of Eq. (2) represents the reconstruction error, which measures the difference between the observed MEG data and the reconstructed MEG signal calculated from the estimated source current. To resolve the ill-posed nature of the MEG inverse problem, a regularization term is added to the error function. Depending on the choice of the regularization matrix $\boldsymbol{\Sigma}_0$, various linear inverse methods can be derived. For example, the minimum norm solution (Hamalainen and Ilmoniemi, 1994) can be derived if $\boldsymbol{\Sigma}_0$ is the N -by- N identity matrix. The parameter β controls the relative weight between the reconstruction error and the regularization term. The minimum solution of Eq. (2) is given by

$$\mathbf{J} = \mathbf{L} \cdot \mathbf{B} = \boldsymbol{\Sigma}_0^{-1} \cdot \mathbf{G}' \cdot (\mathbf{G} \cdot \boldsymbol{\Sigma}_0^{-1} \cdot \mathbf{G}' + \beta^{-1} \mathbf{I}_M)^{-1} \cdot \mathbf{B}, \quad (3)$$

where \mathbf{I}_M denotes the M -by- M identity matrix.

These linear inverse filters can be also derived by using the Bayesian method (Dale et al., 2000; Grave de Peralta Menendez and Gonzalez Andino, 1998; Phillips et al., 2002). The probabilistic model for the source currents can be constructed assuming Gaussian noise for the MEG sensors. Then, the probability distribution, that the magnetic field \mathbf{B} is observed for a given current \mathbf{J} , is given by

$$P(\mathbf{B}|\mathbf{J}) \propto \exp\left(-\frac{1}{2}\beta\|\mathbf{B} - \mathbf{G} \cdot \mathbf{J}\|^2\right), \quad (4)$$

where β corresponds to the inverse of the noise variance. In the Bayesian method (Gelman et al., 1995), a prior assumption on the source currents \mathbf{J} is imposed as a prior probability distribution $P_0(\mathbf{J})$. The Bayesian method calculates the posterior probability distribution for the source current, $P(\mathbf{J}|\mathbf{B})$, which represents the probability that the source current is \mathbf{J} under the observed MEG data \mathbf{B} and the prior information:

$$P(\mathbf{J}|\mathbf{B}) = \frac{P(\mathbf{B}|\mathbf{J})P_0(\mathbf{J})}{P(\mathbf{B})}. \quad (5)$$

¹ In the following, the notation, $m = 1:M$, is used to represent $m = 1, \dots, M$.

The marginal likelihood $P(\mathbf{B})$ appearing in the denominator is defined by

$$P(\mathbf{B}) = \int d\mathbf{J} P(\mathbf{B}|\mathbf{J}) P_0(\mathbf{J}). \quad (6)$$

Maximum a posteriori (MAP) estimates find the optimal parameter (source current value) that maximizes the posterior probability distribution $P(\mathbf{J}|\mathbf{B})$. The linear inverse filters defined by Eq. (3) can be derived as the MAP solution by assuming the Normal distribution prior for the source current with a fixed covariance matrix Σ_0^{-1} ,

$$P_0(\mathbf{J}) \propto \exp\left[-\frac{1}{2} \mathbf{J}' \cdot \Sigma_0 \cdot \mathbf{J}\right]. \quad (7)$$

One can easily see that the maximization of the log-posterior $\log(P(\mathbf{J}|\mathbf{B}))$ is equivalent to the minimization of the error function (2).

The Wiener filter solution is derived by the minimization of the estimated current error, which is defined as the error between the true current and the estimated current. The MAP solution (3) gives the Wiener filter solution, if the prior current covariance matrix Σ_0^{-1} is the true current covariance matrix (Dale et al., 2000). However, the true current covariance is unknown. It has been proposed that the fMRI information can be used to determine the prior current covariance matrix (Dale et al., 2000; Phillips et al., 2002).

Because of the underdetermined character of the MEG inverse problem, that is, $N > M$, a class of the linear inverse methods defined by Eq. (3) satisfies the following constraint (see Appendix A):

$$\sum_{n=1}^N \mathcal{G}_n \leq M. \quad (8)$$

The estimation gain \mathcal{G}_n represents the n th estimated current J_n by Eq. (3) for a single unit dipole source at the n th position and should be one for perfect reconstruction. The inequality (8) implies that the linear inverse method cannot perfectly retrieve more current dipoles than the number of sensors M .

This constraint can be avoided by using the variance information. To simplify the explanation, let us assume that there are active currents in $n = 1:K$ and the current covariance matrix Σ_0^{-1} is given as a diagonal matrix with $(\Sigma_0^{-1})_{n,n} = 1$ (for $n = 1:K$) and $(\Sigma_0^{-1})_{n,n} = 0$ (for $n = (K+1):N$). The estimated currents for null variance points J_n ($n = (K+1):N$) vanish from Eq. (3) with this covariance matrix. Therefore, the estimation gains for $n = (K+1):N$ also vanish, $\mathcal{G}_n = 0$, and the constraint (8) becomes $\sum_{n=1}^K \mathcal{G}_n \leq M$. Consequently, the inequality does not constrain the estimation gains for active dipoles, \mathcal{G}_n ($n = 1:K$), if the number of active dipoles K is less than the number of sensors M . Introduction of activity dependent variance information reduces the effective degree of freedom by suppressing the estimation gains for small variance points, and this makes the estimation of active dipoles more accurate.

However, the source current estimation based on the prior current variance such as the Wiener filter method may fail if the variance information is incorrect.

Hierarchical Bayesian method

We propose a new hierarchical Bayesian method to overcome the above difficulties. In our method, the variances (diagonal

part of the covariance matrix) are considered unknown parameters and estimated from the observed MEG data by introducing a hierarchical prior on the current variance. The fMRI information can be imposed as prior information on the variance distribution rather than the variance itself so that it gives a soft constraint on the variance. The spatial smoothness constraint that neurons within a few millimeter radius tend to fire simultaneously due to the neural interactions can also be implemented as a hierarchical prior.

Hierarchical prior

Let us suppose a time sequence of MEG data $\mathbf{B}_{1:T} \equiv \{\mathbf{B}(t)|t = 1:T\}$ is observed. The MEG inverse problem in this case is to estimate the primary source current $\mathbf{J}_{1:T} \equiv \{\mathbf{J}(t)|t = 1:T\}$ from the observed MEG data $\mathbf{B}_{1:T}$. We assume a Normal prior for the current:

$$P_0(\mathbf{J}_{1:T}|\alpha) \propto \exp\left[-\frac{1}{2} \sum_{t=1}^T \mathbf{J}'(t) \cdot \mathbf{A} \cdot \mathbf{J}(t)\right], \quad (9)$$

where \mathbf{A} is the diagonal matrix with diagonal elements $\alpha = \{\alpha_n|n = 1:N\}$. We also assume that the current variance α^{-1} does not change over period T . The current inverse variance parameter α is estimated by introducing an Automatic Relevance Determination (ARD) hierarchical prior (Neal, 1996):

$$P_0(\alpha) = \prod_{n=1}^N \Gamma(\alpha_n|\bar{\alpha}_{0n}, \gamma_{0nz}),$$

$$\Gamma(\alpha|\bar{\alpha}, \gamma) \equiv \alpha^{-1} \left(\alpha\gamma/\bar{\alpha}\right)^\gamma \Gamma(\gamma)^{-1} e^{-\alpha\gamma/\bar{\alpha}}, \quad (10)$$

where $\Gamma(\alpha|\bar{\alpha}, \gamma)$ represents the Gamma distribution with mean $\bar{\alpha}$ and degree of freedom γ . $\Gamma(\gamma) \equiv \int_0^\infty dt t^{\gamma-1} e^{-t}$ is the Gamma function.

We briefly explain the difference between a conventional Normal prior and the hierarchical prior. In the case of the Normal prior, the value of the inverse variance parameter α_n is given as prior information. For small variance (large α_n value, α_L), the prior probability distribution for the current, $P_0(J_n|\alpha_L)$, is concentrated near zero (Fig. 1A), so large current is highly penalized by the prior information. For large variance (small α_n value, α_S), $P_0(J_n|\alpha_S)$ spreads over large values (Fig. 1B), so the estimated current likely becomes large. In the case of the hierarchical prior, the inverse variance parameter α_n is considered a random variable. Therefore, there are nonzero probabilities for any value of the inverse variance parameter α_n , and the prior probability distribution for α_n is given by the Gamma distribution (10). The hyperparameters γ_{0nz} control the spread of the distribution and represent the reliability (or confidence) of the hierarchical prior. The hyperparameters $\bar{\alpha}_{0n}$ represent the mean value of α_n in the hierarchical prior. If the prior mean α_{0n} is small, the prior probability $P_0(\alpha_n = \alpha_S)$ that α_n takes a small value, α_S (large variance), increases (Fig. 1C). In contrast, the prior probability $P_0(\alpha_n = \alpha_L)$ that α_n takes a large value, α_L (small variance), increases for large $\bar{\alpha}_{0n}$ (Fig. 1D). Therefore, the prior information on the current variance is imposed as a soft constraint in the variance estimation.

When the fMRI data are not available, we have no information on the current variance. Lack of knowledge can be represented as a non-informative prior, which assumes that any parameter value has

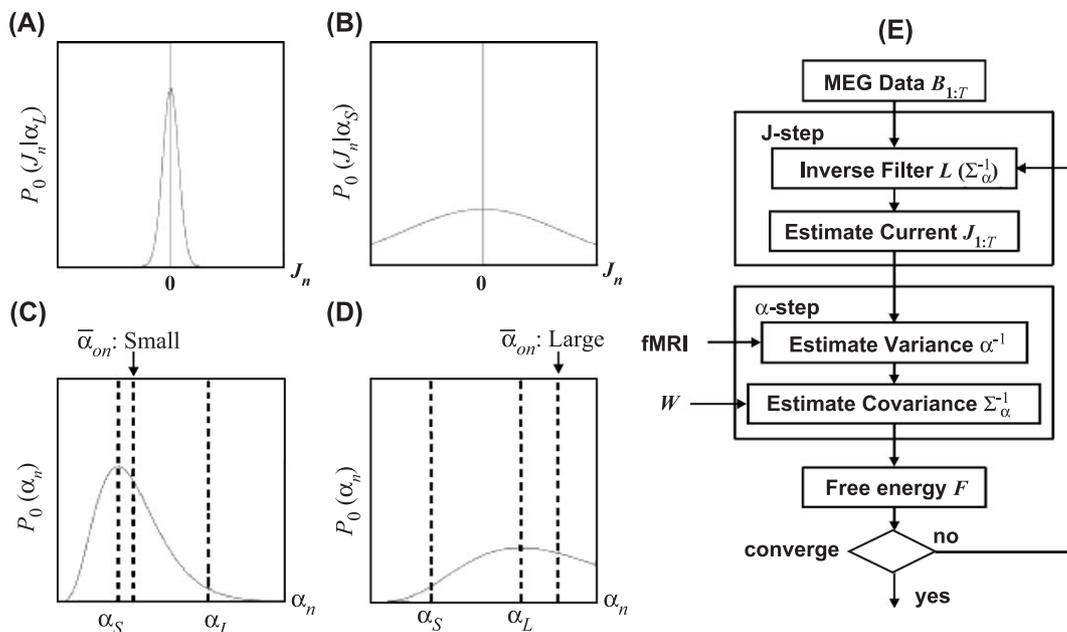


Fig. 1. Hierarchical Bayesian method. (A) Probability distribution of a Normal prior $P_0(J_n|\alpha_L)$ for the current J_n with a large inverse variance (small variance) parameter α_L is plotted. (B) That with a small inverse variance (large variance) parameter α_S . (C) Hierarchical prior distribution $P_0(\alpha_n)$ for the inverse variance parameter α_n with a small prior mean $\bar{\alpha}_{0n}$ is plotted. The probability for small inverse variance α_S is large compared with that for large inverse variance α_L . (D) That with a large prior mean $\bar{\alpha}_{0n}$. The probability for small inverse variance α_S is small compared with that for large inverse variance α_L . (E) Flow chart of the Variational Bayesian (VB) method. In the J-step, the inverse filter $L(\Sigma_\alpha^{-1})$ is recalculated according to (28) using the estimated covariance matrix Σ_α^{-1} in the previous iteration. The current $J_{1:T}$ is then estimated from the observed MEG data $B_{1:T}$ using the inverse filter according to Eq. (29). The current variance α^{-1} is estimated in the α -step by using the estimated time sequence of the current $J_{1:T}$ and the fMRI information according to Eq. (30). The current covariance matrix Σ_α^{-1} is then calculated by using the current variance and the smoothing filter W according to Eq. (31). The free energy is calculated at the end of the steps. If the free energy has not converged, the above J- and α -steps are repeated.

the same probability.² Therefore, we use a non-informative prior for the current inverse variance parameter α_n in this case, that is, $\gamma_{0n\alpha} = 0$ and $P_0(\alpha_n) = \alpha_n^{-1}$. When the fMRI data is available, fMRI information is imposed as the prior mean variance $\bar{\alpha}_{0n}^{-1}$ weighted with the confidence parameter $\gamma_{0n\alpha}$. The prior mean variance $\bar{\alpha}_{0n}^{-1}$ is increased for the fMRI active area, so the probability that α_n takes a small value (large variance) becomes large (Fig. 1C). On the other hand, $\bar{\alpha}_{0n}^{-1}$ is decreased for the fMRI inactive area, so the probability that α_n takes a large value (small variance) becomes large (Fig. 1D).

Smoothness prior

The smoothness constraint can be imposed as the structural constraint on the off-diagonal part of the covariance matrix, which enforces high correlations between neighboring current activity:

$$P_0(\mathbf{J}_{1:T}|\alpha, \lambda) \propto \exp\left[-\frac{1}{2} \sum_{t=1}^T \mathbf{J}'(t) \cdot \Sigma_\alpha \cdot \mathbf{J}(t)\right], \quad (11)$$

where the current covariance matrix Σ_α^{-1} is given by

$$\Sigma_\alpha^{-1} = \mathbf{A}^{-1} + \mathbf{W} \cdot \mathbf{A}^{-1} \cdot \mathbf{W}'. \quad (12)$$

\mathbf{A} and \mathbf{A} are the diagonal matrices with the diagonal elements $\alpha = \{\alpha_n | n = 1:N\}$ and $\lambda = \{\lambda_n | n = 1:N\}$, respectively. The second term in the r.h.s. of Eq. (12) represents the correlation in current activity between the neighboring points. The spatial profile of the

correlation function is characterized by the smoothing filter W . If we assume the Gaussian smoothing filter for W , the correlation in current activity between two points decreases exponentially as the squared distance increases. Parameters α and λ control the diagonal and the off-diagonal part of the current covariance matrix, respectively. As the variance parameter α_n^{-1} increases, the variance of the n th current increases. As the smoothness parameter λ_n^{-1} increases, the correlation between the n th and the neighboring currents increases. Parameters α and λ are estimated by introducing the ARD hierarchical prior for them.

Variational Bayesian method

Because of the hierarchical prior, the estimation problem becomes nonlinear and cannot be solved analytically. Therefore, the approximate posterior distribution is calculated by using the Variational Bayesian (VB) method (Attias, 1999; Sato, 2001). In the VB method (see Appendix B.1), the calculation of the posterior distribution is reformulated as the maximization problem of the free energy for a trial distribution that approximates the true posterior. The optimization problem is solved by imposing constraints on the trial distribution. The resulting algorithm is an iterative procedure. The method alternately estimates the current and the variance from the observed MEG data $B_{1:T}$ and the prior variance information given by the fMRI data. It has been proved (Sato, 2001) that the VB method converges into a local maximum of the free energy and is equivalent to the natural gradient method (Amari, 1998), which has an optimal local convergence property. The maximum free energy provides a model selection criterion and is used to select the most probable surface on which the source

² We assume the probability distribution is flat with respect to $\log(\alpha_n)$.

current is located when there is no prior information on the source position (see Surface model selection). The detailed algorithm is derived in Appendix B.

Here, we explain some of the essential equations of the VB method for the hierarchical prior Eqs. (9) and (10). The algorithmic flow of the VB method is summarized in Fig. 1E (see also Appendix B.4). At the start of the algorithm, the initial value of the estimated variance α_n^{-1} is set to the prior mean variance $\bar{\alpha}_{0n}^{-1}$ and the covariance matrix is calculated as $\Sigma_\alpha^{-1} = \text{diag}(\alpha^{-1})$. Namely, the prior variance information given by the fMRI data is also used as the initial value of the current variance. Consequently, the VB current estimate at the first iteration is equal to the Wiener filter solution. When there is no fMRI information, all of the current variances are set to the same value, and thus the first VB current estimate is given by the minimum norm solution in this case.

In the J-step (Fig. 1E), the inverse filter $L(\Sigma_\alpha^{-1})$ is recalculated using the estimated covariance matrix Σ_α^{-1} . It depends on the observed data $\mathbf{B}_{1:T}$ through the estimated variance α^{-1} . The current is then estimated using the inverse filter (see Eqs. (28) and (29) in Appendix B.4):

$$\mathbf{L}(\Sigma_\alpha^{-1}) = \Sigma_\alpha^{-1} \cdot \mathbf{G}' \cdot \left(\mathbf{G} \cdot \Sigma_\alpha^{-1} \cdot \mathbf{G}' + \beta^{-1} \mathbf{I}_M \right)^{-1},$$

$$\mathbf{J}(t) = \mathbf{L}(\Sigma_\alpha^{-1}) \cdot \mathbf{B}(t). \quad (13)$$

This equation has the same form as the linear inverse Eq. (3). The crucial difference is that the variance of the current is estimated in our method, while it is given as prior information in the linear inverse method.

The variance is estimated in the α -step by using the estimated time sequence of the current $\mathbf{J}_{1:T}$ (Fig. 1E). The update equation for the current variance α_n^{-1} is somewhat complicated (see Eq. (30) in Appendix B.4), but it becomes simple when the algorithm converges:

$$\alpha_n^{-1} = \frac{\gamma_{0nz} \bar{\alpha}_{0n}^{-1} + \frac{T}{2} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{J}_n^2(t) \right)}{\gamma_{0nz} + \frac{T}{2} \left(\mathbf{L}(\Sigma_\alpha^{-1}) \cdot \mathbf{G} \right)_{n,n}}. \quad (14)$$

The first term in the r.h.s. numerator of Eq. (14) represents the contribution of the prior mean variance $\bar{\alpha}_{0n}^{-1}$. It is weighted by a prior confidence factor γ_{0nz} . The second term is the data term and represents the average magnitude of the estimated current. It is weighted by half of the data number $T/2$. Therefore, the estimated variance is a weighted average of the data term and the prior contribution. The covariance matrix is then calculated as $\Sigma_\alpha^{-1} = \text{diag}(\alpha^{-1})$. When the smoothness constraint is imposed by Eq. (11), the covariance matrix is calculated by Eq. (12) using the smoothing filter \mathbf{W} and the estimated smoothness parameter λ^{-1} as well as α^{-1} (see Eq. (31) in Appendix B.4). The free energy is calculated at the end of the steps. If the free energy is not converged, the above J- and α -steps are repeated.

Since the VB method estimates the current variance according to the observed MEG data as well as the fMRI information, the VB method can appropriately estimate the current variance value, even when the fMRI data is biased toward an incorrect variance value. We evaluated this point by using anatomically realistic simulation data, as described in the following sections.

Surface model selection

When structural MRI data is available, the cortical surface obtained from the MRI image data can be used as a surface model \mathcal{M} for current estimates and a dipole source current J_n is assumed in each vertex point on the surface \mathcal{M} . If there is no structural information by MRI, the shape and position of the cortical surface is unknown. In this case, we consider a set of surface models \mathcal{M} with different radii and compare the probability that the source current is on the surface \mathcal{M} . Since there is no prior information on the surface position, any surface may have the same prior probability, that is, $P_0(\mathcal{M}) = \text{constant}$. Then, the probability $P(\mathcal{M}|\mathbf{B}_{1:T})$, that the source current is on the surface \mathcal{M} under the observed data $\mathbf{B}_{1:T}$, is proportional to the marginal likelihood for the surface model \mathcal{M} , $P(\mathbf{B}_{1:T}|\mathcal{M})$:³

$$P(\mathcal{M}|\mathbf{B}_{1:T}) \propto P(\mathbf{B}_{1:T}|\mathcal{M})P_0(\mathcal{M}) \propto P(\mathbf{B}_{1:T}|\mathcal{M}). \quad (15)$$

Namely, we can find the most probable surface \mathcal{M} by finding the surface with the maximum marginal likelihood, which is equivalent to finding the maximum log-marginal likelihood $\log(P(\mathbf{B}_{1:T}|\mathcal{M}))$ because the logarithm is a monotonically increasing function.

The log-marginal likelihood is approximated by the maximum free energy for the surface model \mathcal{M} in the VB method (see Appendix B.1). Therefore, the most probable surface can be determined by finding the surface model with the maximum free energy. This procedure gives more accurate estimation of the source current than using the distributed current model in three-dimensional space.

Results

We evaluated the performance of our VB method to resolve the MEG inverse problem in comparison with conventional linear inverse filters. Basic estimation capabilities were evaluated by conducting simulations for a single dipole source and a single distributed source randomly placed on a cortical surface under three situations: (1) MEG with no other data, (2) MEG with structural MRI data on cortical surfaces, and (3) MEG with both structural MRI and fMRI data. Precision of the source localization was measured by a localization error, which was defined as the Euclidean distance between the true current and the estimated current peak. Accuracy of the current amplitude estimation for a single dipole source was measured by the estimation gain, which was defined as the ratio between the estimated current and the source current at the source position. The spatial profile of the current estimates for a single dipole source was also measured by a point-spread curve, which was defined as the estimated current profile around the source dipole. These measures were averaged over 500 cases randomly generated on the cortical surface.

More realistic estimation performances were also evaluated using simulated visual cortical responses under three other situations: (1) MEG with structural MRI data, (2) MEG with structural MRI and correct fMRI data, and (3) MEG with structural MRI and incorrect fMRI data. To evaluate spatio-temporal patterns

³ Here, the dependence on the surface model \mathcal{M} is explicitly expressed.

of the estimated currents, local currents were defined as the sum of the estimated currents within the areas corresponding to the active source currents. The precision of the estimation was assessed as the estimation gain at the source current peak and the normalized root mean squared error (RMSE) for the temporal sequences of the local current.

MRI-based cortical surface model

A polygon model of cortical surfaces (number of vertex points, 24,917; intervertex distance, 1.5 ± 0.5 mm) was constructed based on MRI image slices of an experimental subject (YO) using BrainVoyager 2000 (Goebel and Max Plank Society, 2000) (Fig. 2A). A single current dipole was assumed at each vertex point perpendicular to the cortical surface. One hundred and ninety-nine gradiometer sensors were assumed to surround the brain (blue dots in Fig. 2A), which reproduced the geometry of the whole-head MEG recording device used in our laboratory (Shimadzu SBI200). For calculation of the lead field, the brain structures were approximated as a sphere with a 100-mm radius, and the Sarvas equation (Sarvas, 1987) was used. In the following simulations, we used this MRI-based cortical surface model as a forward model to generate simulated MEG signals.

VB and linear inverse filter estimates for single dipole and distributed current sources

To investigate the basic capabilities of the VB estimate, we conducted simulations for two types of sources, that is, a single dipole source and a single circular distributed source composed of uniformly distributed dipoles within a 7-mm radius from the source center on the cortical surface. Noiseless and noisy conditions, where MEG signals contain Gaussian sensor noise with N/S of 0.1, were tested under three situations: (1) MEG with no other data, (2) MEG with structural MRI data, and (3) MEG with both structural MRI and fMRI data. Precision of the source localization was measured by the localization error, which was defined as the Euclidean distance between the true current (center) position X_0 and the maximum (absolute value) peak position X_{peak} of the estimated current:

$$\text{Localization Error} = \|X_0 - X_{\text{peak}}\|.$$

A single dipole or a single distributed source was located at one of 500 randomly selected vertices among 24,917 vertices (red points in Figs. 2B and C), and the localization errors for the 500 cases were averaged out, since the estimation performance depends on the depth and orientation of the current sources. The accuracy of

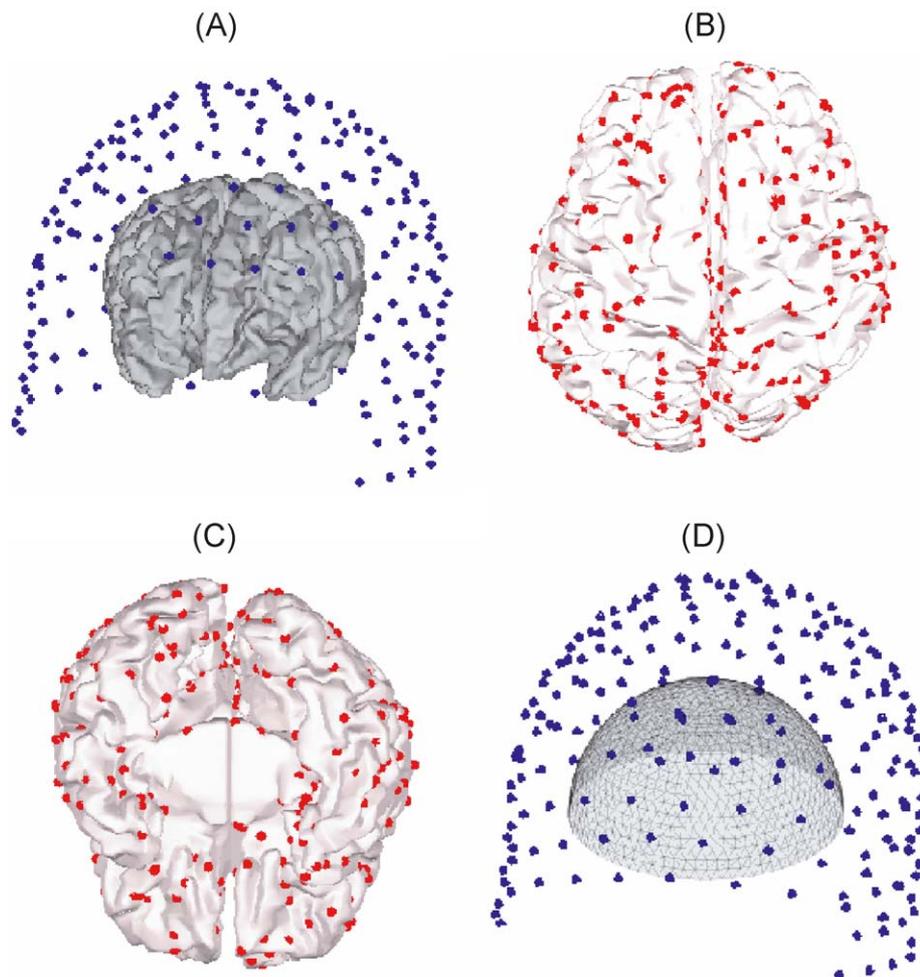


Fig. 2. MRI-based cortical surface model. (A) MRI-based cortical surface model obtained from MRI image data. 199 MEG sensors are also plotted as blue dots. (B) Five hundred points randomly selected from the MRI-based cortical surface model are shown by red dots (Top view). (C) Bottom view. (D) Variable-radius surface model for current estimates is plotted together with 199 MEG sensors.

the amplitude estimation and the spatial profile of the current estimates for a single dipole source were also assessed by averages of the estimation gain and the point-spread curve, respectively. The results were compared with those for the conventional linear inverse filters.

Before going into the detailed results, the estimation procedures of the VB method are explained for the three situations.

MEG with no other information

Variable-radius surface model

In the situation where no other information is available than MEG data, we assumed a variable-radius surface model for current estimates. The surface consisting of 2282 polygons (number of vertex points, 1185; intervertex distance, 5.7 mm at 70 mm radius) with a variable radius R was constructed, and a tangential current dipole with two degrees of freedom was assumed at each vertex point (Fig. 2D). The surface radius was determined so as to maximize the free energy (see Surface Model Selection). We used the hierarchical prior without smoothness constraints Eq. (9) and assumed the non-informative prior for $P_0(\alpha)$.

Single current source

We first explain the VB method for a single sinusoidal current dipole at $X_1 = (\text{radius } r_1 = 74.5 \text{ mm, elevation angle } \theta_1 = 69.4^\circ, \text{ azimuthal angle } \phi_1 = 110.2^\circ)$ of the MRI-based cortical surface model (blue dot in Fig. 3A) under the noisy condition with Gaussian sensor noise ($N/S = 0.1$).

The VB method was applied for each surface with a radius R by repeating the current (J-step) and the variance (α -step) estimations alternately according to Appendix B.4 until the free energy converged to the maximum (Fig. 1E). The free energy obtained for variable R peaked at $R = 75 \text{ mm}$ (Fig. 3C), which roughly agreed with the radius of the current source (74.5 mm). To see the effect of noise, the free energy curve under noiseless condition was also shown in a blue curve of Fig. 3B. The free energy was sharply tuned at the peak position when there was no noise. As the noise amplitude increased, the peak profile of the free energy curve became broader as shown in Fig. 3C. However, the peak almost stayed at the same position and there was no local peak around the maximum peak position in the presence of noise. Therefore, the source depth was safely determined by the free energy maximum. On the other hand, the reconstruction error (green curve in Fig. 3B)

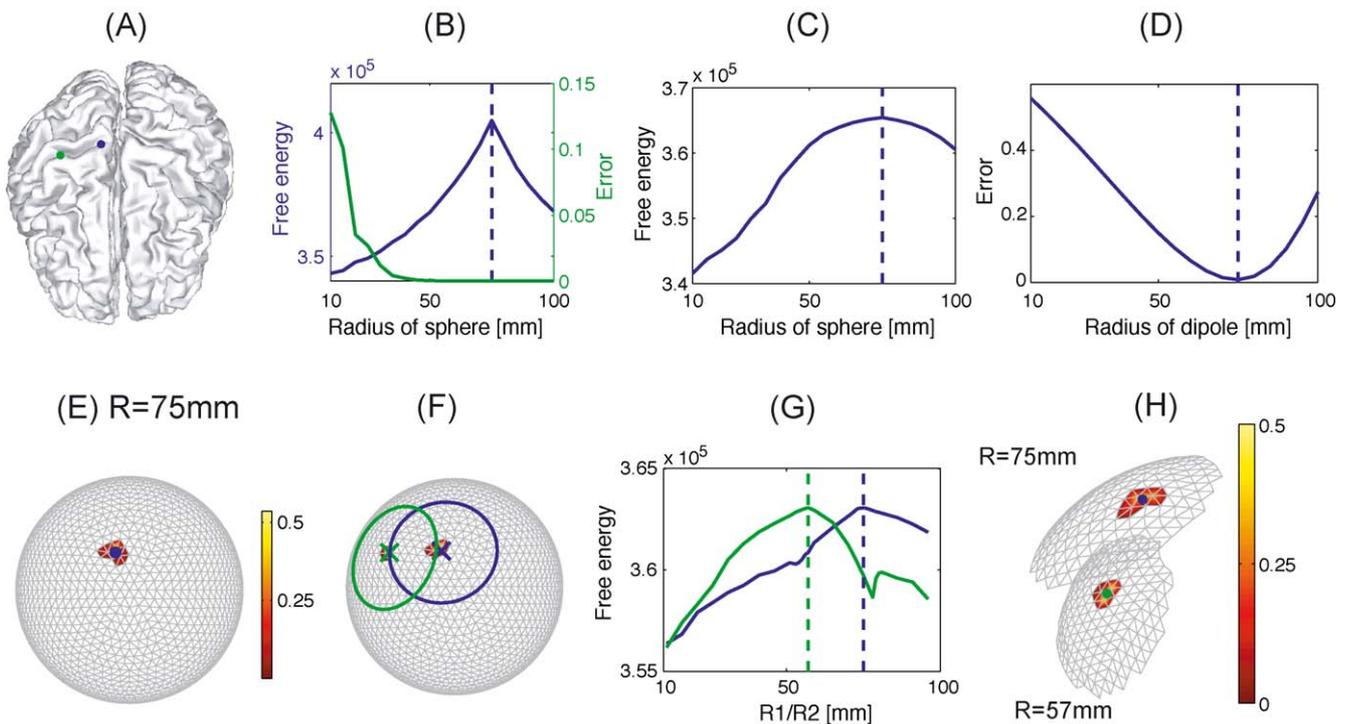


Fig. 3. Single and multiple source estimation. (A) A single dipole source at $X_1 = (74.5 \text{ mm}, 69.4^\circ, 110.2^\circ)$ in the MRI-based cortical surface model is shown by a blue dot. The other dipole source at $X_2 = (57.1 \text{ mm}, 47.2^\circ, 152.3^\circ)$ used in multiple source estimation is also shown by a green dot. (B) Free energy curve for a single source estimation under noiseless condition is plotted by a blue line. The free energy was maximized for each surface with radius R . The free energy peaked at 75 mm, which well agreed with the depth of the current source 74.5 mm. The reconstruction error under noiseless condition is also plotted by a green line. (C) Free energy curve for a single source estimation under noise condition. The free energy maximized for each surface with radius R is plotted. The free energy peaked at 75 mm. (D) The reconstruction error of the dipole fit under noise condition was calculated for each radius. A single dipole was fitted to the observed MEG data for each radius R by constraining the radius of the dipole to R . The error had the minimum at 75 mm. (E) Peak sinusoidal current estimated by VB method under noise condition for the surface with $R = 75 \text{ mm}$ is plotted on a flattened map using pseudocolor representation of current intensity. The source position is depicted by a blue dot. (F) Estimated currents on the maximum free energy surface for two dipole sources are plotted on the flattened map. Two local current peaks were identified near the angular positions of the two sources and depicted by blue and green crosses. Two local surfaces around the local current peaks are shown by blue and green circles. (G) The free energies maximized for a set of two local surfaces with radii R_1 and R_2 are plotted. The free energy as a function of R_1 (R_2) is plotted by a blue (green) line, where R_2 (R_1) is fixed at 57 mm (70 mm). (H) Current estimates for a set of two local surfaces with $R_1 = 75 \text{ mm}$ and $R_2 = 57 \text{ mm}$ are shown by pseudocolor representation. The two source positions are depicted by blue and green dots.

of this distributed source model under the noiseless condition was nearly flat for the radius larger than 50 mm, so it was not a good measure to estimate the source depth. This is because the magnetic field produced by a deep focal source can be reproduced by distributed superficial sources. We also examined the standard single dipole fit for this problem. The reconstruction error (Fig. 3D) of the dipole model under the same noise condition (N/S = 0.1) showed the similar (reversed) tuning curve as the free energy (Fig. 3C).

We then estimated the source current for the maximum free energy surface. The sinusoidal current peak estimated by the VB method for the surface with $R = 75$ mm was correctly focused around the true dipole position (blue dot in Fig. 3E) with a localization error of 2.7 mm.

Multiple current sources

To explain the VB method for a multiple dipole case, we assumed another sinusoidal current dipole of the same current strength at $X_2 = (57.1 \text{ mm}, 47.2^\circ, 152.3^\circ)$ (green dot in Fig. 3A), in addition to that for the single dipole case.

The maximum free energy surface was determined in the same way as for the single current source case. The free energy peaked at $R = 70$ mm. There were two current peaks at \hat{X}_1 and \hat{X}_2 on this surface (blue and green crosses in Fig. 3F), and local surfaces (blue and green circles) were extracted as assemblies of vertex points X within 30° angular coordinate differences from \hat{X}_1 and \hat{X}_2 , that is, $(X \cdot \hat{X}_i) / (|X| |\hat{X}_i|) \geq \cos(30^\circ)$ for $i = 1, 2$. Then, a new surface model was constructed as the set of the two local surfaces with the radii R_1 and R_2 , which were determined by alternately maximizing the free energy with respect to R_1 and R_2 . The free energy was first maximized by varying R_2 while R_1 was fixed at 70 mm. The free energy peaked at $R_2 = 57$ mm (green line in Fig. 3G). Next, the free energy was maximized by varying R_1 while R_2 was fixed at 57 mm. The free energy peaked at $R_1 = 75$ mm (blue line in Fig. 3G). In principle, it is necessary to iterate this procedure until the free energy converges to a maximum value. However, no increase in the free energy was found with further iteration in this case, and R_1 and R_2 were estimated at 75 and 57 mm, respectively. Plots of sinusoidal current peaks estimated for the two local surfaces (R_1 and R_2 , 75 and 57 mm) indicate that the VB current estimates were correctly focused to the two true dipole sources (blue and green dots in Fig. 3H). Localization errors for X_1 and X_2 were 2.7 and 2.6 mm, respectively.

The VB method can recover multiple current sources from MEGs if the estimated current peaks in the maximum free energy surface can be discriminated and correspond to all of the current sources. To specify each source position, a new surface model is constructed from a set of local surfaces for the individual current peaks. The VB estimates can be completed by maximizing free

energy by alternately changing each local surface radius. To investigate whether the VB method can discriminate two sources placed at nearby points, simulations were conducted by randomly generating 350 sets of two dipole sources (maximum distance, 35 mm) in the MRI-based cortical surface model. When the Euclidean distance between the two sources was small, the VB method often estimated only a single current peak (Table 1). The percentages that only a single source was recovered were 81.8%, 40.4%, and 14.6% for the distance ranges 0–5, 5–10, and 10–15 mm, respectively. When only a single current peak was estimated, the localization error for each source was defined as the Euclidean distance between each source and the single estimated current peak. Localization errors for the two sources were averaged over 350 cases according to the range of the distance, that is, 0–5, ..., 30–35 mm (Table 1). Localization errors slightly increased around distances of 5–20 mm. However, the localization errors decreased for distances less than 5 mm, since the single estimated current peak could approximate both sources within 5-mm localization error. Table 1 also shows that the mean localization errors for the two sources remained below 7 mm across the distances of 0–35 mm. These results indicate that the minimal distance for the VB estimates to resolve two current sources is roughly 15 mm if no other information than MEG is available.

MEG with structural MRI information

When structural MRI data were available, the MRI-based cortical surface model for current sources (Fig. 2A) could be used as a surface model for current estimates. The VB estimate was conducted in two steps: (1) low-resolution current estimation using a coarse surface model and (2) high-resolution estimation using a fine surface model. The coarse surface model (number of vertex points, 2494; average intervertex distance, 6 mm) was constructed by reducing the number of vertex points down to one-tenth of that for the MRI-based cortical surface model. A locally distributed current source with a Gaussian profile was assumed at each vertex of the coarse surface model. Then, the current in the coarse surface model Z_L induces the current $J = W_L Z_L$ in the MRI-based cortical surface model, where W_L denotes the Gaussian profile of the distributed currents. The Gaussian profile is given by $(W_L)_{nj} = \exp(-(d_{nj}/\sigma)^2)$, where d_{nj} denotes the shortest path length between the n th vertex point of the MRI-based cortical surface model and the j th vertex point of the coarse surface model. The magnetic field B produced by the coarse model current Z_L is given by $B = G \cdot J = G \cdot W_L \cdot Z_L$. Therefore, the effective lead field matrix for Z_L is given by $G \cdot W_L$. The coarse model current Z_L was estimated by the VB method with the hierarchical prior Eq. (9). The non-informative prior was assumed for $P_0(\alpha_L)$, where α_L was an inverse variance parameter for Z_L . The VB algorithm in Appendix

Table 1
Discriminability of two sources

Distance (mm)	0–5	5–10	10–15	15–20	20–25	25–30	30–35
% of failure	81.8	40.4	14.6	2.0	2.4	0.0	0.0
Localization error	3.62	6.95	6.43	5.50	3.89	3.78	3.04

Discriminability of two sources. Discriminability of two sources for the VB method with no other information was investigated by randomly generating 350 sets of two dipole sources (maximum distance, 35 mm) in the MRI-based cortical surface model. The percentages that only a single source is recovered are listed according to the range of the Euclidean distance between the two sources, that is, 0–5, ..., 30–35 mm. Localization errors for two sources were averaged over 350 sets according to the range of the distance and listed in millimeters.

B.4 was modified such that \mathbf{G} was replaced by $\mathbf{G} \cdot \mathbf{W}_L$, and $\mathbf{W} = 0$. The current in the MRI-based cortical surface model is given by

$$\mathbf{J}(t) = \Sigma_{\alpha}^{-1} \cdot \mathbf{G}' \left(\mathbf{G} \cdot \Sigma_{\alpha}^{-1} \cdot \mathbf{G}' + \beta^{-1} \mathbf{I}_M \right)^{-1} \cdot \mathbf{B}(t),$$

$$\Sigma_{\alpha}^{-1} = \mathbf{W}_L \cdot \mathbf{A}_L^{-1} \cdot \mathbf{W}_L', \quad (16)$$

where $\mathbf{A}_L^{-1} = \text{diag}(\alpha_L^{-1})$ denotes the variance matrix for \mathbf{Z}_L . The comparison between Eqs. (13) and (16) implies that this current model is equivalent to assuming a hierarchical prior with the current covariance matrix, $(\mathbf{W}_L \cdot \mathbf{A}_L^{-1} \cdot \mathbf{W}_L')$.

High-resolution VB estimation was conducted for local peak currents of the coarse surface model. As in the case of MEG for multiple dipoles with no other information, local surfaces were extracted from the MRI-based cortical surface model as an assembly of vertex points within 35 mm from the vertex point of each local current peak. The VB current estimate was performed for a fine surface model constructed by a set of extracted local surfaces using the hierarchical prior, Eqs. (11) and (12), with a non-informative prior. The Gaussian filter for the smoothness constraints was assumed as $\mathbf{W}_{nm} = \exp(-(d_{nm}/\sigma)^2)$, where d_{nm} denotes the shortest path length between the n th and the m th vertex point of the MRI-based cortical surface model. The full width half maximum (FWHM) of the Gaussian filter was assumed to be 8 mm. We tested several Gaussian filters with different FWHM. The biologically plausible value, FWHM = 3.5 mm according to the point spread map of fMRI (Engel et al., 1997), was too small in the present MRI-based cortical surface model. The number of points within a 3.5-mm diameter for each vertex point was very few (average number, 2 ± 1), which caused inhomogeneity in the smoothness constraints. The value FWHM = 8 mm (average neighbor number, 8 ± 2) was selected as the compromise between spatial resolution and smoothness of the estimation results.

The performance of the VB estimates with structural MRI information was compared with that of the conventional linear inverse filters. The minimum norm estimate (MN) (Hamalainen and Ilmoniemi, 1994) and the minimum norm estimate with smoothness constraints (MNC) were tested using the same MRI-based cortical surface model. To avoid the regularization problem of determining the regularization coefficient, simulation was conducted only for the noiseless case. The MN estimate is defined as the linear inverse estimate (3) with $\Sigma_0 = \mathbf{I}_N$ and $\beta^{-1} = 0$. The MNC estimate is defined as that with $(\Sigma_0^{-1})_{nm} = \exp(-(d_{nm}/\sigma)^2)$ (FWHM = 8 mm) and $\beta^{-1} = 0$. This smoothness constraint with Gaussian profile is comparable with the smoothness constraint used in the high-resolution VB estimation.

MEG with MRI and fMRI information

If fMRI data is available in addition to structural MRI data, that information may be used as hierarchical priors. fMRI signals were assumed to uniformly spread out in a circular area of a 15-mm radius around the point source brain activity: the fMRI signal V_n at the n th vertex point was one in the 15-mm radius around the source current (center) position, and $V_n = 0$ for other positions. The VB current estimate was performed in the same way as for the high-resolution VB with MRI information, except for the introduction of fMRI signals into the hierarchical priors defined by Eq. (32) (see Appendix B.4).

The performance of the VB estimates was also compared with that of the Wiener filter estimates (Dale et al., 2000; Kajihara et al., 2004; Phillips et al., 2002), using the same MRI and fMRI information. For the noiseless case, the Wiener filter estimate is given by Eq. (3) with $(\Sigma_0^{-1})_{nm} = V_n V_m \exp(-(d_{nm}/\sigma)^2)$ (FWHM = 8 mm) and $\beta^{-1} = 0$.

Performance improvement by combining MEG with MRI and fMRI

Table 2 compares localization errors of the VB estimates averaged over 500 randomly placed dipoles and distributed sources in the MRI-based cortical surface model. Noiseless and noisy conditions were tested under three situations: (1) MEG with no other data (NO), (2) MEG with MRI (MRI), and (3) MEG with both MRI and fMRI (MRI + fMRI). One can see that localization errors in the VB estimates reduce for all cases as more information becomes available.

Interestingly, the performance of the VB estimate is slightly better under the noisy condition than that of the noiseless condition. In the noisy condition, dynamic sources with sinusoidal waveform were assumed and the time sequence of MEG data was used to estimate the current variance in the VB method. On the other hand, static sources were assumed in the noiseless case, and only a single set of MEG data was used to estimate the current variance. This result shows the usefulness of the time sequence information of MEG data.

The contribution of MRI and fMRI information to improve localization errors for single dipole sources is also shown by scatter plots of localization errors as a function of the lead field norm $\|\mathbf{G}_n\| = \sqrt{(\Sigma_{m=1}^M G_{m,n}^2)}$ (Figs. 4A–C). The lead field norm represents the magnitude of MEG produced by a single dipole source, and it becomes small as the source location becomes deep and the current orientation approaches the radial direction. Local-

Table 2
Localization error for single source

Method		VB			MN	MNC	Weiner
		NO	MRI	MRI + fMRI	MRI	MRI	MRI + fMRI
Dipole source	No Noise	4.85 (8.19)	1.33 (3.49)	0.28 (0.98)	21.6 (31.1)	17.0 (24.6)	1.30 (1.90)
	N/S = 0.1	4.64 (7.72)	0.71 (2.08)	0.11 (0.57)	–	–	–
Distributed source	No Noise	5.90 (7.71)	3.64 (3.30)	3.12 (1.61)	20.1 (28.0)	16.4 (24.0)	1.65 (1.53)
	N/S = 0.1	5.57 (7.44)	2.81 (2.70)	2.40 (1.94)	–	–	–

Summary of the localization error (mm). The localization errors averaged over 500 randomly placed dipole and distributed sources in the MRI-based cortical surface model are listed in millimeters. The standard deviations are listed in parentheses. Noiseless and noisy conditions, where MEG signals contain Gaussian sensor noise with N/S of 0.1, were tested under three different situations depending on the available information: no other data than MEG (NO), MEG with MRI (MRI), and MEG with MRI and fMRI (MRI + fMRI). The estimation methods are the VB method (VB), the minimum norm estimate (MN), the minimum norm estimate with smoothness constraints (MNC), and the Wiener filter method (Wiener). The noisy conditions were not tested for MN, MNC, and Wiener filter methods as indicated by – in the table.

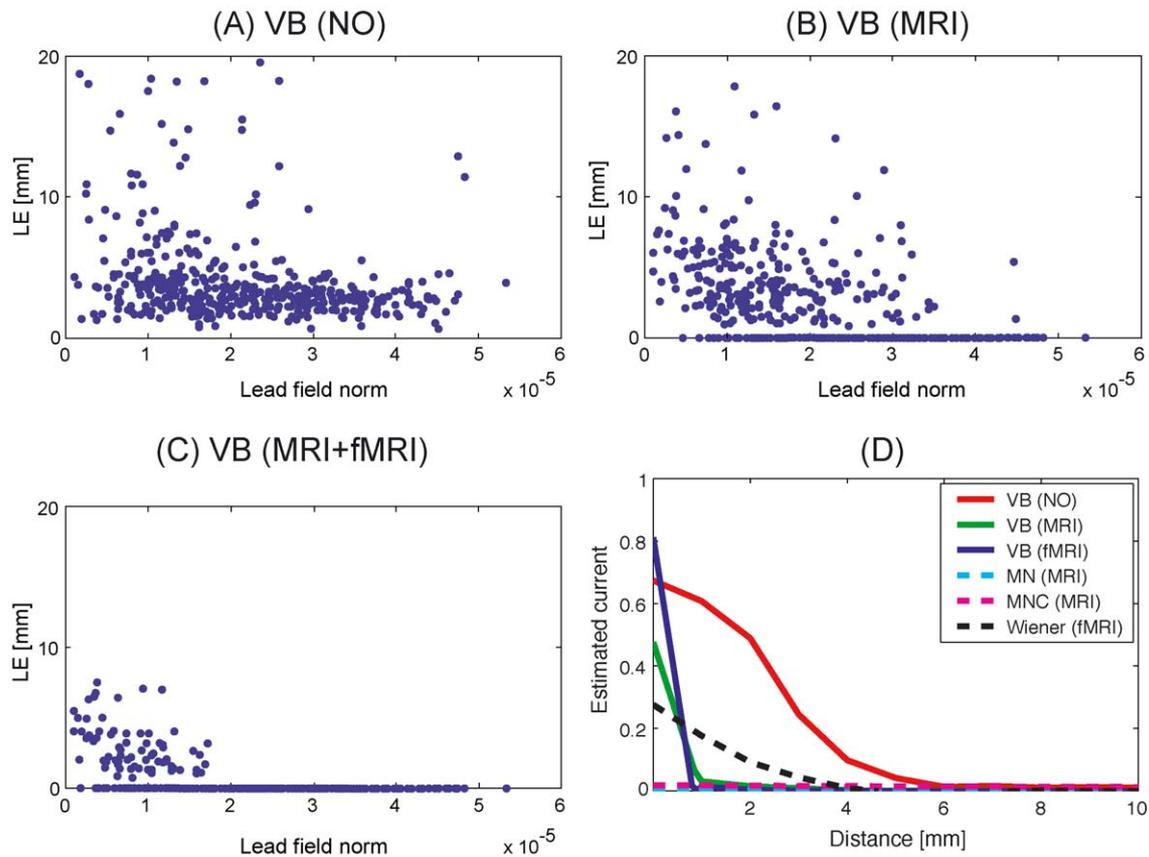


Fig. 4. Localization error for a single dipole source. (A–C) Scatter plots of localization errors for single dipole source as a function of the lead field norm. (A) The VB method for MEG with no other data (NO). (B) That for MEG with MRI (MRI). (C) That for MEG with both MRI and fMRI (MRI + fMRI). (D) Point-spread curves of the current estimates around the source dipole were calculated by averaging over the 500 dipole sources as a function of the Euclidean distance from the source dipole position. Estimation methods are VB estimate with NO other information (solid red line), VB estimate with MRI (solid green line), VB estimate with MRI and fMRI (solid blue line), minimum norm estimate (MN) with MRI (dashed cyan line), minimum norm estimate with smoothness constraints (MNC) with MRI (dashed magenta line), and Wiener filter estimate with MRI and fMRI (dashed black line).

ization errors for dipole sources with small lead field norms were larger than those with large lead field norms in all situations (Figs. 4A–C) because the estimation of the sources with small lead field norms was difficult. The scatter plots also show that introduction of MRI and fMRI information significantly decreases localization errors, especially for large lead field norms. In addition, the positions of dipole sources were perfectly recovered for large lead field norms when both MRI and fMRI information were imposed.

Table 2 also shows that the precision of the VB estimates with MRI information is far better than those of the MN and MNC estimates both for single and distributed sources. The precision of the VB estimates with MRI and fMRI information is comparable with that of the Wiener filter, that is, that of the VB estimates is slightly better than that of the Wiener filter for a single dipole source, but the converse is true for a single distributed source (Table 2). It should be noted that the localization error of less than 7 mm for the distributed source implies that the source position is correctly estimated inside the distributed source, since the radius of the distributed source is 7 mm. Therefore, the differences of localization errors less than 7 mm do not have much meaning for the distributed source, unless the detailed spatial profile of the estimated currents are examined.

The spatial profile of the estimated currents for single dipole source was assessed as the point-spread curve. It was defined by the estimated current profile around the source dipole and averaged

across the 500 dipole sources as a function of the Euclidean distance from the source dipole position. If the current estimation is perfect, the value of the point-spread curve at the origin, which is equal to the estimation gain, should be one and that for positive distance should be zero. Fig. 4D shows the point-spread curve of the VB estimates for three situations: (1) MEG with no other data (NO), (2) MEG with MRI (MRI), and (3) MEG with both MRI and fMRI (MRI + fMRI). The point-spread curve for NO case (red line in Fig. 4D) was rather broadly spread with FWHM = 5 mm. The estimation gain (height of the point-spread curve at the origin) was 0.68 in this case. The point-spread curves for MRI and MRI + fMRI cases (green and blue lines in Fig. 4D) were sharply tuned at the source position with FWHM = 1 mm, and the estimation gains were 0.48 and 0.81, respectively. These results show that the introduction of MRI and fMRI information significantly increases the spatial resolution of the VB method.

The point-spread curves of the MN and the MNC methods (dashed cyan and magenta lines in Fig. 4D) were almost flat and nearly overlapped with the x -axis, which indicated very poor resolution of these methods. The point-spread curve of the Wiener filter estimate (dashed black line in Fig. 4D) was worse than that of the VB estimate for MRI + fMRI (FWHM = 4 mm; estimation gain, 0.27).

These results show that the precision of the source localization for the VB method is comparable with that for the Wiener filter

method when structural MRI and correct fMRI data are available. On the other hand, the estimated current profile of the Wiener filter method have much larger spread than that of the VB method for a single dipole source. Both measures are much better for the VB method than for the MN and the MNC methods when only structural MRI data is available.

VB and linear inverse filter estimates for simulated cortical responses

Comparison of the VB performance for single and distributed current sources among the three situations indicates that the MRI and fMRI information effectively improves the spatial resolution as well as the reliability of the VB estimates. However, mismatches could occur between MEG and fMRI information for several reasons, including noise artifacts and the slow hemodynamic response of fMRI. We investigated how these mismatches affect the VB and the Wiener filter estimates for MEGs by conducting simulations reproducing realistic visual cortical responses.

We identified cortical areas in a subject's (MA) brain including V1v, V1d, V2v, V2d, V3, and VP (Fig. 5A). This was based on the cortical retinotopy determined by fMRI using a rotating pole and expanding/shrinking annulus stimuli (Engel et al., 1994; Linden et al., 1999; Sereno et al., 1995). We simulated sequential activities evoked in visual cortical areas by lower-quadrant visual field

stimulation. Sinusoidal currents uniformly distributed in vertex points within a circular area (radius, 5 mm) of V1v were assumed and those in V2v and VP were activated with 50- and 100-ms delays to those in V1v, respectively (Figs. 5C and D). White Gaussian sensor noise with N/S = 0.1 was added to the generated MEG signals.

To evaluate spatio-temporal patterns of the estimated currents, we defined local V1v, V2v, and VP currents as the sum of the estimated currents within a 5-mm radius of the local peak current positions in the V1v, V2v, and VP areas, respectively. The precision of the estimation was assessed as the estimation gains for the V1v, V2v, and VP local currents at 50, 100, and 150 ms, respectively, where each source current became maximum. The normalized root mean squared errors (RMSEs) were also calculated for the temporal sequences of the local V1v, V2v, and VP currents. The estimation gain should be one and RMSE should be zero for perfect estimation.

MEG with no fMRI information

We first investigated the performance of the VB estimate for the simulated visual cortical responses, in the situation where MEGs were supplemented with only structural MRI information. The VB estimate was conducted according to the method of high-resolution estimation for MEG combined with MRI. Time-lapse current maps using pseudocolor representation revealed that the VB estimate

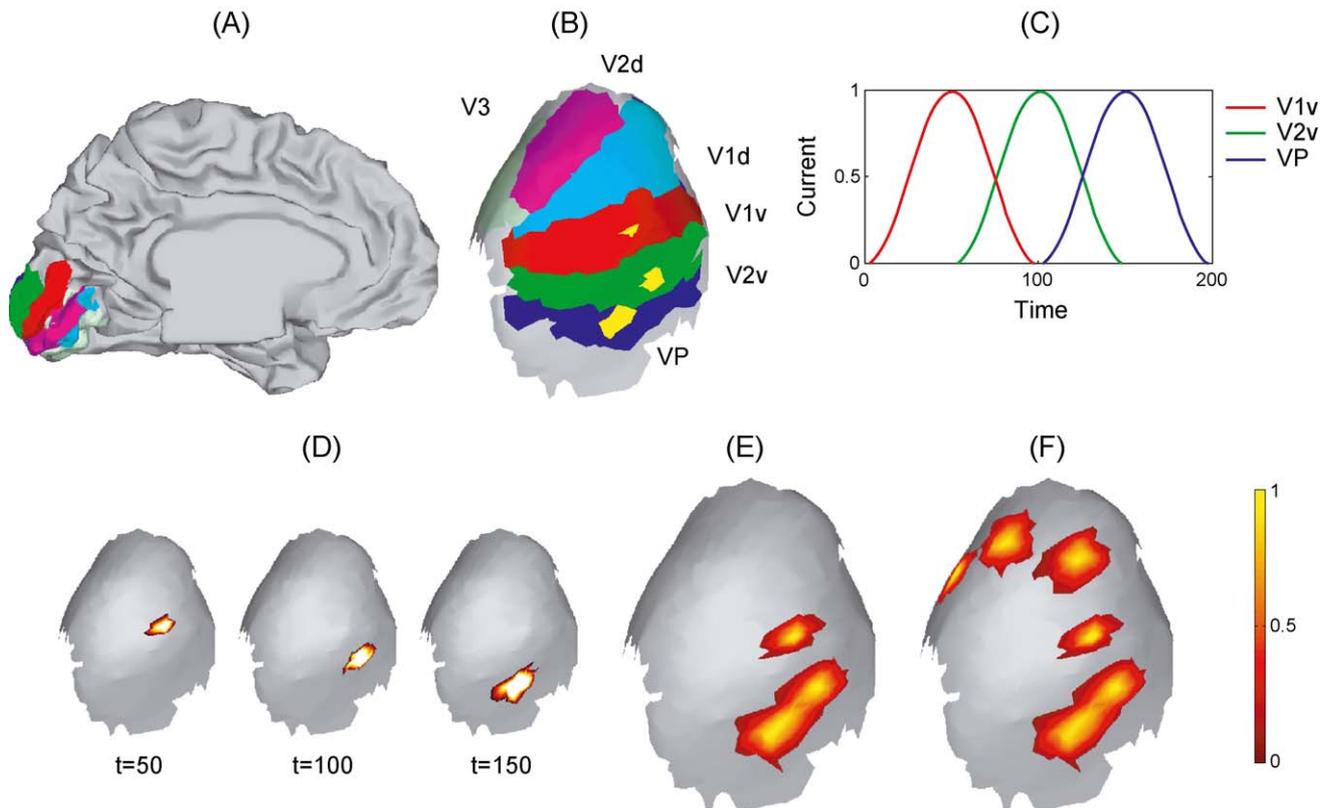


Fig. 5. Simulated visual responses. (A) The cortex model used for the simulated visual responses. The visual area is depicted on the left hemisphere. (B) The visual areas are shown by using an inflated map of the cortex. Visual areas V1v, V1d, V2v, V2d, VP, and V3 are shown by different colors. The generated source currents with a 5-mm radius in the V1v, V2v, and VP areas are shown in yellow. (C) Time course plot of the V1v (red line), V2v (green line), and VP (blue line) currents. Total current strength of each source is shown. Abscissa denotes the time in ms. (D) Time-lapse map of source currents at $t = 50$, 100, and 150 ms are shown on the inflated map. (E) Simulated fMRI activity given as the sum of the source signal amplitude and 0.1 N/S white Gaussian noise convolved with a Gaussian smoothing filter (FWHM = 6 mm). (F) False-positive fMRI signal including V1d, V2d, and V3 activities in addition to the true fMRI signals in V1v, V2v, and VP.

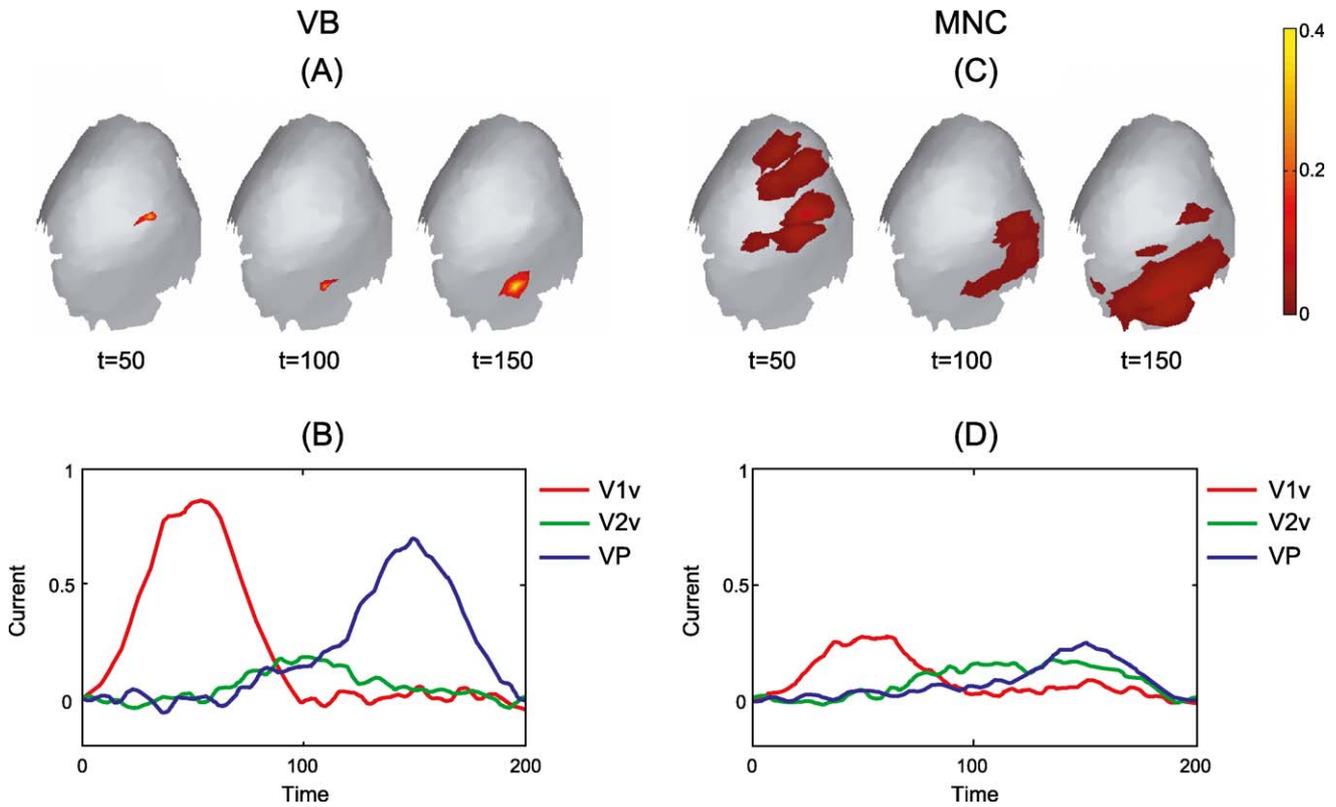


Fig. 6. Estimated current with no fMRI information. (A and C) Time-lapse current maps estimated by the VB method (A) and the MNC method (C) with no fMRI information are plotted using pseudocolor representation at $t = 50, 100,$ and 150 ms. (B and D) Time course plot of the V1v (red line), V2v (green line), and VP (blue line) currents estimated by the VB method (B) and the MNC method (D) with no fMRI information are plotted. Each current is defined as a current sum within a 5-mm radius of the local peak current positions in the V1v, V2v, and VP areas.

virtually recovered V1v and VP responses but left V2v responses undetectable (Fig. 6A). Correspondingly, the time plots for local currents indicated a rather fine match between the estimated and true source currents assumed in V1v and VP (compare red and blue traces in Fig. 6B with those in Fig. 5C). However, the estimated V2v current was far smaller than the true source current assumed in V2v (compare green trace in Fig. 6B with those in Fig. 5C). These findings can be confirmed by the estimation gains and RMSEs for VB with the no fMRI situation in Table 3. The current estimates for V2v were difficult, since the lead field for V2v was much smaller (roughly 0.1) compared with those for V1v and VP currents, and V2v has a very close distance (less than 2 mm) to VP.

The performances of MNC were very poor. The time-lapse maps revealed that the estimated currents broadly spread out from the areas in V1v, V2v, and VP for which current sources were assumed (Fig. 6C). The time plots showed much smaller peaks for V1v, V2v, and VP than those for the VB estimates (Fig. 6D). In fact, the estimation gains for MNC were very small and RMSEs were large as can be seen in Table 3. In addition, the time course of V2v currents was more broadened than that of the true source currents (compare green trace in Fig. 6D with those in Fig. 5C). These findings indicate that the MNC estimate is not feasible for the MEG of realistic visual cortical responses. The performance of the VB method is much better than that of the

Table 3
Accuracy of spatio-temporal current estimates

Method	VB			MNC		Wiener	
	Information	No fMRI	Correct fMRI	Incorrect fMRI	No fMRI	Correct fMRI	Incorrect fMRI
Estimation gain	V1v	0.85	0.95	0.96	0.27	0.83	0.93
	V2v	0.19	0.86	0.86	0.15	0.58	0.24
	VP	0.70	0.87	0.88	0.24	0.95	1.06
RMSE	V1v	0.14	0.06	0.05	0.72	0.20	0.35
	V2v	0.82	0.21	0.20	0.84	0.51	0.59
	VP	0.34	0.12	0.12	0.77	0.34	0.55

Accuracy of spatio-temporal current estimates. The performance of the VB method was investigated using simulated visual cortical responses under three situations: MEG with MRI data (No fMRI), MEG with MRI and correct fMRI data (Correct fMRI), and MEG with MRI and incorrect fMRI data (Incorrect fMRI). The precision of the estimation was assessed as the estimation gains for the V1v, V2v, and VP local currents at 50, 100, and 150 ms, respectively. The normalized root mean squared errors (RMSEs) for the temporal sequences of the local V1v, V2v, and VP currents were also listed in the table. The MNC method with no fMRI information and the Wiener filter method with correct fMRI and incorrect fMRI information were also tested.

MNC estimate, but the VB estimate with no fMRI information is still insufficient to recover the fine dynamic structure of visual responses.

MEG combined with correct fMRI information

Next, we examined whether the correct fMRI information improves the VB and the linear filter performance for the same simulated cortical responses. fMRI signals were assumed to be the sum of the source signal and 0.1 N/S Gaussian white noise convolved with a Gaussian smoothing filter for fMRI imaging (FWHM = 6 mm) (Fig. 5E). The VB estimate was conducted in the same way as for the situation of MEG combined with MRI and fMRI. The time-lapse maps and time plots indicated that the VB estimates nearly correctly recovered the current activities in V1v, V2v, and VP (compare Figs. 7A,B with Figs. 5D,C). The estimated currents were slightly smaller than the true source currents assumed in these cortical areas. The estimation gains for V1v, V2v, and VP currents were 0.95, 0.86, and 0.87, and RMSEs for V1v, V2v, and VP currents were 0.06, 0.21, and 0.12, respectively (Table 3).

The performances of the VB estimates were compared with those of the Wiener filter (Dale et al., 2000; Kajihara et al., 2004; Phillips et al., 2002). The Wiener filter estimates were conducted as Eq. (3) with $(\Sigma_0^{-1})_{nm} = CV_n V_m \exp(-(d_{nm}/\sigma)^2)$ (FWHM = 8 mm) for the same MEGs and fMRI signals. The coefficient C was determined so as to maximize the correlation between the fMRI signals and the temporal average of the estimated currents

(Kajihara et al., 2004). The Wiener filter estimates recovered activities in V1v, V2v, and VP (compare Fig. 7C with Fig. 5D), but V2v current was rather smaller than the true source current assumed in V2v. The estimation gains for V1v, V2v, and VP currents were 0.83, 0.58, and 0.95, and RMSEs for V1v, V2v, and VP currents were 0.20, 0.51, and 0.34, respectively (Table 3). In addition, the time plots show a tendency of some oscillations in time courses of V1v, V2v, and VP currents (compare Fig. 7D with Fig. 5C). Consequently, correct fMRI information almost perfectly improved the performance of the VB estimate, but this improvement was still insufficient for the linear inverse filter.

MEG combined with incorrect fMRI information

We further examined how false-positive fMRI signals affect the VB estimate. The VB estimate was conducted for the same simulated cortical responses as those for the correct fMRI situation using incorrect fMRI information. fMRI signals were assumed to contain false-positive signals in V1d, V2d, and V3 as strong as the correct fMRI signals in V1v, V2v, and VP in addition to the correct ones in these areas (Fig. 5F). The performance of the VB estimate remained almost unaffected by false-positive fMRI signals as can be seen from the estimation gains and RMSEs in Table 3. The time-lapse map and time course plot of the VB estimates were almost the same as those for the correct fMRI situation (compare Figs. 8A,B with Figs. 7A,B). No false currents were estimated in V1d, V2d, and V3, where false-positive fMRI signals were assumed (compare Fig. 8A with Fig. 5D).

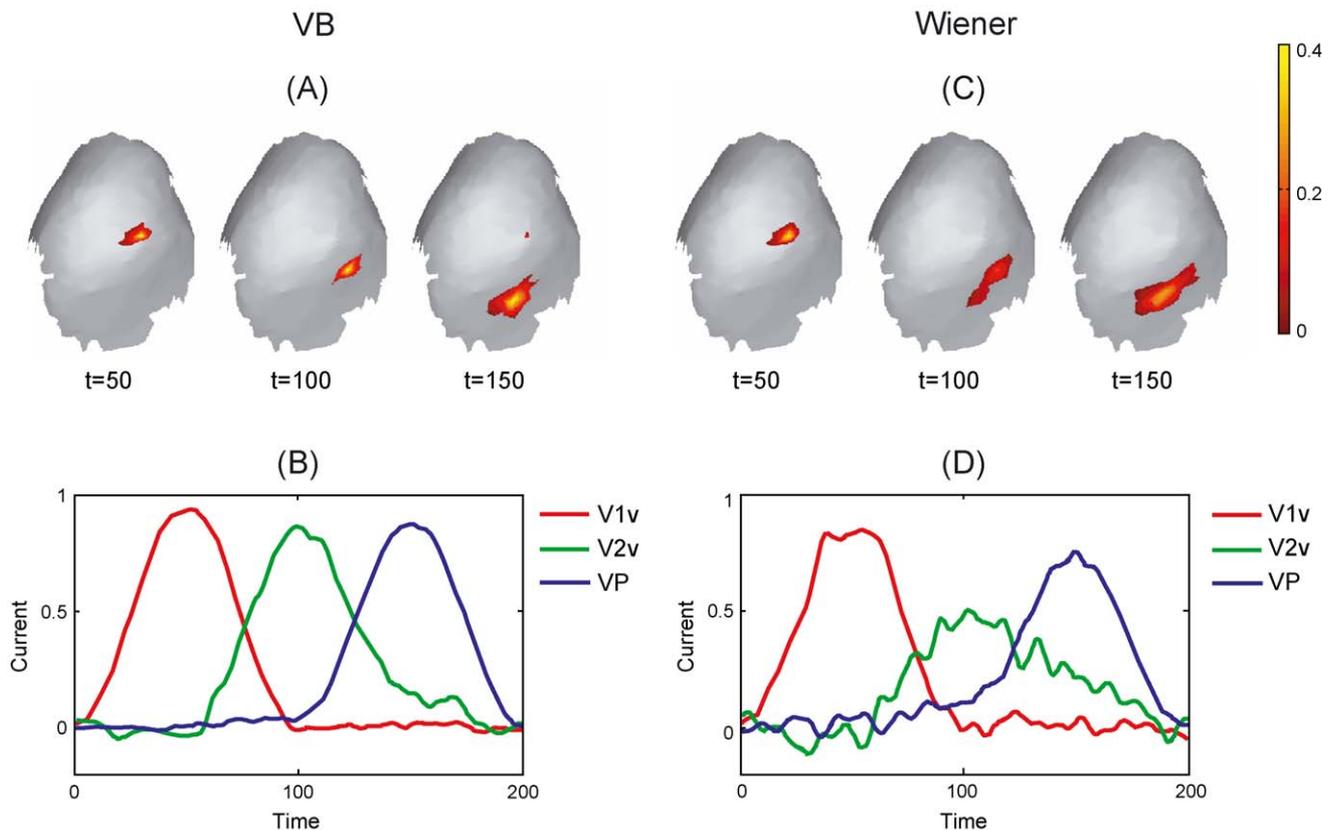


Fig. 7. Estimated current with correct fMRI information. (A and C) Time-lapse current maps estimated by the VB method (A) and the Wiener filter method (C) with correct fMRI information are plotted at $t = 50, 100,$ and 150 ms. (B and D) Time course plot of the V1v (red line), V2v (green line), and VP (blue line) currents estimated by the VB method (B) and the Wiener filter method (D) with correct fMRI information.

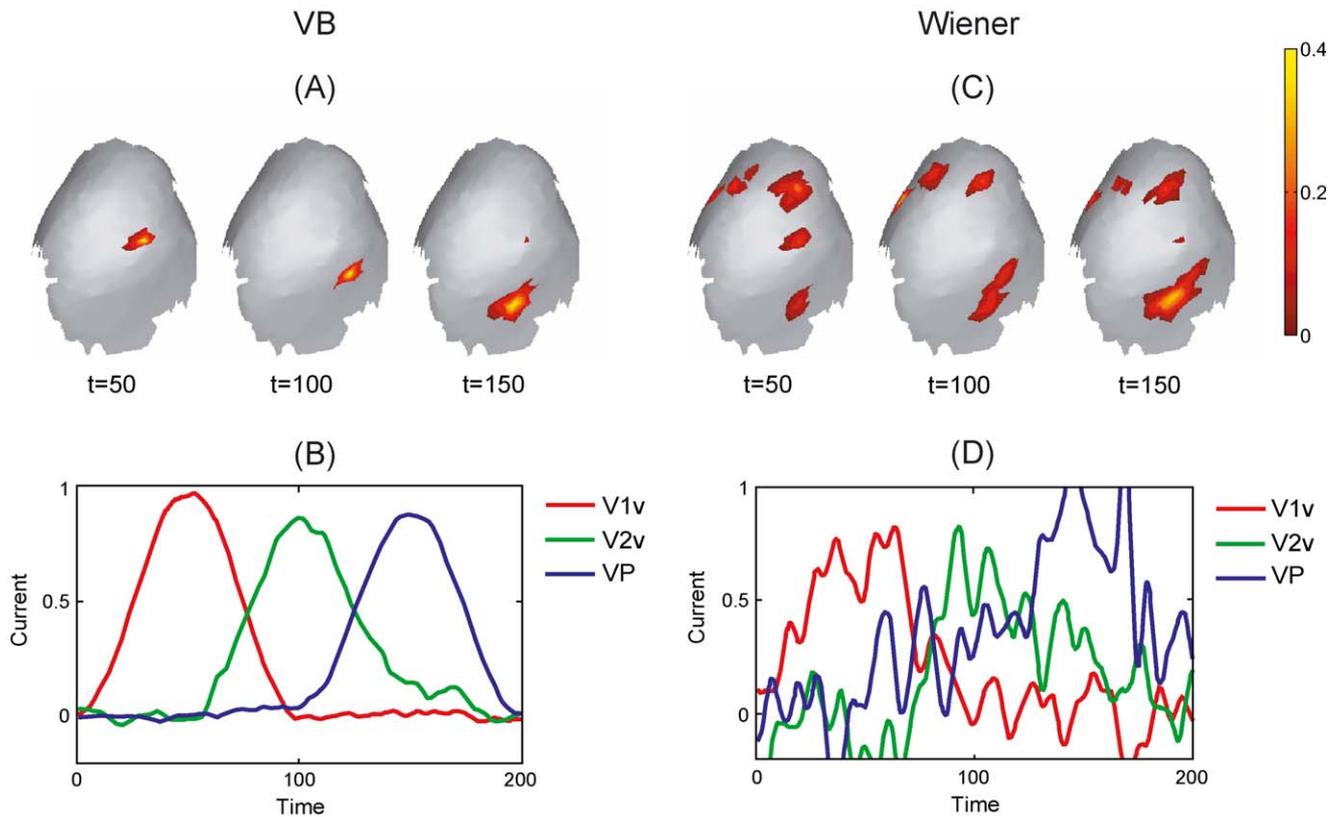


Fig. 8. Estimated current with false-positive fMRI information. (A and C) Time-lapse current maps estimated by the VB method (A) and the Wiener filter method (C) with false-positive fMRI information are plotted at $t = 50, 100,$ and 150 ms. (B and D) Time course plot of the V1v (red line), V2v (green line), and VP (blue line) currents estimated by the VB method (B) and the Wiener filter method (D) with false-positive fMRI information.

In contrast, the Wiener filter estimate was seriously affected by false-positive fMRI signals. The time-lapse map indicated that substantial currents were estimated around the area of V1d, V2d, and V3, where false-positive signals were assumed (compare Fig. 8C with Fig. 5F). Correspondingly, the time course of V1v, V2v, and VP currents became much more oscillatory than those for the correct fMRI situation (compare Fig. 8D with Fig. 7D) due to increased ambiguity caused by false-positive signals. As a result, RMSEs became larger than those for the correct fMRI case (Table 3).

These findings show that the Wiener filter estimates are sensitive to false-positive fMRI signals but not the VB estimates.

Separation of simultaneously active currents

The spatial resolution of the VB and the Wiener filter estimates were also examined. The resolution matrix defined by Eq. (18) (see Appendix A) was calculated using the optimized inverse filters for MEGs of the simulated visual responses with the correct fMRI. The n th row of the resolution matrix represents the point-spread profiles of the current estimates for a single dipole source at the n th vertex point. The spatial profiles of the current estimates for simultaneously active V1v, V2v, and VP currents (Fig. 9A) were calculated by summing the rows of the resolution matrix corresponding to the active vertex points of the V1v, V2v, and VP. Profile maps indicate that both the VB and the Wiener filter methods nearly correctly recover the source current profiles in V1v, V2v, and VP (compare Figs. 9B and C with Fig. 9A).

Likewise, the spatial profiles of the current estimates for simultaneously active V1v, V2v, and VP (Fig. 9A) were determined

according to the resolution matrix of the VB inverse filter optimized for the MEGs combined with the false-positive fMRI (Fig. 9D). Profile maps indicate that the VB estimates for simultaneously active V1v, V2v, and VP correctly recover the source current profiles (compare Figs. 9E with A) under the false-positive fMRI information. Conversely, the current estimates based on the resolution matrix of the Wiener filters optimized for the same set of MEG and incorrect fMRI as those for the VB estimates were rather seriously affected by the false-positive fMRI signals (compare Figs. 9F with D).

MEG combined with insufficient fMRI information

fMRI also may not provide correct information for traveling cortical responses. We studied the performance of the VB estimate assuming that a uniform circular current source (radius, 5 mm) stayed in V1 for the 200-ms period of time and traveled (distance, 50 mm; velocity, 0.5 m/s) to the other region in V1 (Fig. 10A). This may correspond to V1 responses, where a spot near the fixation point moves toward the upper-right direction.

fMRI signals were given as a sum of the temporal average of the source currents and 0.1 N/S Gaussian white noise convolved with a Gaussian smoothing filter (FWHM = 6 mm) (Fig. 10A). Time-lapse maps indicate that the VB estimates for MEGs generated by the traveling V1 responses almost recovered the spatio-temporal patterns of the traveling source currents (compare Figs. 10C with B). However, the estimates were slightly deformed and weaker than the true traveling sources. On the contrary, the Wiener filter estimates almost failed to recover the spatio-temporal patterns of the traveling source currents (compare Figs. 10D with B).

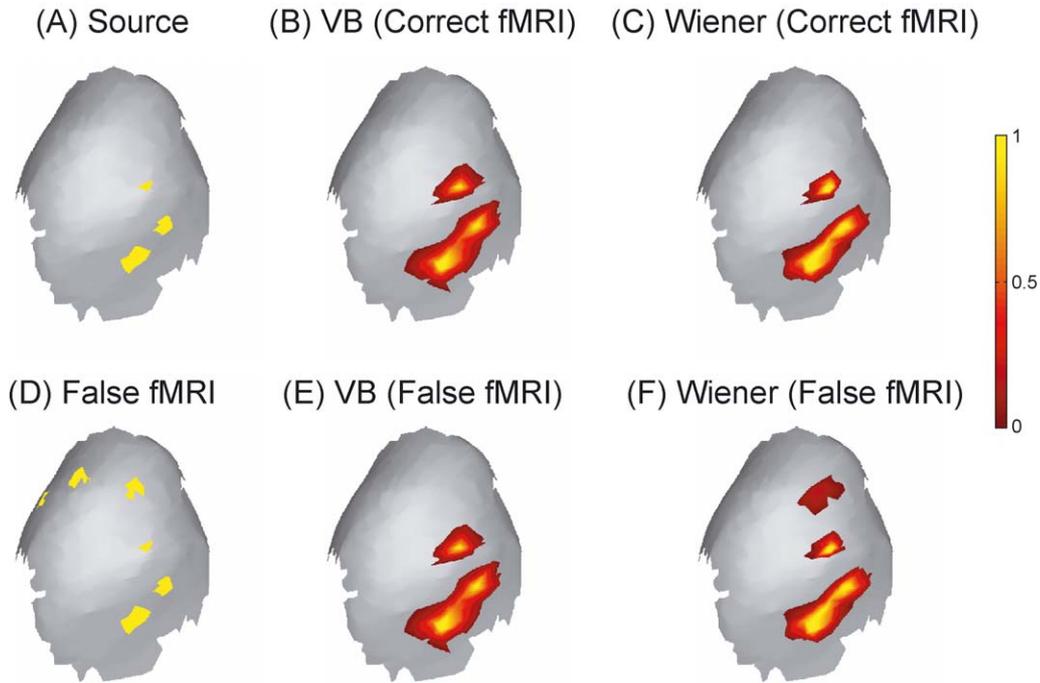


Fig. 9. Current profiles for simultaneously active V1v, V2v, and VP currents. Spatial profiles of the current estimates for simultaneously active V1v, V2v, and VP currents are shown on the inflated map of the visual area. (A) Simultaneously active currents in V1v, V2v, and VP. (B and C) Current profiles estimated by the VB method (B) and the Wiener filter method (C) with correct fMRI information. (D) False-positive fMRI signals in V1d, V2d, and V3 in addition to the correct fMRI signals in V1v, V2v, and VP. (E and F) Current profiles estimated by the VB method (E) and the Wiener filter method (F) with false-positive fMRI information.

A static source, which corresponded to the large fMRI signal, was recovered most of the time. These results show the advantage of the VB estimate over the Wiener filter estimate for spatio-temporally changeable brain activities.

Discussion

We proposed a novel hierarchical Bayesian method to resolve the MEG inverse problem, which gives a unified framework for three possible situations: (1) MEG with no other data (NO), (2) MEG with structural MRI data on cortical surfaces (MRI), and (3) MEG with both structural MRI and fMRI data (MRI + fMRI). Our method estimates the current variance from the time sequence of the MEG data by introducing a hierarchical prior. The fMRI information can be imposed as prior information on the variance distribution rather than the variance itself so that it gives a soft constraint on the variance. The spatial smoothness constraint can also be incorporated as a hierarchical prior. The smoothness constraint reduces the effective degree of freedom in the estimation model, and thus it reduces the ill-posedness of the inverse problem.

Because of the hierarchical prior, the estimation problem becomes nonlinear, so the approximate posterior distribution of the source current is calculated by using the Variational Bayesian (VB) method. An important difference between the VB method and the MAP estimate used for the linear inverse methods is that the VB estimate calculates the posterior probability distribution of the source current while the MAP estimate calculates the optimal current value that maximizes the posterior distribution. The VB method takes account of the spread of the posterior distribution in

estimating the current variance, so robust estimation for the current variance can be achieved. In the MEG inverse problem, the number of unknown parameters is much larger than the number of MEG sensors M . The parameter estimation of such a system often suffers from overfitting to the observed data. The VB method can avoid such overfitting because the model complexity penalty term in the free energy (see Eq. (22) in Appendix B.1) freezes unnecessary degrees of freedom by suppressing the estimation gain for small variance points. The advantages of the VB estimate over the MAP estimate were also shown by Penny et al. (2003) for fMRI data analysis. It is also noted that the current variances are estimated from T time samples of the MEG data. Therefore, the effective number of data for variance estimation becomes $(T \times M)$. This improves the accuracy of the variance estimation.

Since the VB method solves the nonlinear estimation problem, the VB solution converges into a local maximum of the free energy. Therefore, the converged solution depends on the initial value of the estimated current variance. Fortunately, we have reasonable candidates for the initial value of the current variance. When there is no fMRI information, the initial value of the current variance is set to the same value for all dipole currents. This setting corresponds to the minimum norm solution, that is, the VB current estimate at the first iteration is given by the minimum norm solution. The VB method, then, improves the accuracy of the current estimate by estimating the current variance from the time sequence of the MEG data. A single hyperparameter a_0 controls the initial variance as $(\alpha_{0n}^{-1} = a_0/G_{ave}^2)$, which is normalized by the average of the lead field norm $G_{ave}^2 = (\sum_{n=1}^N \sum_{m=1}^M G_{m,n}^2)/N$. In all the simulations of this paper, $a_0 = 10$ was used. When there is fMRI information, it is also used as the initial value of the current variance (see

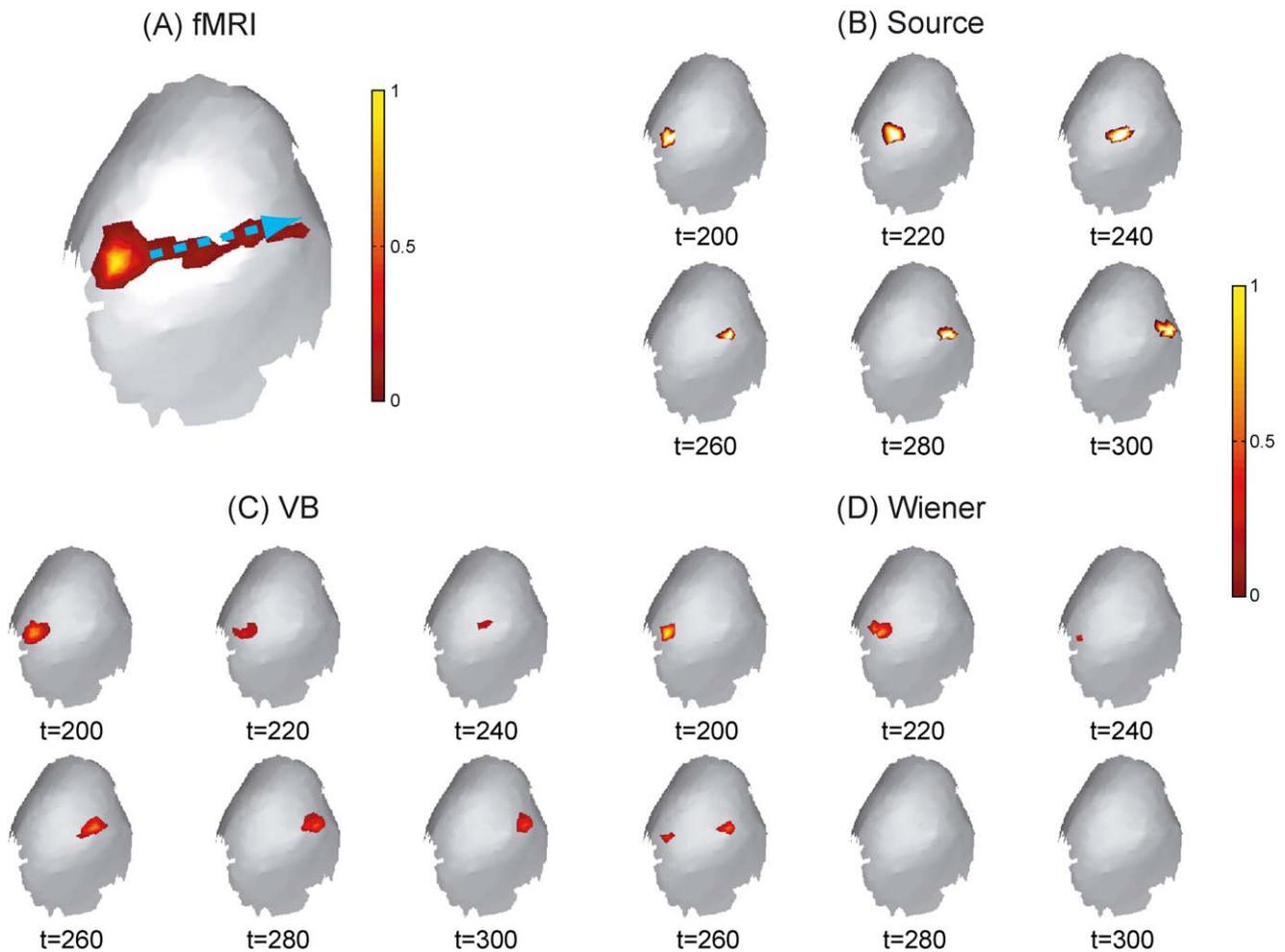


Fig. 10. Traveling wave responses in visual area. (A) Traveling wave responses schematically shown by an arrow, assuming that a uniform circular current source (radius, 5 mm) stayed in V1 for a 200-ms period of time and traveled (distance, 50 mm; velocity, 0.5 m/s) to the other region in V1. Simulated fMRI activity given as the sum of the temporal average of the source currents and 0.1 N/S Gaussian white noise convolved with a Gaussian smoothing filter (FWHM = 6 mm) is also shown using pseudocolor representation. (B) Time-lapse map of the source currents at $t = 200, 220, \dots, 300$ ms. (C and D) The time-lapse map of the currents estimated by the VB method (C) and the Wiener filter method (D).

Variational Bayesian method). Therefore, the VB current estimate at the first iteration is given by the Wiener filter solution in this case. The initial variance is controlled by two hyperparameters a_{\min} and a_{\max} as defined by (32) in the Appendix B.4. They were set as $a_{\min} = 10^{-3}$ and $a_{\max} = 10$ in all the simulations of this paper. The above hyperparameter values were determined by the maximum free energy criterion applied to one of the simulated MEG data (see the procedure explained in the Appendix B.4). The range of the searched values were $1 \leq a_{\max} \leq 10^5$ and $1 \leq a_{\max}/a_{\min} \leq 10^5$. The obtained solutions were not sensitive to small changes of the hyperparameter values.

There are other factors, which affect the estimation accuracy. In the simulations, we used the same cortical surface model both for the forward magnetic field calculation and for the inverse source current estimation. However, in the analysis of real experimental data, there may be inaccuracy in the cortical surface estimation. Since the cortical surfaces have highly folded structures, small errors in the cortical position may cause large errors in the normal direction and then cause large lead field errors. Therefore, high accuracy of the cortical surface estimation is required when the dipole direction is imposed as the structural

prior information. The effective radius (FWHM) of the smoothness (Gaussian) filter also affects the estimation performance as explained in the MEG with Structural MRI Information section. Too small filter radius disables the smoothness constraint while too large filter radius smears the current estimation. An appropriate value may depend on the resolution and the accuracy of the extracted cortical surfaces.

As pointed out previously, fMRI activity may not precisely correspond to the MEG source current activity. In particular, the fMRI activity corresponds to an average of several thousands of MEG time series data because of the time resolution difference between MEG and fMRI. When higher brain functions are examined, several different processes may occur within several seconds, and these activities are averaged out in the fMRI data. Consequently, the fMRI active areas may include inactive current areas in the MEG time slice data, and the fMRI activity for the active current may be smeared by the temporal averaging. When the fMRI activity is used to determine the source current variance such as for the Wiener filter method, the fMRI information may mislead the source current estimation in these situations. To deal with these situations, we divide the entire MEG time series into

several short periods in which the brain activity changes moderately, and the VB method is applied to the MEG data for each period. Although the VB estimate starts from the Wiener filter solution, the current variance estimation using the short-period MEG data may improve the accuracy of the source current estimation.

An advantage of the VB method is the model selection capability. If there is no structural MRI information, we consider a variable-radius surface model for current estimates and the most probable surface is determined by the surface with the maximum free energy. When there was no noise, the free energy was sharply tuned at the peak position (Fig. 3B). As the noise amplitude increased, the peak profile of the free energy curve became broader as shown in Fig. 3C. However, the peak almost stayed at the same position and there was no local peak around the maximum peak position in the presence of noise. Therefore, the source depth was safely determined by the free energy maximum. We also examined the standard single dipole fit for the single dipole source problem shown in Fig. 3. The reconstruction error (Fig. 3D) of the dipole model under the same noise condition showed the similar (reversed) tuning curve as the free energy (Fig. 3C).

In the simulations, the VB method iterated the current estimation until the free energy converged. Therefore, the processing time would be an important concern in practice. The VB estimate with the hierarchical prior Eq. (9) (50 MEG time samples and 2494 vertex points) required about 1 min on a XEON (2.2 GHz) processor, while that with the smoothness prior Eq. (11) (200 MEG time samples and 819 vertex points) required 6 min.

Summary

The advantages of the VB method over the linear filters were confirmed in simulation studies using a single source and realistic visual responses. The simulation results can be summarized as follows.

- (1) When the same information is given, the VB method achieves better performance than the conventional linear inverse filter methods.
- (2) When the correct information is given, the performance of the VB method improves as more information becomes available.
- (3) When inaccurate fMRI information is given, performance degradation of the current estimate is much less for the VB method than for the Wiener filter method.

Consequently, the VB method provides a unified framework to resolve the MEG inverse problem for three possible situations: MEGs combined with no other information, with only MRI, and with both MRI and fMRI. The first version may be useful for analysis of MEGs whose current sources are in deep brain structures such as the thalamus and basal ganglia because reliable estimation of dipole locations and orientations based on MRI is difficult. The second version may also be useful for situations in which fMRI measurement using the same experimental conditions as those for MEG is difficult. Finally, the third version provides the most reliable solution of the MEG inverse problem. It is important that the VB method may correctly resolve the inverse problem even if fMRI signals fail to convey the precise spatio-temporal

patterns of brain activities due to the slow hemodynamic response, which would commonly be the case in many experimental protocols.

In this article, we have concentrated on the MEG inverse problem with fMRI information. It should be mentioned that our method can also be applied to electroencephalography (EEG) and that fMRI information can be replaced by positron emission tomography (PET) data.

Acknowledgments

We would like to thank Dr. Daniel Callan for careful reading of the manuscript. This research was supported in part by the National Institute of Information and Communications Technology (NICT).

Appendix A. Resolution matrix and estimation gain

We define the resolution matrix and estimation gain in this Appendix and derive the inequality on the estimation Eq. (8). In the absence of noise, the magnetic field \mathbf{B} produced by the current \mathbf{J} is given by $\mathbf{B} = \mathbf{G} \cdot \mathbf{J}$. The estimated current $\hat{\mathbf{J}}$ is written as

$$\hat{\mathbf{J}} = \mathbf{L} \cdot \mathbf{B} = \mathbf{L} \cdot \mathbf{G} \cdot \mathbf{J} \quad (17)$$

using the linear inverse method (3). The resolution matrix \mathbf{R} and the estimation gain \mathcal{G} (Grave de Peralta Menendez and Gonzalez Andino, 1998) are defined by

$$\mathbf{R} = \mathbf{L} \cdot \mathbf{G}, \quad \mathcal{G} = \{R_{n,n} | n = 1:N\}. \quad (18)$$

The (n,k) component of the resolution matrix $R_{n,k}$ represents the n th estimated current when the unit current dipole is applied for the k th current, that is, $J_k = 1$ and $J_l = 0 (l \neq k)$. Therefore, the resolution matrix \mathbf{R} should be the identity matrix in the current space for perfect reconstruction. Specifically, the estimation gain $\mathcal{G}_n (n = 1:N)$ should be one. However, the following simple calculation shows the relation:

$$\sum_{n=1}^N \mathcal{G}_n = \text{Tr} \mathbf{R} = M - \beta^{-1} \text{Tr} \left(\mathbf{G} \cdot \Sigma_0^{-1} \cdot \mathbf{G}' + \beta^{-1} \mathbf{I}_M \right)^{-1} \leq M. \quad (19)$$

This equation holds because both $\beta^{-1} \mathbf{I}_M$ and $\mathbf{G} \cdot \Sigma_0^{-1} \cdot \mathbf{G}'$ are positive semidefinite matrices. The inequality (19) was derived for the first time in this paper, as far as we know.

Appendix B. Variational Bayesian method for hierarchical prior

In the following, we explain the VB method for the hierarchical prior with the smoothness constraint, (11) and (12). The hierarchical prior, (9) and (10), is a special case of the hierarchical prior with the smoothness constraint, and it is obtained by setting $\mathbf{W} \equiv 0$.

B.1. Variational Bayesian method

First, we explain the main ideas of the VB method using the hierarchical prior, (9) and (10), with a single MEG data set \mathbf{B} . The

objective of the Bayesian estimation is to calculate the posterior probability distribution of \mathbf{J} for the observed data \mathbf{B} :

$$P(\mathbf{J}|\mathbf{B}) = \int d\alpha P(\mathbf{J}, \alpha | \mathbf{B}),$$

$$P(\mathbf{J}, \alpha | \mathbf{B}) = \frac{P(\mathbf{J}, \alpha, \mathbf{B})}{P(\mathbf{B})},$$

$$P(\mathbf{J}, \alpha, \mathbf{B}) = P(\mathbf{B}|\mathbf{J})P_0(\mathbf{J}|\alpha)P_0(\alpha),$$

$$P(\mathbf{B}) = \int d\mathbf{J} d\alpha P(\mathbf{J}, \alpha, \mathbf{B}).$$

The calculation of the marginal likelihood $P(\mathbf{B})$ cannot be done analytically. In the VB method, the calculation of the joint posterior $P(\mathbf{J}, \alpha | \mathbf{B})$ is reformulated as the maximization problem of the free energy. The free energy for a trial distribution $Q(\mathbf{J}, \alpha)$ is defined by

$$\begin{aligned} F(Q) &= \int d\mathbf{J} d\alpha Q(\mathbf{J}, \alpha) \log \left[\frac{P(\mathbf{J}, \alpha, \mathbf{B})}{Q(\mathbf{J}, \alpha)} \right] \\ &= \log(P(\mathbf{B})) - \text{KL}[Q(\mathbf{J}, \alpha) \| P(\mathbf{J}, \alpha | \mathbf{B})]. \end{aligned} \quad (20)$$

Eq. (20) implies that the maximization of the free energy $F(Q)$ is equivalent to the minimization of the Kullback–Leibler distance (KL-distance) defined by

$$\text{KL}[Q(\mathbf{J}, \alpha) \| P(\mathbf{J}, \alpha | \mathbf{B})] = \int d\mathbf{J} d\alpha Q(\mathbf{J}, \alpha) \log(Q(\mathbf{J}, \alpha) / P(\mathbf{J}, \alpha | \mathbf{B})).$$

This measures the difference between the true joint posterior $P(\mathbf{J}, \alpha | \mathbf{B})$ and the trial distribution $Q(\mathbf{J}, \alpha)$. Since the KL-distance reaches its minimum at zero when the two distributions coincide, the joint posterior can be obtained by maximizing the free energy $F(Q)$ with respect to the trial distribution Q . In addition, the maximum free energy gives the log-marginal likelihood $\log(P(\mathbf{B}))$.

The optimization problem can be solved using a factorization approximation restricting the solution space⁴ (Attias, 1999; Sato, 2001):

$$Q(\mathbf{J}, \alpha) = Q_J(\mathbf{J})Q_\alpha(\alpha). \quad (21)$$

Under the factorization assumption (21), the free energy can be written as

$$\begin{aligned} F(Q) &= \langle \langle \log P(\mathbf{J}, \alpha, \mathbf{B}) \rangle \rangle_J \langle \rangle_\alpha - \langle \log Q_J(\mathbf{J}) \rangle_J - \langle \log Q_\alpha(\alpha) \rangle_\alpha \\ &= \langle \log P(\mathbf{B}|\mathbf{J}) \rangle_J - \text{KL}[Q_J(\mathbf{J})Q_\alpha(\alpha) \| P_0(\mathbf{J}|\alpha)P_0(\alpha)], \end{aligned} \quad (22)$$

where $\langle \cdot \rangle_J$ and $\langle \cdot \rangle_\alpha$ represent the expectation values with respect to $Q_J(\mathbf{J})$ and $Q_\alpha(\alpha)$, respectively. The first term in the second Eq. (22) corresponds to the negative sign of the expected reconstruction error. The second term (KL-distance) measures the difference between the prior and the posterior and corresponds to the effective

degree of freedom that can be well specified from the observed data. Therefore, (negative sign of) the free energy can be considered a regularized error function with a model complexity penalty term. The maximum free energy is obtained by alternately maximizing the free energy with respect to Q_J and Q_α . In the first step (J-step), the free energy $F(Q)$ is maximized with respect to Q_J while Q_α is fixed. The solution is given by

$$Q_J(\mathbf{J}) \propto \exp[\langle \log P(\mathbf{J}, \alpha, \mathbf{B}) \rangle_\alpha]. \quad (23)$$

In the second step (α -step), the free energy $F(Q)$ is maximized with respect to Q_α while Q_J is fixed. The solution is given by

$$Q_\alpha(\alpha) \propto \exp[\langle \log P(\mathbf{J}, \alpha, \mathbf{B}) \rangle_J]. \quad (24)$$

The above J- and α -steps are repeated until the free energy converges.

B.2. Hierarchical prior with smoothness constraints

We assume the following probabilistic model for the source currents:

$$P(\mathbf{B}|\mathbf{J}, \beta) \propto \exp \left[-\frac{1}{2} \beta (\mathbf{B} - \mathbf{G} \cdot \mathbf{J})' \cdot \Sigma_G \cdot (\mathbf{B} - \mathbf{G} \cdot \mathbf{J}) \right], \quad (25)$$

where $(\beta \Sigma_G)^{-1}$ denotes the covariance matrix of the sensor noise. Σ_G^{-1} is the normalized covariance matrix satisfying $\text{Tr}(\Sigma_G^{-1}) = M$, and β^{-1} is the average noise variance.

The direct application of the VB method to Eqs. (11) and (12) cannot be solved in a closed form because the inverse covariance matrix Σ_α is a nonlinear function of α and λ . To get the closed form solution, we rewrite the smoothness prior by introducing an auxiliary internal variable $\mathbf{Z}(t) = \{Z_n(t) | n = 1:N\}$:

$$\begin{aligned} P_0(\mathbf{J}_{1:T} | \mathbf{Z}_{1:T}, \alpha, \beta) & \\ & \propto \exp \left[-\frac{1}{2} \beta \sum_{t=1}^T (\mathbf{J}(t) - \mathbf{W} \cdot \mathbf{Z}(t))' \cdot \mathbf{A} \cdot (\mathbf{J}(t) - \mathbf{W} \cdot \mathbf{Z}(t)) \right], \\ P_0(\mathbf{Z}_{1:T} | \lambda, \beta) & \propto \exp \left[-\frac{1}{2} \beta \sum_{t=1}^T \mathbf{Z}'(t) \cdot \mathbf{A} \cdot \mathbf{Z}(t) \right]. \end{aligned} \quad (26)$$

The inverse noise variance β is also introduced in Eq. (26) as a coefficient of \mathbf{A} and \mathbf{A} to get the closed form solution. Therefore, the definitions of α and λ in this Appendix are different from those in the text by a factor β . Integration over \mathbf{Z} in the prior (26) can be accomplished as

$$\begin{aligned} P_0(\mathbf{J}_{1:T} | \alpha, \beta, \lambda) &= \int d\mathbf{Z}_{1:T} P_0(\mathbf{J}_{1:T} | \mathbf{Z}_{1:T}, \alpha, \beta) P_0(\mathbf{Z}_{1:T} | \lambda, \beta) \\ & \propto \exp \left[-\frac{1}{2} \beta \sum_{t=1}^T \mathbf{J}'(t) \cdot \Sigma_\alpha \cdot \mathbf{J}(t) \right]. \end{aligned}$$

Namely, the prior (26) is equivalent to the smoothness prior (11) and (12). Parameters α and λ are estimated by introducing the ARD hierarchical prior,

$$P_0(\alpha) = \prod_{n=1}^N \Gamma(\alpha_n | \bar{\alpha}_{0n}, \gamma_{0\alpha n}), \quad P_0(\lambda) = \prod_{n=1}^N \Gamma(\lambda_n | \bar{\lambda}_{0n}, \gamma_{0\lambda n}).$$

⁴ The VB method with this factorization assumption gives more accurate approximation than the mean field approximation (MFA), which is widely used in the field of statistical physics (Parisi, 1988), because MFA has a stronger factorization assumption, that is, that all of the variable components can be factorized.

The inverse noise variance parameter β is also estimated by assuming the non-informative prior:

$$P_0(\beta) = \beta^{-1}.$$

B.3. Free energy

To calculate the joint posterior $P(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda} | \mathbf{B}_{1:T})$, the free energy for the trial distribution $Q(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda})$ is defined as

$$F(Q) = \int d\mathbf{J}_{1:T} d\mathbf{Z}_{1:T} d\beta d\boldsymbol{\alpha} d\boldsymbol{\lambda} Q(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda}) \times \log \left[\frac{P(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda}, \mathbf{B}_{1:T})}{Q(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda})} \right], \quad (27)$$

$$P(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda}, \mathbf{B}_{1:T}) = P(\mathbf{B}_{1:T} | \mathbf{J}_{1:T}, \beta) \times P_0(\mathbf{J}_{1:T} | \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta) P_0(\mathbf{Z}_{1:T} | \boldsymbol{\lambda}, \beta) \times P_0(\beta) P_0(\boldsymbol{\alpha}) P_0(\boldsymbol{\lambda}).$$

In this case, we assume the factorization

$$Q(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \boldsymbol{\alpha}, \beta, \boldsymbol{\lambda}) = Q_J(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \beta) Q_\alpha(\boldsymbol{\alpha}, \boldsymbol{\lambda}).$$

The maximum free energy can be obtained by alternately maximizing the free energy $F(Q)$ with respect to Q_J and Q_α . In the first step (J-step), the free energy $F(Q)$ is maximized with respect to Q_J . The maximum solution of Q_J is a Normal-Gamma distribution:

$$Q_J(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T}, \beta) = Q_{J|\beta}(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T} | \beta) Q_\beta(\beta).$$

The conditional posterior distribution $Q_{J|\beta}(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T} | \beta)$ is the Normal distribution for $(\mathbf{J}_{1:T}, \mathbf{Z}_{1:T})$ with the noise variance parameter β^{-1} . The posterior distribution for β , $Q_\beta(\beta)$, is the Gamma distribution. The expectation values of these parameters are calculated in the J-step. In the second step (α -step), the free energy $F(Q)$ is maximized with respect to Q_α . The maximum solution of Q_α is the Gamma distribution for $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$. The expectation values of $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ are calculated in the α -step.

B.4. VB algorithm

The VB algorithm for the hierarchical prior with the smoothness constraint, (11) and (12), is summarized here (Fig. 1E). All of the quantities in the following equations represent the expectation value with respect to the posterior distribution. In the J-step (Fig. 1E), the inverse filter $\mathbf{L}(\boldsymbol{\Sigma}_\alpha^{-1})$ is calculated using the estimated covariance matrix $\boldsymbol{\Sigma}_\alpha^{-1}$ in the previous iteration:

$$\begin{aligned} \boldsymbol{\Sigma}_J &= \mathbf{G}' \cdot \boldsymbol{\Sigma}_G \cdot \mathbf{G} + \boldsymbol{\Sigma}_\alpha, \\ \boldsymbol{\Sigma}_B &= \mathbf{G} \cdot \boldsymbol{\Sigma}_\alpha^{-1} \cdot \mathbf{G}' + \boldsymbol{\Sigma}_G^{-1}, \\ \mathbf{L}(\boldsymbol{\Sigma}_\alpha^{-1}) &= \boldsymbol{\Sigma}_J^{-1} \cdot \mathbf{G}' \cdot \boldsymbol{\Sigma}_G = \boldsymbol{\Sigma}_\alpha^{-1} \cdot \mathbf{G}' \cdot \boldsymbol{\Sigma}_B^{-1}. \end{aligned} \quad (28)$$

The expectation values of the current $\mathbf{J}(t)$ and the noise variance β^{-1} are estimated using the inverse filter (28).

$$\begin{aligned} \mathbf{J}(t) &= \mathbf{L}(\boldsymbol{\Sigma}_\alpha^{-1}) \cdot \mathbf{B}(t), \\ \mathbf{Z}(t) &= \mathbf{A}^{-1} \cdot \mathbf{W}' \cdot \boldsymbol{\Sigma}_\alpha \cdot \mathbf{J}(t), \\ \boldsymbol{\Upsilon}_\beta &= \frac{1}{2} \text{NT}, \\ \boldsymbol{\Upsilon}_\beta \beta^{-1} &= \frac{1}{2} \sum_{t=1}^T \left[(\mathbf{B}(t) - \mathbf{G} \cdot \mathbf{J}(t))' \cdot \boldsymbol{\Sigma}_G \cdot (\mathbf{B}(t) - \mathbf{G} \cdot \mathbf{J}(t)) \right. \\ &\quad \left. + (\mathbf{J}(t) - \mathbf{W} \cdot \mathbf{Z}(t))' \cdot \mathbf{A} \cdot (\mathbf{J}(t) - \mathbf{W} \cdot \mathbf{Z}(t)) \right. \\ &\quad \left. + \mathbf{Z}'(t) \cdot \mathbf{A} \cdot \mathbf{Z}(t) \right]. \end{aligned} \quad (29)$$

In the α -step (Fig. 1E), the expectation values of the variance parameters are estimated as

$$\begin{aligned} \gamma_{nz} &= \gamma_{0nz} + \frac{T}{2}, \quad \gamma_{n\lambda} = \gamma_{0n\lambda} + \frac{T}{2}, \\ \gamma_{nz} \alpha_n^{-1} &= \gamma_{0nz} \alpha_n^{-1} + \frac{T}{2} \alpha_n^{-1} \left[1 - (\mathbf{A}^{-1} \cdot \mathbf{G}' \cdot \boldsymbol{\Sigma}_B^{-1} \cdot \mathbf{G})_{n,n} \right] \\ &\quad + \frac{1}{2} \beta \sum_{t=1}^T (\mathbf{J}(t) - \mathbf{W} \cdot \mathbf{Z}(t))_n^2, \\ \gamma_{n\lambda} \lambda_n^{-1} &= \gamma_{0n\lambda} \lambda_n^{-1} + \frac{T}{2} \lambda_n^{-1} \\ &\quad \times \left[1 - (\mathbf{A}^{-1} \cdot \mathbf{W}' \cdot \mathbf{G}' \cdot \boldsymbol{\Sigma}_B^{-1} \cdot \mathbf{G} \cdot \mathbf{W})_{n,n} \right] \\ &\quad + \frac{1}{2} \beta \sum_{t=1}^T Z_n^2(t). \end{aligned} \quad (30)$$

The covariance matrix $\boldsymbol{\Sigma}_\alpha^{-1}$ is then recalculated as

$$\boldsymbol{\Sigma}_\alpha^{-1} = \mathbf{A}^{-1} + \mathbf{W} \cdot \mathbf{A}^{-1} \cdot \mathbf{W}'. \quad (31)$$

The above J- and α -steps are repeated until the free energy converges. In the simulations, we further imposed a constraint $\boldsymbol{\lambda} = \kappa \boldsymbol{\alpha}$, which implies that the correlation in the current activities is proportional to the current variance, to reduce the number of unknown parameters. Under this constraint, the expectation values of $\boldsymbol{\alpha}$, κ , and $\boldsymbol{\lambda}$ are calculated in the α -step by applying the following update rule after the update of (30):

$$\begin{aligned} \alpha_n^{-1} &= (\gamma_{nz} \alpha_n^{-1} + \gamma_{n\lambda} \lambda_n^{-1} \kappa) / (\gamma_{nz} + \gamma_{n\lambda}), \\ \kappa^{-1} &= \frac{1}{N} \sum_{n=1}^N \alpha_n / \lambda_n, \\ \lambda_n &= \kappa \alpha_n. \end{aligned}$$

fMRI information was imposed as

$$\alpha_{0n}^{-1} = (a_{\max} V_n^2 + a_{\min}) / G_{\text{ave}}^2 \quad (32)$$

where V_n denotes the fMRI signal for the n th current. It is normalized by the average of the lead field norm $G_{\text{ave}}^2 = \left(\sum_{n=1}^N \sum_{m=1}^M G_{m,n}^2 \right) / N$. The hyperparameters a_{max} and a_{min} control the maximum and minimum of the prior mean variance, respectively. The optimal value of the hyperparameters (a_{max} , a_{min}) can be determined by maximizing the free energy with respect to (a_{max} , a_{min}). (1) The free energy is maximized for a given MEG signal and a sampled value of (a_{max} , a_{min}). (2) Then, the optimal value of (a_{max} , a_{min}) that maximizes the free energies obtained for the sampled values in the first step is determined.

B.5. Posterior distribution for the source current

After the free energy converges, the posterior distribution for the current $\mathbf{J}_{1:T}$ is obtained by integrating the internal variable $\mathbf{Z}_{1:T}$ and the inverse noise variance parameter β :

$$Q(\mathbf{J}_{1:T}) = \frac{|\Sigma_{\mathbf{J}}|^{T/2}}{(2\pi\gamma_{\beta})^{NT/2}} \frac{\Gamma(\gamma_{\beta} + NT/2)}{\Gamma(\gamma_{\beta})} \times \left[1 + \frac{\tilde{\beta}}{2\gamma_{\beta}} \right. \\ \left. \times \sum_{t=1}^T (\mathbf{J}(t) - \tilde{\mathbf{J}}(t))' \cdot \Sigma_{\mathbf{J}} (\mathbf{J}(t) - \tilde{\mathbf{J}}(t)) \right]^{-(\gamma_{\beta} + NT/2)},$$

where the expectation value of the current and the inverse noise variance obtained in the above algorithm are denoted by $\tilde{\mathbf{J}}$ and $\tilde{\beta}$ to distinguish them from the variables \mathbf{J} and β , respectively. The distribution is the student t distribution. By using this posterior distribution, it is possible to calculate the confidence interval and t value statistics.

References

- Ahlfors, S.P., Simpson, G.V., Dale, A.M., Belliveau, J.W., Liu, A.K., Korvenoja, A., Virtanen, J., Huotilainen, M., Tootell, R.B.H., Aronen, H.J., Ilmoniemi, R.J., 1999. Spatiotemporal activity of a cortical network for processing visual motion revealed by MEG and fMRI. *J. Neurophysiol.* 82, 2545–2555.
- Aine, C., Huang, M., Stephen, J., Christner, R., 2000. Multistart algorithms for MEG empirical data analysis reliably characterize locations and time courses of multiple sources. *NeuroImage* 12, 159–172.
- Amari, S., 1998. Natural gradient works efficiently in learning. *Neural Comput.* 10, 251–276.
- Attias, H., 1999. Inferring parameters and structure of latent variable models by variational Bayes. *Proc. 15th Conference on Uncertainty in Artificial Intelligence*. Morgan-Kaufman, pp. 21–30.
- Baillet, S., Garnero, L., 1997. A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Trans. Biomed. Eng.* 44, 374–385.
- Bandettini, P.A., 2000. The temporal resolution of functional MRI. In: Moonen, C.T.W., Bandettini, P.A. (Eds.), *Functional MRI*. Springer, Berlin, pp. 205–220.
- Belliveau, J.W., Kennedy, D.N., McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., Rosen, B.R., 1991. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* 254, 716–719.
- Churchland, P.S., Sejnowski, T.J., 1988. Perspectives on cognitive neuroscience. *Science* 242, 741–745.
- Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J. Cogn. Neurosci.* 5, 162–176.
- Dale, A.M., Liu, A.K., Fischl, B.R., Buchner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67.
- Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., Shadlen, M.N., 1994. fMRI measurements of human visual cortex. *Nature* 369, 525.
- Engel, S.A., Glover, G.H., Wandell, B.A., 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex* 7, 181–192.
- Fujimaki, N., Hayakawa, T., Nielsen, M., Kosche, T.R., Miyauchi, S., 2002. An fMRI-constrained MEG source analysis with procedures for dividing and grouping. *NeuroImage* 17, 324–343.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. *Bayesian Data Analysis*. Chapman & Hall, London.
- Goebel, R., Max Planck Society, 2000. *Brain Voyager 2000*, Version 4.2. <http://www.brainvoyager.de>.
- Grave de Peralta Menendez, R., Gonzalez Andino, S., 1998. A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Trans. Biomed. Eng.* 45, 440–448.
- Hamalainen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42.
- Hamalainen, M.S., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Modern Phys.* 65, 413–497.
- Hari, R., 1991. On brain's magnetic responses to sensory stimuli. *J. Clin. Neurophysiol.* 8, 157–169.
- Heinze, H.J., Mangun, G.R., Burchert, W., Hinrichs, H., Scholz, M., Munte, T.F., Gos, A., Scherg, M., Johannes, S., Hundscheidt, H., Gazzaniga, M.S., Hillyard, S.A., 1994. Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature* 372, 543–546.
- Huang, M., Aaron, R., Shiffman, C.A., 1997. Maximum entropy method for magnetoencephalography. *IEEE Trans. Biomed. Eng.* 44, 98–102.
- Kajihara, S., Ohtani, Y., Goda, N., Tanigawa, M., Toyama, K., 2004. Wiener filter magnetoencephalography of visual cortical activities. *Brain Topogr.* (in press).
- Linden, D.E.J., Kallenbach, U., Heinecke, A., Singer, W., Goebel, R., 1999. The myth of upright vision. A psychophysical and functional imaging study of adaptation to inverting spectacles. *Perception* 28 (4), 469–481.
- Liu, A.K., Belliveau, J.W., Dale, A.M., 1998. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8945–8950.
- Logothetis, N.K., Pauls, J., Augth, M., Trinth, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157.
- Mosher, J.C., Lewis, P.S., Leahy, R.M., 1992. Multiple dipole modelling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* 39, 541–557.
- Neal, R.M., 1996. *Bayesian Learning for Neural Networks*. Springer-Verlag, New York.
- Nunez, P.L., 1981. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford Univ. Press, New York.
- Ogawa, S., Lee, T.-M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast-dependent oxygenation. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9868–9872.
- Parisi, G., 1988. *Statistical Field Theory*. Perseus Books.
- Pascual-Marqui, R.D., 1999. Review of methods for solving the EEG inverse problem. *Int. J. Bioelectromagn.* 1, 75–86.
- Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* 18, 49–65.
- Penny, W., Kiebel, S., Friston, K., 2003. Variational Bayesian inference for fMRI time series. *NeuroImage* 19, 727–741.
- Phillips, C., Rugg, M.D., Friston, K.J., 2002. Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. *NeuroImage* 16, 678–695.

- Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* 32, 11–22.
- Sato, M., 2001. On-line model selection based on the variational Bayes. *Neural Comput.* 13, 1649–1681.
- Schmidt, D.M., George, J.S., Wood, C.C., 1999. Bayesian inference applied to the electromagnetic inverse problem. *Hum. Brain Mapp.* 7, 195–212.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B.H., 1995. Borders of multiple visual areas in human revealed by functional magnetic resonance imaging. *Science* 268, 889–893.
- Toyama, K., Yoshikawa, K., Yoshida, Y., Kondo, S., Tomita, Y., Takanashi, Y., Ejima, Y., Yoshizawa, S., 1999. A new method for magnetoencephalography (MEG): three dimensional magnetometer-spatial filter system. *Neuroscience* 91, 405–415.
- Uutela, K., Hamalainen, M.S., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* 10, 173–180.
- Wang, J.-Z., Williamson, S., Kaufman, J., 1992. Magnetic source images determined by a lead-field analysis: the unique minimum-norm least-squares estimation. *IEEE Trans. Biomed. Eng.* 39, 665–675.