

## Cortical current source estimation from electroencephalography in combination with near-infrared spectroscopy as a hierarchical prior

Takatsugu Aihara <sup>a</sup>, Yusuke Takeda <sup>b</sup>, Kotaro Takeda <sup>a,c</sup>, Wataru Yasuda <sup>a</sup>, Takanori Sato <sup>a,d</sup>, Yohei Otaka <sup>e</sup>, Takashi Hanakawa <sup>f</sup>, Manabu Honda <sup>f</sup>, Meigen Liu <sup>g</sup>, Mitsuo Kawato <sup>a</sup>, Masa-aki Sato <sup>b,\*</sup>, Rieko Osu <sup>a</sup>

<sup>a</sup> ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan

<sup>b</sup> ATR Neural Information Analysis Laboratories, Kyoto 619-0288, Japan

<sup>c</sup> Clinical Research Center, National Hospital Organization Murayama Medical Center, Tokyo 208-0011, Japan

<sup>d</sup> Department of Electrical, Electronics and Information Engineering Faculty of Engineering, Nagaoka University of Technology, Niigata 940-2188, Japan

<sup>e</sup> Tokyo Bay Rehabilitation Hospital, Chiba 275-0026, Japan

<sup>f</sup> National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo 187-8502, Japan

<sup>g</sup> Department of Rehabilitation Medicine, Keio University, Tokyo 160-8582, Japan

### ARTICLE INFO

#### Article history:

Received 28 April 2011

Revised 26 September 2011

Accepted 30 September 2011

Available online 14 October 2011

#### Keywords:

Variational Bayesian Multimodal

EncephaloGraphy (VBMG)

Electroencephalography (EEG)

Near-infrared spectroscopy (NIRS)

### ABSTRACT

Previous simulation and experimental studies have demonstrated that the application of Variational Bayesian Multimodal EncephaloGraphy (VBMG) to magnetoencephalography (MEG) data can be used to estimate cortical currents with high spatio-temporal resolution, by incorporating functional magnetic resonance imaging (fMRI) activity as a hierarchical prior. However, the use of combined MEG and fMRI is restricted by the high costs involved, a lack of portability and high sensitivity to body-motion artifacts. One possible solution for overcoming these limitations is to use a combination of electroencephalography (EEG) and near-infrared spectroscopy (NIRS). This study therefore aimed to extend the possible applications of VBMG to include EEG data with NIRS activity as a hierarchical prior. Using computer simulations and real experimental data, we evaluated the performance of VBMG applied to EEG data under different conditions, including different numbers of EEG sensors and different prior information. The results suggest that VBMG with NIRS prior performs well, even with as few as 19 EEG sensors. These findings indicate the potential value of clinically applying VBMG using a combination of EEG and NIRS.

© 2011 Elsevier Inc. All rights reserved.

### Introduction

Magnetoencephalography (MEG) and electroencephalography (EEG) provide direct measures of the magnetic and electrical fields caused by neural activity, respectively, with high temporal resolution. The spatial resolution of these techniques, however, is relatively poor, because MEG/EEG signals measured on the scalp surface are comprised of a mixture of signals originating from a large number of cortical areas. Because many different source configurations can generate the same distribution of magnetic/electric fields on the scalp, it is necessary to solve an ill-posed inverse problem to improve the spatial resolution (Michel, et al., 2004). Inverse procedures are commonly classified as dipole and distributed source methods. Dipole methods (Hari, 1991; Mosher, et al., 1992) solve the inverse problem by approximating brain activity using a small number of current dipoles. Although the methods provide good estimates 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 dipole methods. On the other hand, distributed methods assume a large number of current dipoles distributed in the brain (Hamalainen, et al., 1993). In a linear approach to resolving the inverse problem, several prior assumptions can be used such as the minimum norm method. Unfortunately, the prior assumptions are insufficient to fully resolve the ill-posed nature of the inverse problem, meaning that the spatial resolution of these methods is still rather low. Attempts have been made to overcome these limitations using functional information obtained with other imaging modalities, such as functional magnetic resonance imaging (fMRI) (for a review of EEG-fMRI integration, see Rosa, et al., 2010). These 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 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. However, source current estimation based on these methods may fail if the fMRI data contain incorrect information. In recent years, there are several attempts to overcome these difficulties (Daunizeau, et al., 2005; Grova, et al., 2008; Henson, et al., 2010).

\* Corresponding author at: 2-2-2 Hikaridai, Keihanna Science City, Kyoto 619-0288, Japan. Fax: +81 774 95 1259.

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

We previously proposed Variational Bayesian Multimodal EncephaloGraphy (VBMG) as a hierarchical Bayesian estimation method that incorporates fMRI activity information as a hierarchical prior to overcome the vulnerability to inaccuracies in fMRI information (Sato et al., 2004). In this method, the variance of the source current at each source location is considered to be an unknown parameter and is 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, such that it places a soft constraint on the variance. Combined MEG and fMRI measurements confirmed that VBMG could be used to estimate brain activity with high spatio-temporal resolution in both computer simulations (Sato et al., 2004) and real experiments involving visual (Yoshioka et al., 2008) and motor tasks (Toda et al., 2011).

We have been postulating that VBMG with a combination of MEG and fMRI can make a substantial contribution to neuroscience research. However, studies combining MEG and fMRI are limited by the high sensitivity of these methods to body-motion artifacts, a lack of portability, and high costs. Some or all of these factors reduce the suitability of this combined method for measuring brain activity in infants or patients, for investigating brain activity related to human motor control in realistic situations (e.g., in sitting or standing positions), and for daily use, e.g., monitoring day-to-day changes in brain activities during stroke recovery or motor learning. One possible solution for overcoming these limitations is to use a combination of EEG and NIRS, rather than MEG and fMRI. The reasoning behind this approach is that both EEG and MEG record brain electrical activity, whereas fMRI and NIRS measure blood-flow-related signals in the brain. EEG and NIRS have lower sensitivity to body-motion artifacts, higher portability and lower costs, compared to MEG/fMRI. We therefore applied VBMG to EEG data by incorporating NIRS activity as a hierarchical prior.

In the present study, we initially conducted computer simulations to investigate the performance of VBMG applied to EEG data. VBMG was applied to simulated EEG data with different types of prior information (no prior, correct prior and incorrect prior) to investigate the effects of prior information on estimation accuracy. The effects of the resolution of prior information on estimation accuracy were also investigated because the different imaging modalities, providing prior information, have different spatial resolutions. In addition, the effects of the number of EEG sensors on estimation accuracy were investigated, because the use of fewer sensors is advantageous from the subject's point of view, particularly for infants and patients.

We then conducted real experiments in subjects performing a motor task and applied VBMG to real EEG data to confirm the results of the computer simulations. We compared the performance of VBMG with fMRI prior, NIRS prior and no prior, and investigated the effects of the number of EEG sensors for both fMRI and NIRS priors.

A motor task was used for the following reasons. First, brain activity associated with movement, especially with the preparation and execution of movements, has been well studied. Regarding the spatial patterns of motor-related brain activity, the existence of laterality and somatotopy in the sensorimotor area is well known; for example, right finger movements activate the hand area of the contralateral (i.e., left) sensorimotor area. Regarding the temporal pattern of brain activity for voluntary movements, movement-related potentials (MRPs) recorded using EEG are known to be divided into three sub-components: (1) the Bereitschaftspotential (BP), a slowly rising negative component starting 1000–1500 ms before electromyography (EMG) onset; (2) the negative slope (NS), a steeper increase in negativity starting around 400 ms before EMG onset; and (3) the motor potential (MP), a further increase in negativity appearing around EMG onset and peaking just after EMG onset. The earliest component, the BP, shows widespread and symmetrical scalp distribution, suggesting that it may reflect cortical activity related to early movement

preparation, arising predominantly from the supplementary motor area (SMA). In contrast, the later components, the NS and MP, show clear lateralization towards the hemisphere contralateral to the movement, suggesting that the NS and MP reflect cortical activity associated with late preparation and execution, arising predominantly from the contralateral primary motor area (M1) (Cunnington et al., 1996). Second, as described previously, we propose that VBMG with a combination of EEG and NIRS can provide a suitable approach for monitoring day-to-day changes in brain activity during recovery of motor function after stroke. One of the aims of the present study was therefore to lay the foundations for such future studies.

The motor task in the present study was a repetitive right-finger movement task. The movement repetition rate was slow (less than 0.3 Hz) in the EEG experiments, which were conducted using an event-related design, and rapid (as fast as possible) in both the fMRI and NIRS experiments, which were conducted using a block design. Previous studies have reported that the primary sensorimotor areas (SM1) and SMA are activated during slow-rate movements, reflecting not only the execution and somatosensory information processing stages, but also the preparation stage, whereas SM1 is predominantly activated during fast-rate movements, mainly reflecting the late preparation, execution and somatosensory information processing stages (Kunieda et al., 2000; Toma et al., 1999). Brain activity of the late preparation, execution and somatosensory information processing stages (i.e., SM1 activity) is thus commonly observed during both movements. The present study therefore focused on these stages.

## Materials and methods

### Computer simulations

Computer simulations were conducted to investigate the performance of VBMG applied to EEG data.

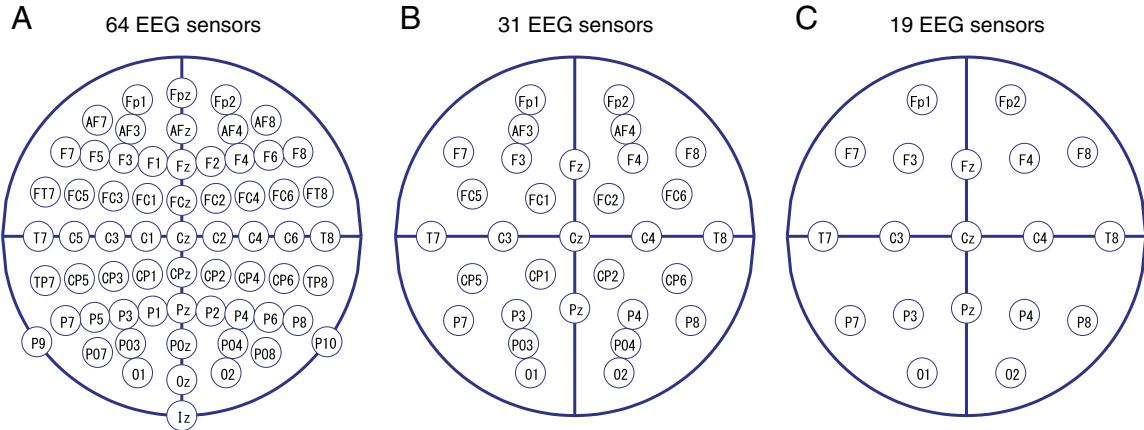
### Simulated EEG data

A polygon model of cortical surfaces (20,004 vertex points) was constructed from a T1 structural image (for MRI data, see [MRI experiments and analysis](#)) of an experimental subject (RO) using FreeSurfer software (Dale et al., 1999). A single current dipole was assumed at each vertex point perpendicular to the cortical surface. The brain structures were approximated as a three-layer model by identifying three boundaries, for the cerebrospinal fluid (CSF), skull and scalp. These surfaces obtained by FreeSurfer were slightly modified using gray/white/CSF segmentation by SPM2 (Wellcome Department of Cognitive Neurology, UK) and morphological operations. This MRI-based three-layer model was used in the following simulations to calculate a lead field, which converts cortical currents to EEG data. We assumed that the conductivities of the CSF, skull and scalp were 0.33, 0.0042, and 0.33 S/m (Waberski, et al., 1998), respectively.

We assumed two single dipole sources ( $S_1$  and  $S_2$ ), both located at a cortical vertex that was randomly selected from 20,004 vertices. The waveform of every source current was the product of a sinusoidal wave (4-Hz for  $S_1$  and 10-Hz for  $S_2$ ) and Hanning window. Both sources became active at 0 ms, peaked at 200 ms (peak amplitudes 10-nAm), and returned to the baseline level at 400 ms. We tested 500 different pairs of  $S_1$  and  $S_2$ .

We assumed that every cortical dipole exhibits background activity, which is assumed to be independent among dipoles. The background activity was assumed to be a Gaussian white noise with zero mean and standard deviation (SD) 0.2-nAm (1/50 of the peak amplitude of the source current).

We assumed that the head was surrounded by 64, 31, or 19 EEG sensors ([Fig. 1](#)) whose positions were determined based on real experimental sensor-position data from the subject (for the position of EEG sensors, see [EEG experiments and preprocessing](#)). Simulated



**Fig. 1.** Placement of EEG sensors according to the International 10–20 system. (A) 64 sensors, (B) 31 sensors, (C) 19 sensors. The positions of these sensors in the simulations were determined on the basis of real experimental data.

EEG data were generated using the forward model (see Eq. (1) in Appendix A). We assumed that every EEG sensor has sensor noise, which was assumed to be independent among sensors. The sensor noise was assumed to be a Gaussian white noise with zero mean and SD of 1.5  $\mu\text{V}$ . Signal to noise ratio (SNR) can be defined as:

$$\text{SNR} = S / \sqrt{N_{bg}^2 + N_{sn}^2},$$

where  $S$  is the maximum amplitude of EEG originated from cortical source activities,  $N_{bg}$  is the variance of EEG originated from cortical background activities, and  $N_{sn}$  is the variance of sensor noise. In the above settings of source current amplitude, background activity intensity, and sensor noise level, SNR was 0.81, averaged over 500 source pairs. This value is similar to the SNR obtained in the experimental data described below.

#### Variational Bayesian Multimodal EncephaloGraphy (VBMEG)

In VBMEG, the prior current variance is considered as a random variable and a hierarchical prior distribution for the prior current variance is introduced (see Appendix A for details of estimation). The activity information, such as that derived from fMRI/NIRS, is imposed not on the current variance itself, but on the hierarchical prior distribution for the current variance through two hyper-parameters: the variance magnification parameter ( $m_0 \geq 1$ ) and the confidence parameter ( $\gamma_0 \geq 0$ ). The magnification parameter ( $m_0$ ) controls the relative amplitude of the prior current variance; large  $m_0$  increases the prior current variance, especially for dipoles with large activities (i.e., large  $t$ -values in the case of fMRI prior) and encourages the estimated current amplitude to increase. The  $m_0$  can be theoretically derived if both the source and background currents are known. In the setting described in Simulated EEG data,  $m_0$  is 674 in theory. The confidence parameter ( $\gamma_0$ ) controls the width of the prior distribution and the variance of the prior distribution is inversely proportional to  $\gamma_0$ ; large  $\gamma_0$  narrows the prior distribution around the mean value, and the estimation depends more critically on the activity data. Thus, the activity information can give a soft constraint on the current variance (for details of VBMEG, see Sato et al., 2004; Yoshioka et al., 2008).

Related work in multimodal fusion (e.g., Grova et al., 2008) has reported that a major issue arises in imposing prior constraints on the EEG/MEG inverse problem when the prior information<sup>1</sup> is ‘wrong’ or distorted, because of false positive or missing sources. In addition, VBMEG is sensitive to the values of hyper-parameters

(Yoshioka et al., 2008). Therefore, the first step of the simulations was to investigate the effects of the types of prior information, the values of hyper-parameters, and their interactions on the estimation accuracy, with both the resolution of the prior information and the number of EEG sensors fixed. We used four different types of prior information; no prior, missing prior, correct prior, and false positive prior. The ‘no prior’<sup>2</sup> condition indicates that  $V_n = 1$  for all dipoles (i.e., uniform spatial prior), where  $V_n$  is the prior value at the  $n$ -th vertex. The ‘correct prior’ condition indicates that  $V_n = 1$  for dipoles within  $R$ -mm radius around both source dipoles (i.e.,  $S_1$  and  $S_2$ ), and  $V_n = 0$  for other dipoles. The ‘false positive prior’ condition indicates that  $V_n = 1$  for dipoles within  $R$ -mm radius around  $S_1$ ,  $S_2$  and a randomly selected dipole, and  $V_n = 0$  for other dipoles. The ‘missing prior’ condition indicates that  $V_n = 1$  for dipoles within  $R$ -mm radius around only  $S_1$ , and  $V_n = 0$  for other dipoles. This imposes false negative information on the active source,  $S_2$ . The value of  $R$  corresponds to the spatial resolution of the prior information; the smaller the  $R$  value, the higher the spatial resolution. Three different values were used for each hyper-parameter:  $m_0 = 1, 100, 1000$ , and  $\gamma_0 = 1, 10, 1000$ . Thus, nine combinations of hyper-parameters were tested. Here, the resolution of the prior information,  $R$ , was fixed to six, roughly corresponding to the resolution of the fMRI scanning. The number of EEG sensors was fixed to the maximum value of 64.

In the second step, we investigated the effects of spatial resolution of prior information, the number of EEG sensors, and their interactions on estimation accuracy. The following values of  $R$  were used:  $R = 6, 12$ , and  $18$ . The larger value of  $R$  may correspond to the case of NIRS prior, because the spatial resolution of NIRS is worse than that of fMRI. The numbers of EEG sensors were 64, 31, and 19. Here, hyper-parameters were fixed to the following values:  $m_0 = 100$  and  $\gamma_0 = 10$ , according to the result of the first step of the present simulation study. The type of prior was ‘correct prior’.

#### Evaluation of estimation accuracy

In the first step of the simulation, we used four evaluation metrics to access the performance of VBMEG.

The first metric, the area under curve (AUC), is based on the receiver operating characteristic (ROC) curve approach (Daunizeau, et al., 2005; Grova et al., 2006, 2008; Ou et al., 2010). For each combination of true sources, the AUC was calculated in the following way. Normalized energy for the estimated current distribution for the  $k$ -th dipole was defined as:  $E_{\text{estimated}}(k) = \|Z_{\text{estimated}}^k\|^2 / \max_k (\|Z_{\text{estimated}}^k\|^2)$ , where

<sup>1</sup> Throughout this paper, ‘prior information’ refers to prior functional or activity information, because we were only interested in the effects of functional (activity) information. Note that VBMEG incorporates both structural and functional (activity) information as priors.

<sup>2</sup> Throughout this paper, ‘no prior’ indicates that no functional MRI/NIRS information was used as a hierarchical prior. Note that structural MRI information, however, was incorporated as a prior in these cases.

$Z_{\text{estimated}}^k$  is the estimated current averaged around peak time (195–205 ms; it should be noted that both  $S_1$  and  $S_2$  reach peak at 200 ms) of the  $k$ -th dipole. On the other hand, normalized energy for the true current distribution for the  $k$ -th dipole,  $E_{\text{true}}(k)$ , was defined as:  $E_{\text{true}}(k)=1$  if the distance between the  $k$ -th dipole and  $S_1$  or  $S_2$  was less than  $R$  mm (it should be noted that  $R$  is the resolution of prior information), and  $E_{\text{true}}(k)=0$  otherwise. For a threshold  $\beta$  chosen in the interval [0, 1] (to be precise,  $\beta=1.2^{-200}, 1.2^{-199}, \dots, 1.2^{-2}, 1.2^{-1}$ ), we considered the  $k$ -th dipole to be active if its energy  $E_{\text{estimated}}(k)\geq\beta$ . By comparing the estimated energy,  $E_{\text{estimated}}(k)$ , with the true energy,  $E_{\text{true}}(k)$ , for each dipole, we were able to quantify the amount of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) for each threshold  $\beta$ . Sensitivity and specificity were then estimated as follows:

$$\text{sensitivity}(\beta)=\frac{\text{TP}(\beta)}{\text{TP}(\beta)+\text{FN}(\beta)}, \text{ and } \text{specificity}(\beta)=\frac{\text{TN}(\beta)}{\text{TN}(\beta)+\text{FP}(\beta)}.$$

ROC curves were then obtained by plotting sensitivity ( $\beta$ ) against (1-specificity( $\beta$ )) for different thresholds  $\beta$ . The area under the ROC curve (AUC), a well-known criterion to assess detection accuracy, was then calculated according to the trapezium rule. An AUC value of 1 indicates the perfect detection performance and lower values indicate lower performance. Potential biases of specificity estimation (due to the imbalance between the active and the inactive dipoles) were corrected as in Grova et al. (2006) (see Appendix B for the method of the bias correction). The AUCs for  $S_1$  and  $S_2$  were averaged across 500 combinations of sources.

The second metric was the correlation coefficient ( $R_{\text{st}}$ ) between spatio-temporal patterns of the true and estimated currents, calculated across the whole activated duration (400 ms):

$$\overline{\langle Z_{\text{true}} \rangle} = \frac{1}{T} \sum_t \frac{1}{K} \sum_k Z_{\text{true}}^k(t), \quad \overline{\langle Z_{\text{estimated}} \rangle} = \frac{1}{T} \sum_t \frac{1}{K} \sum_k Z_{\text{estimated}}^k(t), \\ R_{\text{st}} = \frac{\sum_t \sum_k \left\{ Z_{\text{true}}^k(t) - \overline{\langle Z_{\text{true}} \rangle} \right\} \cdot \left\{ Z_{\text{estimated}}^k(t) - \overline{\langle Z_{\text{estimated}} \rangle} \right\}}{\sqrt{\sum_t \sum_k \left\{ Z_{\text{true}}^k(t) - \overline{\langle Z_{\text{true}} \rangle} \right\}^2} \cdot \sqrt{\sum_t \sum_k \left\{ Z_{\text{estimated}}^k(t) - \overline{\langle Z_{\text{estimated}} \rangle} \right\}^2}},$$

where  $T$  is the number of time points in the analysis window (400 ms in this case),  $K$  is the number of cortical dipoles (20,004 in this case),  $Z_{\text{true}}^k(t)$  is the true current at the  $k$ -th dipole and time  $t$ , and  $Z_{\text{estimated}}^k(t)$  is the estimated current at the  $k$ -th dipole and time  $t$ . This metric tests whether the spatio-temporal pattern of the estimated currents is similar to that of the true currents.

The third metric was the correlation coefficient ( $R_s$ ) between spatial patterns of the true and estimated currents averaged around peak time (195–205 ms):

$$\overline{Z_{\text{true}}} = \frac{1}{K} \sum_k Z_{\text{true}}^k, \quad \overline{Z_{\text{estimated}}} = \frac{1}{K} \sum_k Z_{\text{estimated}}^k, \\ R_s = \frac{\sum_k \left( Z_{\text{true}}^k - \overline{Z_{\text{true}}} \right) \cdot \left( Z_{\text{estimated}}^k - \overline{Z_{\text{estimated}}} \right)}{\sqrt{\sum_k \left( Z_{\text{true}}^k - \overline{Z_{\text{true}}} \right)^2} \cdot \sqrt{\sum_k \left( Z_{\text{estimated}}^k - \overline{Z_{\text{estimated}}} \right)^2}},$$

where  $Z_{\text{true}}^k$  and  $Z_{\text{estimated}}^k$  are the temporal averages (around peak time; 195–205 ms) of the true and estimated currents, respectively. This metric shows whether the spatial pattern of the estimated currents is similar to that of the true currents.

The fourth metric was the root mean squared error (RMSE) to assess the accuracy of estimation of the current amplitude over the all dipoles (Daunizeau et al., 2005):

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_t \frac{1}{K} \sum_k (Z_{\text{true}}^k(t) - Z_{\text{estimated}}^k(t))^2}.$$

This was also calculated across the whole activated duration (400 ms).

In the second step of the simulation, we applied the four metrics used in the first step. In addition, we used localization error and estimation gain as two other metrics. The former reflects the precision of the source localization and the latter reflects the accuracy of the amplitude estimation at the source position. For each dipole, amplitude was averaged around peak time (195–205 ms). Localization error was defined as the distance between true and estimated sources, where the estimated source was the dipole that had the maximum of the averaged amplitude. For both true and estimated sources, we calculated the sum of the averaged amplitudes from dipoles within a 6 mm radius around the source. Estimation gain was defined as the ratio of the sum of the averaged amplitudes between true and estimated sources.

#### Application of VBMG to real EEG data

We conducted experiments to confirm whether the findings obtained in the simulation study were applicable to real experimental data. We investigated performance of VBMG with different types of prior information (fMRI prior, NIRS prior and no prior) and different numbers of EEG sensors (64, 31 and 19).

#### Subjects

Two subjects (TA and RO; aged 32–41, one male and one female) took part in the fMRI, NIRS and EEG experiments. Both subjects gave written informed consent for the experimental procedures, which were approved by the ATR Human Subject Review Committee.

#### MRI experiments and analysis

Structural and functional MRI data were recorded using a 3T Magnetom Trio MRI scanner (Siemens, Munich, Germany). The acquisition parameters for T1-weighted images were as follows: repetition time 2 s, time of echo 4.38 ms, flip angle 8°, slice thickness 1 mm, field of view 256×256 mm, imaging matrix 256×256 and 208 slices. The acquisition parameters for echo-planar images (EPs) were as follows: repetition time 3 s, time of echo 30 ms, flip angle 60°, slice thickness 3 mm, field of view 192×192 mm, imaging matrix 64×64 mm and 44 slices.

The fMRI experiments consisted of alternating blocks of rest and task periods (18 s each). During a rest period, the word 'rest' was presented at the center of the display and subjects remained quiet. During a task period, subjects conducted repetitive movements of the right index finger (brisk extension and flexion at the metacarpophalangeal joints) as fast as possible, according to the instruction 'right index finger' presented at the center of the display. There were eight (subject RO) or nine (subject TA) blocks.

A polygon cerebral cortex model was constructed for each subject using FreeSurfer software (Dale et al., 1999) from the T1 structural image of the subject. The number of vertices for each subject was 20,004. A single current dipole was assumed at each vertex point perpendicular to the cortical surface. For the calculation of the lead field, the brain structures were approximated as a three-layer model by identifying three boundaries, for CSF, skull and scalp. These surfaces obtained by FreeSurfer were slightly modified using gray/white/CSF segmentation by SPM2 and morphological operations. This MRI-based three-layer model was used as a forward model, describing the relationship between cortical currents and EEG data, in the following analysis for source current estimation from EEG using VBMG. The conductivities of the CSF, skull and scalp used in the analysis of real data were the same as those used in the simulations.

fMRI signals elicited by the movements were analyzed using SPM2 to obtain prior information on the variances of the cortical currents. Head motion was corrected and the images were smoothed using an 8-mm full-width at half maximum (FWHM) Gaussian filter. The

voxel *t*-values were calculated for the difference between the task and rest periods. The resulting *t*-values were thresholded ( $p < 0.001$ , uncorrected for multiple comparisons), normalized to have a maximum value of 1, then mapped to the cortical vertices for the prior information of source current estimation, as described previously (Fujiwara et al., 2009; Toda et al., 2011; Yoshioka et al., 2008).

#### NIRS experiments and analysis

Subjects were seated in a comfortable reclining armchair. NIRS recording was performed using an optical imaging system (FOIRE-3000, Shimadzu Co., Japan). Bilateral SM1, bilateral SMA, and bilateral dorsal premotor area (PMA) were covered with 16 emitters and 12 detectors which provided 43-channel recording of changes in the concentrations of oxygenated hemoglobin and deoxygenated hemoglobin ( $\Delta[\text{OxyHb}]$  and  $\Delta[\text{DeoxyHb}]$ , respectively). For the 43-channel recording, each distance between one emitter and one detector was set at 3 cm (long distance channel: long-ch). To reduce artifacts related to skin blood flow, four short-distance (1.5 cm) source-detector channels (short distance channel: short-ch) were also placed onto the bilateral frontal and parietal lobes (Fig. 2). It should be noted that signals measured from short distance channels mainly originated from surface tissues (Okada et al., 1997).

The NIRS experiments consisted of alternating blocks of rest–task–rest periods (15 s each). The start of the task period was indicated by a single click sound, while the end was indicated by two clicks. During a rest period, subjects were presented with a fixation point and remained quiet. During a task period, subjects performed repetitive movements of the right index finger (the same as those in the fMRI experiment) as rapidly as possible. Each session consisted of six blocks.

After the NIRS experiment, probe positions were measured using a stylus marker (FASTRAK; Polhemus, U.S.A.) to allow subsequent co-registration of EEG and NIRS results.

Three near-infrared beams (wavelengths 780, 805, and 830 nm) were irradiated and detection beams sampled every 130 ms were used to calculate  $\Delta[\text{OxyHb}]$  and  $\Delta[\text{DeoxyHb}]$ . The present study focused on  $\Delta[\text{OxyHb}]$  because it is a more sensitive parameter of cerebral blood flow (Hoshi et al., 2001; Strangman et al., 2002). In each long-distance recording channel (43 channels), the time-series data for  $\Delta[\text{OxyHb}]$  were smoothed using a Gaussian kernel of 4 s FWHM. The baseline trend for  $\Delta[\text{OxyHb}]$  in each task block was then removed using linear-fitting to the signals of baseline intervals (10 s rest period just before the task onset and 5 s rest period after 10 s of the task offset). To normalize the amplitude of  $\Delta[\text{OxyHb}]$ , the smoothed and trend-reduced data were transformed to z-score values using the SD

during rest periods in the rest blocks (10 s before the task and 5 s after 10 s of the task).

The following process was used to remove the skin blood flow artifacts from long-ch data. The four short-ch data were normalized to z-scores using the same process as for long-ch. Principal component analysis (PCA) was used to estimate the skin blood flow artifact. In this study, we considered the first component to represent this artifact. We computed multiple regression coefficients on the basis of a model in which data from long-ch was modeled as a linear sum of hemodynamic response function, skin blood flow artifact (the first principal component) and a constant. We then subtracted the artifact and bias components from the long-ch data to produce artifact-free  $\Delta[\text{OxyHb}]$  data.

For the long-ch data, the peak value of the artifact-free  $\Delta[\text{OxyHb}]$  during the task period was regarded as NIRS activity. The NIRS activity was normalized by dividing the peak value in each channel by the maximum value of the 43 channels. The normalized NIRS activity was mapped onto the cerebral cortex using Fusion software (Shimadzu Co., Japan) for the prior information of VBMG, according to the following method. First, the position of the every probe was superimposed on the model of the scalp, and the position of every NIRS channel was estimated as the center point between the corresponding emitter and detector. Second, the cortical projection point of each NIRS channel position was determined according to the method described by Okamoto et al. (2004) (see Appendix C). Third, the normalized activity data for each NIRS channel was applied to cortical vertices using the interpolation method described in Takeuchi et al. (2009) (see Appendix D). Note that vertices more than 2 cm away from all of the cortical projection points had no NIRS data.

#### EEG experiments and preprocessing

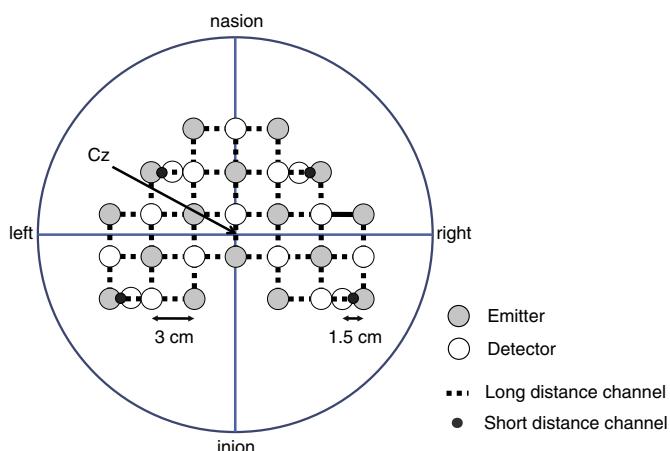
Subjects were seated in a comfortable reclining armchair and performed brisk, self-initiated movements (extension and flexion) of the right index finger (more than about 3-s inter-movement interval). They completed six sessions, each consisting of 20 movements, resulting in a total of 120 movements. Subjects took brief (about 1 min) rests between sessions.

EEG was recorded at a sampling rate of 1024 Hz with a 64-electrode cap, referenced to the Common Mode Sense (CMS) active electrode. A Biosemi Active Two system was used for amplification and analog-to-digital conversion of the EEG signals. Electrooculography (EOG) data were also collected to control for eye movements.

Movement onset<sup>3</sup> was detected using an optical sensor (GP2Y0A21YK; SHARP, Japan); the subject's finger blocked the light in the resting state and unblocked it when he or she lifted the finger. The electrical output of the optical sensor unit was sent to the EEG system for signal averaging. Data from a later additional experiment revealed that movement onset detected by the optical sensor was delayed by  $134 \pm 23$  ms relative to that detected by EMG in subject TA.

EEG sensor positions were measured before the EEG experiment using the stylus marker to allow subsequent co-registration of EEG, fMRI and NIRS results.

EEG signals were first passed through a band-pass filter (0.2–100 Hz) and a notch filter (50 Hz), and down-sampled to 250 Hz. Note that BP was eliminated using a high-pass filter of 0.2 Hz, because it reflects the early preparation stage and was therefore beyond the scope of this study. EEG signals between 1500 ms before and 200 ms after movement onset detected by the optical sensor were then extracted and baseline-corrected using the first 500 ms as a baseline. Finally, outlier trials and sensors were excluded from the analysis according to the following steps: (1) determining a threshold so that 95% of all absolute EEG signals are below the threshold, (2) excluding trials where absolute EEG signals



**Fig. 2.** NIRS probe positions. Emitters and detectors are indicated by gray and white circles, respectively. Long- and short-distance channels are shown by dotted lines and black points, respectively.

<sup>3</sup> In the present paper, 'movement onset' refers to movement onset detected by the optical sensor signal; 'EMG onset' refers to movement onset detected by EMG.

for >25% of sensors exceeded the threshold, (3) excluding sensors where absolute EEG signals for >25% of trials exceeded the threshold, (4) excluding trials where absolute EEG signals for at least one sensor were >120  $\mu$ V. Consequently, the number of available trials was reduced from 120 to approximately 80 trials for each subject. The number of available sensors was 64 for subject TA and 62 for subject RO.

EEG signals after the above preprocessing were used for both the calculation of event-related potentials (ERPs) and for source current estimation with VBMEG.

#### Settings for VBMEG

As shown in the result section below, the hyper-parameter values  $m_0 = 100$  and  $\gamma_0 = 10$  performed relatively well in most situations. Therefore, we used these values for all cortical dipoles. In addition to the cortical dipoles, artifact dipoles located at the center of the left and right eyeballs were assumed in the estimation with VBMEG. The results of previous studies (Fujiwara et al., 2009; Morishige et al., 2009) suggested that the incorporation of artifact dipoles allowed the effective removal of eye-movement artifacts from EEG data. Hyper-parameter values of the prior estimate of the current variance of  $10^2$  [ $\text{nAm}^2$ ] and  $\gamma_0 = 10^{4.5}$  were used for artifact dipoles, based on a previous study (Morishige et al., 2009).

The EEG time series from 1000 ms before to 200 ms after the movement initiation was analyzed to study the brain activity involved in the late preparation, execution and somatosensory information-processing stages of finger lifting (movement duration was about 500 ms). The time series data from 1500 ms to 1000 ms before the movement initiation were used to estimate the baseline of the current variance. Because the pattern of cortical activity was assumed to change according to the phase of movement (e.g., preparation, execution), the EEG time series was divided into 100-ms time-windows with 50-ms overlaps. Source current estimation with VBMEG was conducted for each series of divided EEG data under the assumption that the current variance was temporally invariant within each time window. Thus, the current variance changed every 50 ms, but the current was estimated every 4 ms (because the sampling rate was 250 Hz).

As described above, estimations with VBMEG were conducted for the combinations of three numbers of EEG sensors<sup>4</sup> (64-ch, 31-ch, 19-ch; Fig. 1) and three prior information types (fMRI prior, NIRS prior, no prior).

#### Evaluation of the performance of source current estimation with VBMEG

Because it is impossible to determine the true source current in real situations, the estimation accuracy measures that require true source positions and/or current amplitudes, such as the localization error, were not available for the real experimental data. We therefore evaluated the performance of VBMEG as follows. First, the estimated current for the combination of 64-ch EEG and fMRI was used as a reference, because previous studies demonstrated that VBMEG with fMRI prior yielded physiologically acceptable results in the case of MEG (Toda et al., 2011; Yoshioka et al., 2008). Then, variants of three metrics used in the simulation study were calculated by regarding the reference currents as substitutes of true currents. The first metric was the correlation coefficient ( $R_{st}$ ) between the spatio-temporal pattern of the estimated current under the reference condition (i.e., the combination of 64-ch EEG and fMRI) and that under the comparison condition. The second metric was the correlation coefficient ( $R_s$ ) between the temporally-averaged (during the 200 ms period just before movement onset) spatial pattern of the estimated current under the reference condition and that under the comparison condition. The third metric was the RMSE between the estimated current under the reference condition and that under the comparison

condition. RMSE was calculated over the analysis time-windows (from 1000 ms before to 200 ms after onset).

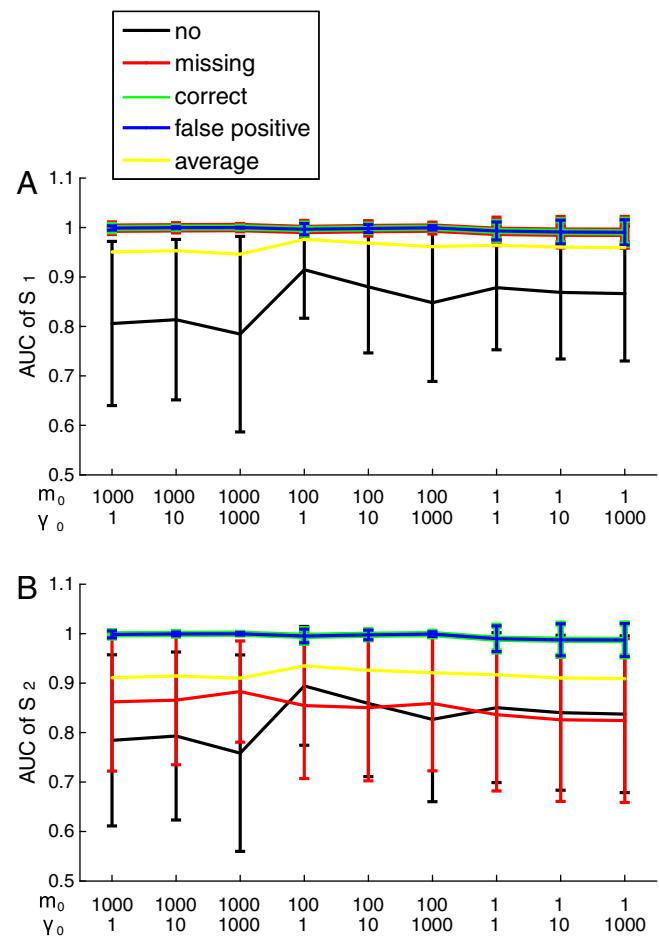
## Results

### Results of computer simulations

First, we describe the results of the first step simulation, in which we investigated the effects of the types of prior information (no, missing, correct, and false positive priors) and the values of the hyper-parameters ( $m_0$ ,  $\gamma_0$ ), and their interactions. Then, we describe the results of the second step simulation, where we investigated the effects of the number of EEG sensors (64ch, 31ch, 19ch) and the resolution of the prior information ( $R = 6, 12, 18$ ), and their interactions.

### Results of the first step

Fig. 3 shows the results of ROC analysis. For the first source,  $S_1$ , two-way ANOVA revealed significant effects of the types of prior information [ $F(3, 17,999) = 4184.99, p < 0.05$ ] and the combinations of hyper-parameters [ $F(8, 17,999) = 29.84, p < 0.05$ ], and a significant interaction between these factors [ $F(24, 17,999) = 42.99, p < 0.05$ ]. For the second source,  $S_2$ , two-way ANOVA revealed significant effects of the types of prior information [ $F(3, 17,999) = 3090.02, p < 0.05$ ] and the combinations of hyper-parameters [ $F(8, 17,999) = 13.03, p < 0.05$ ], and a significant interaction between these factors [ $F(24, 17,999) = 26.28, p < 0.05$ ]. When the correct or false positive prior

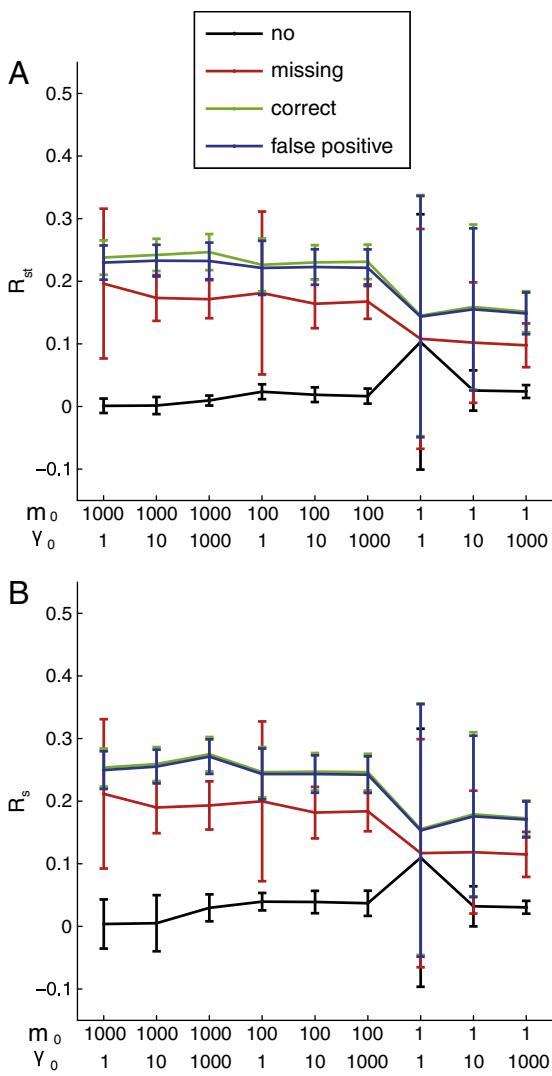


**Fig. 3.** Results of ROC analysis for the first step simulation. The Areas Under Curves (AUCs) for  $S_1$  (A) and  $S_2$  (B) are plotted against the combinations of hyper-parameters, for no prior (black line), missing prior (red line), correct prior (green line), and false positive prior (blue line). The yellow line is the average of all prior types. Error bars indicate SDs.

<sup>4</sup> EEG data were obtained with a 64-electrode cap for both the 31-ch and 19-ch cases with no additional experiments.

was used, the AUC was almost 1 (i.e., perfect detection) for both  $S_1$  (Fig. 3A, green or blue line) and  $S_2$  (Fig. 3B, green or blue line), regardless of the values of hyper-parameters tested in the present study (multiple comparison,  $p > 0.05$ ). When the missing prior was used, the AUC was almost 1 for  $S_1$ , which had correct active prior information, regardless of the values of hyper-parameters (multiple comparison,  $p > 0.05$ ; Fig. 3A, red line). However, the AUC for  $S_2$ , which had false inactive prior information, decreased to about 0.85 and changed with the values of hyper-parameters; roughly speaking,  $m_0 = 1$ , which is far smaller than the theoretical value of 674 in the present setting, yielded slightly lower AUC values (multiple comparison,  $p < 0.05$ ; Fig. 3B, red line). With no prior information, the AUC values for both  $S_1$  and  $S_2$  were both around 0.85 and changed with the values of hyper-parameters; roughly speaking,  $m_0 = 1000$  yielded lower AUC values (multiple comparison,  $p < 0.05$ ; Figs. 3A and B, black lines). Taken together, these results indicate that an overly large or small value of  $m_0$  (i.e.,  $m_0 = 1$  or 1000) can yield incorrect estimates, depending on the type of prior information.

Fig. 4 shows correlation coefficients ( $R_{st}$  and  $R_s$ ) between true and estimated currents against the combination of hyper-parameters.



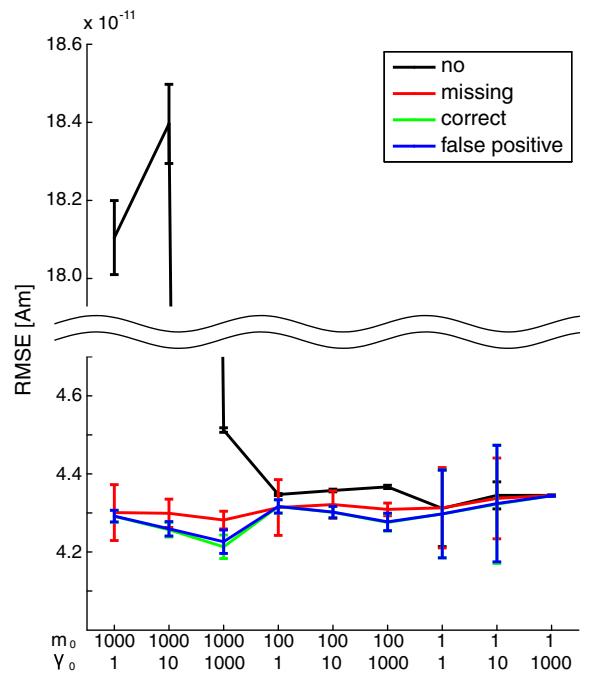
**Fig. 4.** Results of correlation analysis for the first step simulation. (A) Correlation coefficients between true and estimated spatio-temporal current patterns,  $R_{st}$ , are plotted against the combinations of hyper-parameters. (B) Correlation coefficients between true and estimated spatial current patterns,  $R_s$ , are plotted against the combinations of hyper-parameters. The line colors are the same as those in Fig. 3. Error bars indicate SDs.

Both  $R_{st}$  and  $R_s$  showed similar tendencies. For  $R_{st}$ , a two-way ANOVA showed significant effects of the types of prior information [ $F(3, 17,999) = 4797.31, p < 0.05$ ] and the combinations of hyper-parameters [ $F(8, 17,999) = 186.74, p < 0.05$ ] and a significant interaction between these factors [ $F(24, 17,999) = 79.22, p < 0.05$ ]. For  $R_s$ , there were significant effects of the types of prior information [ $F(3, 17,999) = 4971.55, p < 0.05$ ] and the combinations of hyper-parameters [ $F(8, 17,999) = 209.96, p < 0.05$ ] and a significant interaction between these factors [ $F(24, 17,999) = 77.64, p < 0.05$ ]. Both correlations were far higher when the prior was available compared to when it was unavailable, even if the prior included wrong information. When the prior was available, both correlations were lower with  $m_0 = 1$ , the value furthest away from the theoretical value of 674, and the SDs of both correlations were larger with  $\gamma_0 = 1$ .

Fig. 5 shows the RMSE against the combination of hyper-parameters. Two-way ANOVA revealed significant effects of the types of prior information [ $F(3, 17,999) = 2.8696 \times 10^6, p < 0.05$ ] and the combinations of hyper-parameters [ $F(8, 17,999) = 1.1897 \times 10^6, p < 0.05$ ], and a significant interaction between these factors [ $F(24, 17,999) = 1.2042 \times 10^6, p < 0.05$ ]. When the prior was not available, errors were markedly large with  $m_0 = 1000$  regardless of the value of  $\gamma_0$  (Fig. 5).

The effect of the types of prior information can be summarized as follows. The performance of VBMEG was markedly better when the prior information was available compared to when it was unavailable, even if it contained incorrect information. Estimation with VBMEG was relatively robust for the false positive prior but vulnerable to the missing prior, especially for the source with false inactive prior (i.e.,  $S_2$ ).

It is difficult to determine the best combination of hyper-parameters because the estimation performance was not severely affected by the combination of hyper-parameters when prior information was available. Regarding the magnification parameter,  $m_0$ , we observed the following tendency. Overly small values of  $m_0$  (i.e.,  $m_0 = 1$ ) sometimes yielded incorrect estimates, even when the correct prior was available (e.g., see Fig. 4). The largest value of  $m_0$  (i.e.,  $m_0 = 1000$ ) yielded good estimates when the prior was available, but yielded



**Fig. 5.** The root mean squared errors (RMSE) between true and estimated spatio-temporal current patterns are plotted against the combinations of hyper-parameters, for all types of priors. The line colors are the same as those in Fig. 3. Error bars indicate SDs.

incorrect estimates when the prior was not available (see Fig. 5). These findings suggest that a moderate value of  $m_0$  (i.e.,  $m_0=100$ ) should be adopted. Regarding the confidence parameter,  $\gamma_0$ , the smallest value of  $\gamma_0$  (i.e.,  $\gamma_0=1$ ) sometimes yielded unstable estimates (i.e., SDs of evaluation metrics were large; e.g., see Fig. 4). Because the spatial resolution and localization accuracy of NIRS are limited by measurement geometry to 3 cm, a lower value of  $\gamma_0$  should be used in the experimental data analysis. Therefore, we chose to use  $[m_0, \gamma_0]=[100, 10]$  instead of  $[m_0, \gamma_0]=[100, 1000]$  in both the second step simulation and experimental data analysis. An additional reason for adopting  $[m_0, \gamma_0]=[100, 10]$  is that this combination has yielded valid results in previous studies (Fujiwara et al., 2009; Toda et al., 2011).

#### Results of the second step

Fig. 6 shows the results of ROC analysis for  $S_1$  (Fig. 6A) and  $S_2$  (Fig. 6B). The AUC value was very close to 1 (perfect detection), regardless of the resolution of prior information or the number of EEG sensors. For  $S_1$ , two-way ANOVA revealed a significant effect of the resolution of prior information [ $F(2, 4499)=200.69, p<0.05$ ], but revealed no significant effect of the number of EEG sensors [ $F(2, 4499)=0.68, p=0.5068$ ] or a significant interaction between these factors [ $F(4, 4499)=1, p=0.4074$ ]. Also for  $S_2$ , two-way ANOVA revealed a significant effect of the resolution of prior information [ $F(2, 4499)=109.26, p<0.05$ ], but did not reveal a significant effect of the number of EEG sensors [ $F(2, 4499)=0.45, p=0.6393$ ] or a significant interaction between these factors [ $F(4, 4499)=1.54, p=0.1872$ ]. These results suggest that, when the correct prior was available, detection improved with the resolution of prior information, whereas it hardly changed with the number of EEG sensors.

Fig. 7A shows the results for  $R_{st}$ . Two-way ANOVA revealed significant effects of the resolution of prior information [ $F(2, 4499)=25,122.9,$

$p<0.05$ ] and the number of EEG sensors [ $F(2, 4499)=117.68, p<0.05$ ], but interaction between these factors was not significant [ $F(4, 4499)=1.32, p=0.2617$ ]. The correlation became lower as the spatial resolution of the prior worsened and as the number of EEG sensor decreased (multiple comparison,  $p<0.05$ ). The lowest correlation,

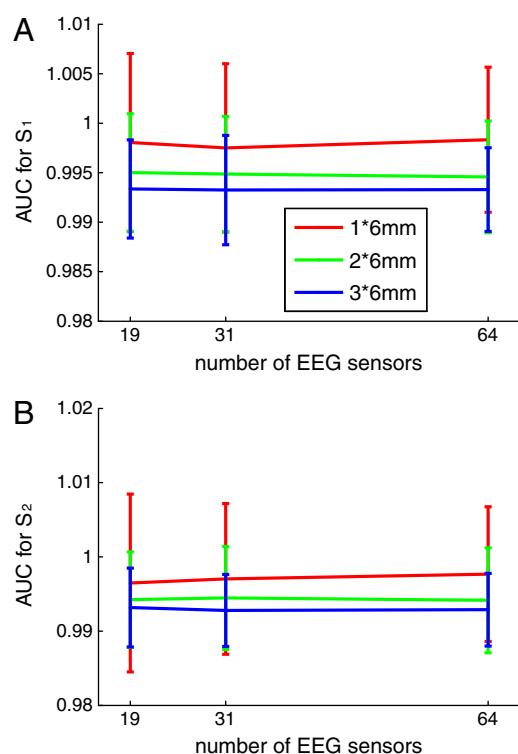


Fig. 6. Results of ROC analysis for the second step simulation. The AUCs for  $S_1$  (A) and  $S_2$  (B) are plotted against the number of EEG sensors, for  $R=6$  (red),  $R=12$  (green), and  $R=18$  (blue). Error bars indicate SDs.

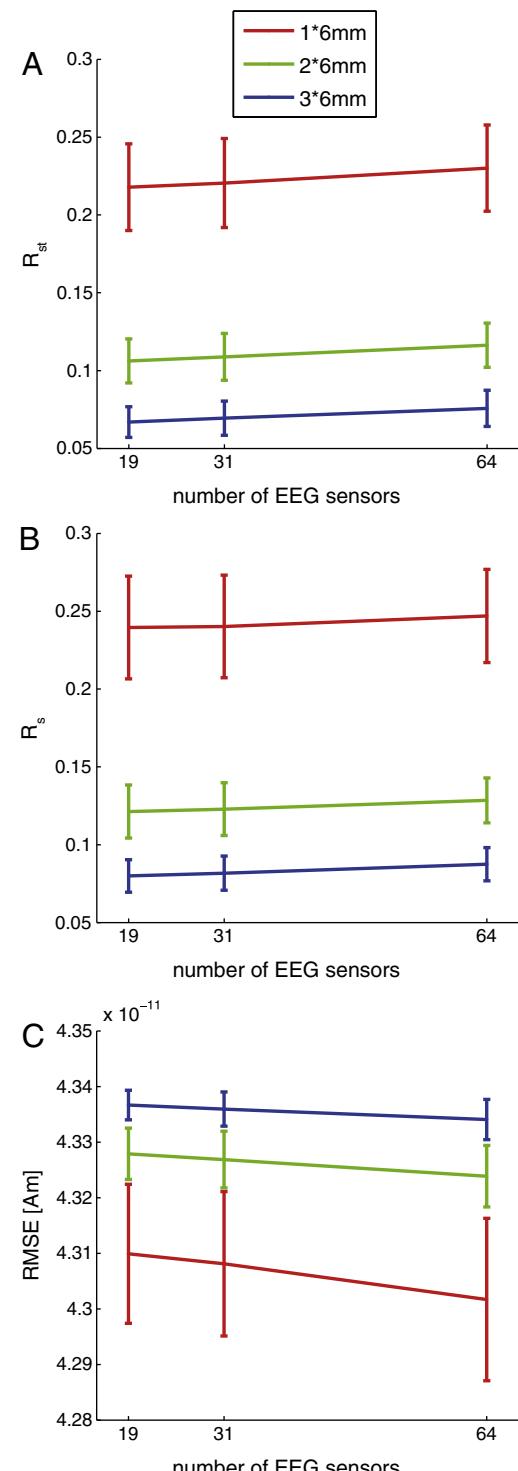
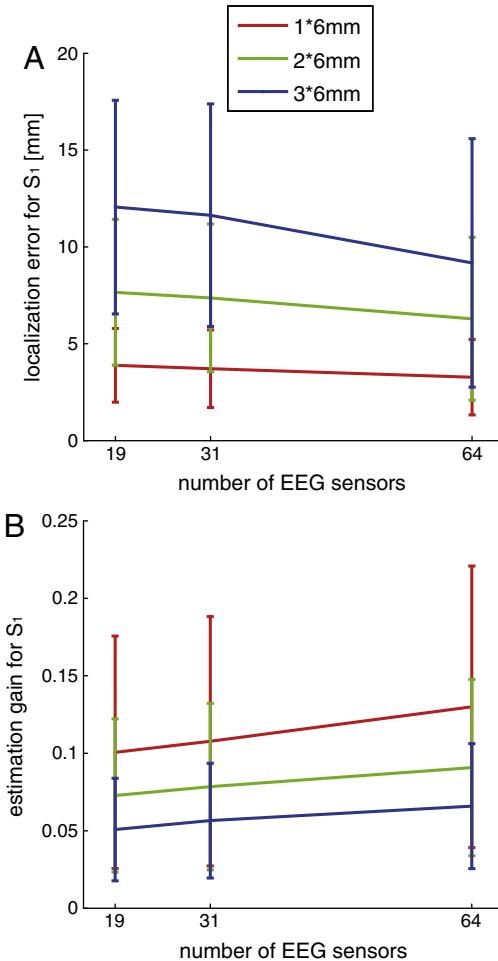


Fig. 7. Results of correlation analysis and RMSE analysis for the second step simulation. The evaluation metrics are plotted against the number of EEG sensors. The line colors are the same as those in Fig. 6. (A) The correlation coefficients of spatio-temporal patterns,  $R_{st}$ . (B) The correlation coefficients of spatial patterns,  $R_s$ . (C) The root mean squared errors, RMSE. Error bars indicate SDs.

$R_{st} = 0.0670$  (0.0098), was obtained for the combination of the 19-ch EEG and a prior with  $R = 18$ . However, in the first step, the average (SD) of  $R_{st}$  was only 0.0188 (0.0118) when the prior was not available even if the number of EEG sensors was the maximum, 64. Therefore,  $R = 18$  with 19 EEG sensors led to markedly higher performance. Fig. 7B shows the results for  $R_s$ . Two-way ANOVA revealed significant effects of the resolution of prior information [ $F(2, 4499) = 21,958.09, p < 0.05$ ] and the number of EEG sensors [ $F(2, 4499) = 49.62, p < 0.05$ ], but interaction between these factors was not significant [ $F(4, 4499) = 0.15, p = 0.965$ ]. Just like the  $R_{st}$  case, the correlation became lower as the resolution worsened and as the number of EEG sensors decreased (multiple comparison,  $p < 0.05$ ). The lowest correlation,  $R_s = 0.080$  (0.0104), was obtained for the combination of the 19-ch EEG and prior with  $R = 18$ . However, in the first step, the average (SD) of  $R_s$  was only 0.0389 (0.0178) when the prior was not available even if the number of EEG sensors was the maximum, 64. Therefore,  $R = 18$  with 19 EEG sensors was far better. Fig. 7C illustrates the results of the RMSE. Two-way ANOVA revealed significant effects of the resolution of prior information [ $F(2, 4499) = 4574.81, p < 0.05$ ] and the number of EEG sensors [ $F(2, 4499) = 139.82, p < 0.05$ ] and a significant interaction between these factors [ $F(4, 4499) = 16.7, p < 0.05$ ]. RMSE became larger as the resolution of the prior worsened and as the number of EEG sensors decreased (multiple comparison,  $p < 0.05$ ). The largest RMSE was  $4.3367 \times 10^{-11}$  ( $2.6579 \times 10^{-14}$ ) [Am] for the combination of 19-ch EEG and prior with  $R = 18$ . However, considering that the average (SD) of RMSE was  $4.3575 \times 10^{-11}$  ( $3.4765 \times 10^{-14}$ ) [Am] when no prior was available, using the prior with  $R = 18$  produced markedly better performance. Taken together, the performance of VBMEG worsened as the resolution of the prior became lower and as the number of EEG sensors decreased, but even the worst combination of prior resolution and the number of sensors was far better than the combination of no prior and the densest EEG as far as we investigated. The effect of the number of EEG sensors was small as compared to that of the resolution of the prior.

Fig. 8 shows the results of additional evaluation metrics, localization error and estimation gain, for  $S_1$ . For the localization error of  $S_1$ , two-way ANOVA revealed significant effects of the resolution of prior information [ $F(2, 4499) = 1117.89, p < 0.05$ ] and the number of EEG sensors [ $F(2, 4499) = 62.01, p < 0.05$ ] and significant interaction between these factors was not significant [ $F(4, 4499) = 11.23, p < 0.05$ ]. The localization error became larger as the resolution of the prior worsened and as the number of EEG sensors decreased (multiple comparison,  $p < 0.05$ ). But the effect of the number of EEG sensors was small as compared to that of the resolution of the prior. For the estimation gain of  $S_1$ , two-way ANOVA showed significant effects of the resolution of prior information [ $F(2, 4499) = 312.95, p < 0.05$ ] and the number of EEG sensors [ $F(2, 4499) = 46.77, p < 0.05$ ], but interaction between these factors was not significant [ $F(4, 4499) = 2.35, p = 0.0522$ ]. The estimation gain decreased as the resolution of the prior worsened and as the number of EEG sensors decreased (multiple comparison,  $p < 0.05$ ). Again, the effect of the number of EEG sensors was small as compared to that of the resolution of the prior. Because both  $S_1$  and  $S_2$  had correct active prior information,  $S_2$  showed similar results (see Tables 1 and 2). From Table 1, it can be seen that the localization error was less than the spatial resolution of the prior if the correct prior is available. From Table 2, it can be seen that the estimation gain was 0.12 even when the source current was estimated with 64-ch EEG using the correct prior with the highest resolution ( $R = 6$ ).

The results of the second step can be summarized as follows. Estimation accuracy decreased as both the resolution of prior information and the number of EEG sensors, but the resolution had a larger effect than the number of sensors. However, estimation with the combination of the lowest density EEG (i.e., 19-ch) and prior resolution (i.e.,  $R = 18$ ) performed far better than estimation using the densest EEG (i.e., 64-ch) without prior.



**Fig. 8.** Results of the second step of the simulation. (A) The localization errors for  $S_1$  are plotted against the number of EEG sensors. (B) The estimation gains for  $S_1$  are plotted against the number of EEG sensors. The line colors are the same as those in Fig. 6. Error bars indicate SDs.

#### Results of human scanning experiments

##### EEG, fMRI and NIRS activity

Experimental data for subject RO are presented in Fig. 9. Fig. 9A shows the SMA, the PMA, the M1, and the primary somatosensory area ( $S_1$ ) determined using the Automated Anatomical Labeling (AAL) in the WFU PickAtlas (Maldjian et al., 2003) from structural MRI. These areas were selected as the regions of interest (ROI) for later analysis, because they are related to movement preparation/execution and somatosensory information processing.

**Table 1**

Localization error.

Number of EEG sensors	19	31	64
$S_1$			
$R = 6$	3.89 (1.90)	3.72 (2.00)	3.28 (1.95)
$R = 12$	7.77 (3.77)	7.37 (3.81)	6.29 (4.20)
$R = 18$	12.06 (5.52)	11.64 (5.74)	9.17 (6.42)
$S_2$			
$R = 6$	4.02 (1.88)	3.95 (2.21)	3.44 (2.09)
$R = 12$	7.59 (3.78)	7.54 (3.99)	6.44 (4.06)
$R = 18$	12.36 (5.67)	12.10 (5.79)	9.64 (6.28)

Mean localization errors from 500 combinations of sources ( $S_1$  and  $S_2$ ) are listed in millimeters. SDs are listed in parentheses.

**Table 2**  
Estimation gain.

Number of EEG sensors	19	31	64
<b>S<sub>1</sub></b>			
R = 6	0.10 (0.08)	0.11 (0.08)	0.13 (0.09)
R = 12	0.07 (0.05)	0.08 (0.05)	0.09 (0.06)
R = 18	0.05 (0.03)	0.06 (0.04)	0.07 (0.04)
<b>S<sub>2</sub></b>			
R = 6	0.09 (0.07)	0.10 (0.07)	0.12 (0.09)
R = 12	0.07 (0.05)	0.07 (0.05)	0.09 (0.06)
R = 18	0.05 (0.03)	0.05 (0.03)	0.07 (0.04)

Mean estimation gains from 500 combinations of sources ( $S_1$  and  $S_2$ ) are listed. SDs are listed in parentheses.

**Fig. 9B** shows the ERPs recorded at C3 and C4 of the International 10–20 system. MRPs began to increase around 600 ms before movement onset and peaked at around movement onset. The observation that the early and late components of the MRPs were lateralized towards the hemisphere contralateral to the movement, combined with the finding that the movement onset detected by the optical sensor was delayed by about 130 ms relative to that detected by EMG, suggests that the early and late components of the observed MRPs represented NS and MP, respectively. **Fig. 9C** shows the fMRI activity for right index finger movements; the movements activated the contralateral SM1 area. These data were used as prior information for later current estimation with VBMEG. **Fig. 9D** shows NIRS activity mapped to the cortical surface using Fusion software. Again, these data were also used as prior information for later current estimation. Increases in OxyHb were observed in bilateral motor-related areas. The results revealed weak lateralization towards the hemisphere

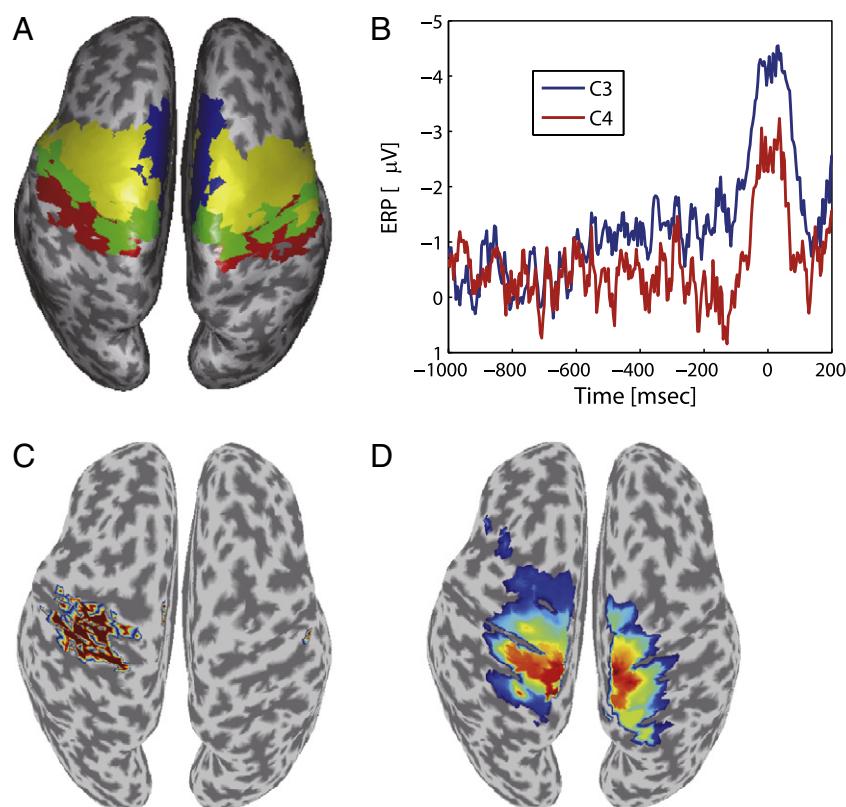
contralateral to the movement. Similar results were also found for subject TA.

#### Cortical current estimated by VBMEG

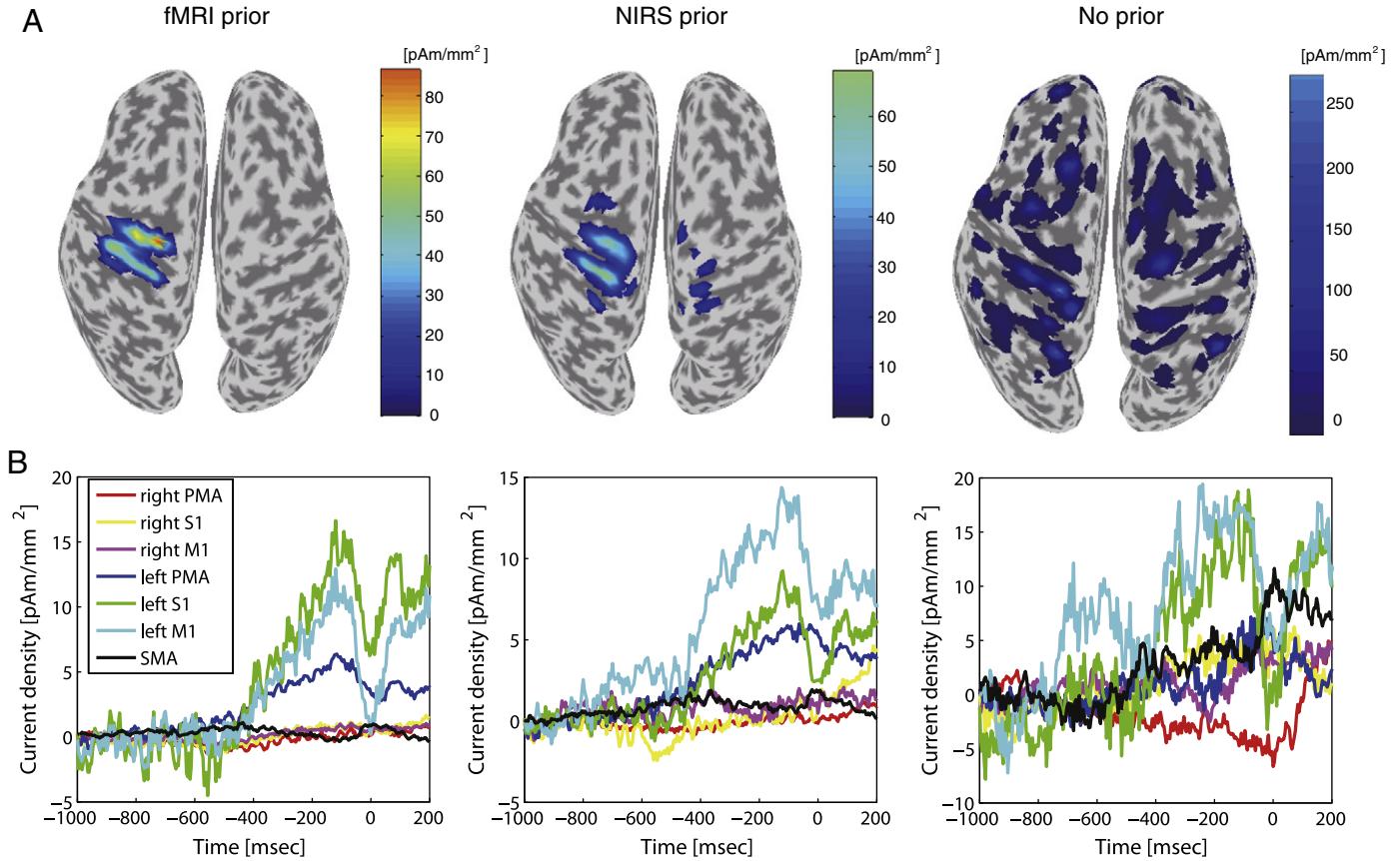
The spatial pattern of the estimated cortical current was visualized by averaging the estimated current density of each dipole over the 200 ms before movement onset to focus on the brain activities involved with movement execution. The temporal pattern of the estimated current was visualized by averaging the estimated current density within each area of SMA, PMA, M1 and S1.

**Fig. 10** shows results of VBMEG applied to 64-ch EEG data with fMRI prior (left), NIRS prior (middle), and no prior (right) for subject RO. **Figs. 10 A** and **B** show the spatial and temporal patterns, respectively.

Regarding the results of the fMRI prior case, the spatial pattern (**Fig. 10A**, left) was similar to the fMRI activity (**Fig. 9C**), confirming the validity of the estimation. The temporal pattern (**Fig. 10 B**, left) shows that current densities in the left (i.e., contralateral) PMA, M1 and S1 began to increase around 500 ms before movement onset and peaked just before movement onset. These results were consistent with the findings of previous studies that used other source localization methods with MEG data (Cheyne et al., 2006). The amplitudes of the estimated currents were higher in left M1 and S1 than in SMA and PMA, consistent with the previous studies investigating the source current from MEG during voluntary movements (Cheyne et al., 2006; Huang et al., 2004; Toda et al., 2011). In addition, the order of the estimated amplitude was comparable to that in previous studies on MEG source current in sensorimotor (Cheyne et al., 2006; Huang et al., 2004; Toda et al., 2011) and visual (Yoshioka et al., 2008) regions. Thus, the results of VBMEG applied to 64-ch EEG



**Fig. 9.** Experimental data for subject RO. (A) The supplementary motor, premotor, primary motor, and primary somatosensory areas are represented by blue, yellow, green and red, respectively. (B) Event-related potentials (ERPs) for right index finger movements at C3 and C4 of the International 10–20 system. (C) fMRI activity used for prior information. (D) NIRS activity used for prior information. Note that we mapped the NIRS measurements onto the cortical surface without solving the inverse problem.



**Fig. 10.** Results of VBMEG under different priors. Estimations with fMRI prior, NIRS prior, and no prior are shown from left to right. (A) Spatial patterns for estimated currents averaged over a 200 ms duration just before movement onset. (B) Temporal patterns for estimated currents averaged within each area. PMA; premotor area, S1; primary somatosensory area, M1; primary motor area, SMA; supplementary motor area.

with fMRI prior were physiologically reasonable, supporting its use as a reference.

Regarding the NIRS prior case, the spatial pattern (Fig. 10A, middle) was similar to both the fMRI activity (Fig. 9C) and the spatial pattern for the case with fMRI-prior information (Fig. 10A, left), suggesting the validity of estimation with VBMEG. The temporal pattern (Fig. 10B, middle) was similar to that for cases using fMRI information as a prior, in accord with several previous reports (Cheyne et al., 2006; Huang et al., 2004; Toda et al., 2011). In addition, the order of the estimated amplitude was comparable to that in previous studies (Cheyne et al., 2006; Huang et al., 2004; Toda et al., 2011; Yoshioka et al., 2008). Thus, VBMEG applied to 64-ch EEG with NIRS prior performed relatively well.

Regarding the results of the no prior case, the spatial pattern (Fig. 10 A, right) showed that estimated current sources were located in widespread areas in addition to the activated areas indicated in the fMRI experiment (mainly left SM1, see Fig. 9C) and no clear laterality was observed. Furthermore, the temporal pattern (Fig. 10B, right) for each area was noisy, though the activities of the left M1 and S1 increased during the 500-ms period before movement onset. Thus, the performance of VBMEG with no prior information was rather poor, in accord with previous MEG study (Yoshioka et al., 2008) and the present simulation experiment.

The quantitative evaluation of the experimental data is as follows. First, we describe the effect of the types of prior information (see Fig. 10 for the qualitative evaluation). Table 3 shows that the current density pattern with the NIRS prior resembled the reference pattern (i.e., the pattern for the combination of 64-ch EEG and fMRI prior) more than the pattern with the prior in terms of spatio-temporal

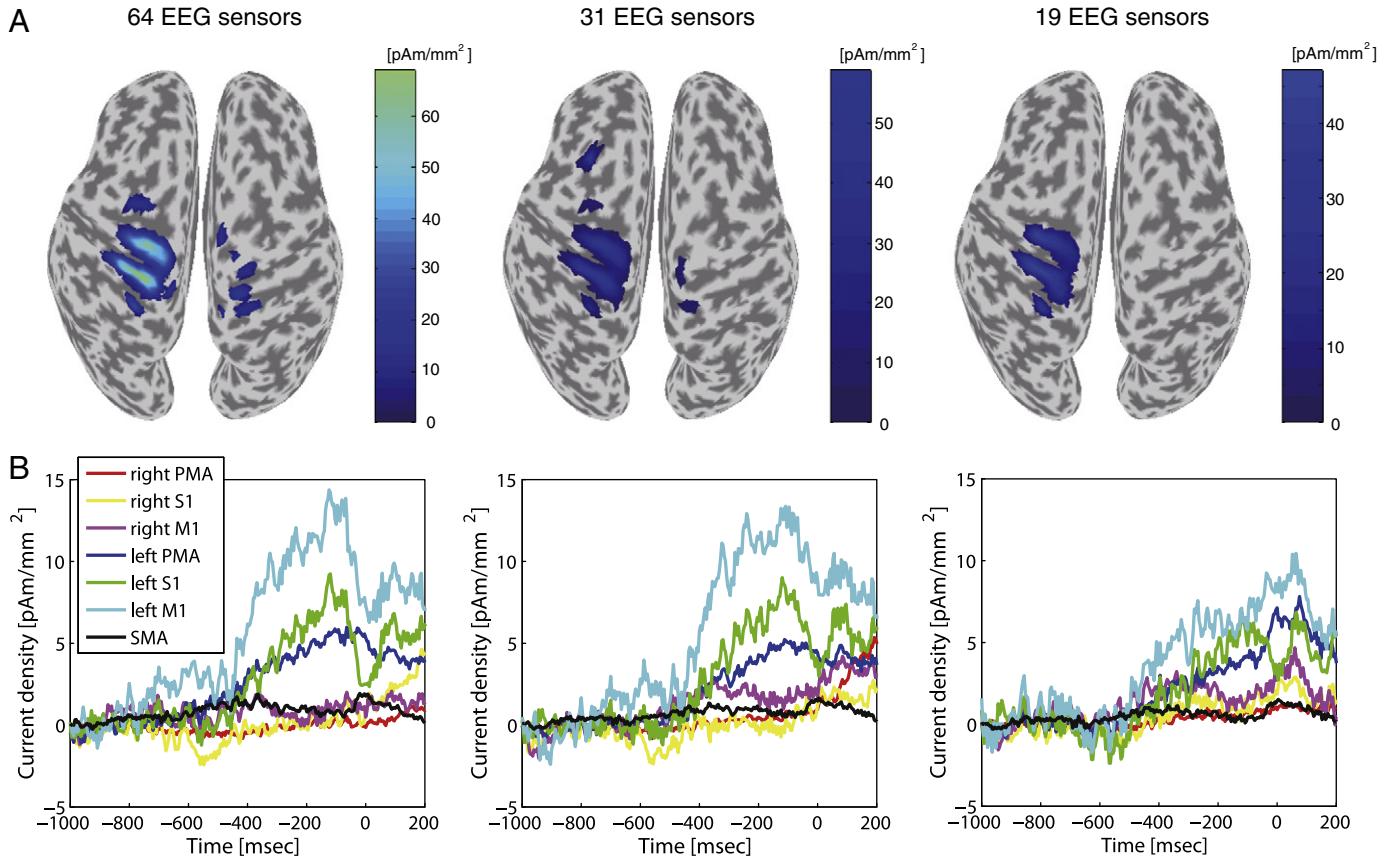
( $R_{st}$ ) and spatial correlations ( $R_s$ ) and root mean squared errors (RMSE). This was true regardless of the number of EEG sensors. In addition, estimation using the fMRI prior performed better than that using the NIRS prior, for both 19-ch and 31-ch EEG, in terms of all of these metrics using the reference pattern. These findings are consistent with the results of the second step simulation study, indicating that estimation accuracy was increased by the use of prior information (even when the spatial resolution of prior information was low and/or when the prior included incorrect information), compared with the estimation accuracy with no prior information, and that

**Table 3**

Comparison of the current density pattern for each combination with the reference pattern.

Evaluation metrics	$R_{st}$	$R_s$	RMSE [ $\text{pAm/mm}^2$ ]
fMRI prior			
31-ch EEG	0.76 (0.23)	0.80 (0.23)	5.82 (6.07)
19-ch EEG	0.44 (0.17)	0.53 (0.09)	12.0 (4.21)
NIRS prior			
64-ch EEG	0.40 (0.14)	0.45 (0.13)	5.34 (2.32)
31-ch EEG	0.18 (0.07)	0.37 (0.15)	9.49 (4.81)
19-ch EEG	0.18 (0.06)	0.36 (0.12)	9.13 (4.99)
No prior			
64-ch EEG	0.07 (0.03)	0.08 (0.04)	103.08 (83.18)
31-ch EEG	0.02 (0.00)	0.03 (0.01)	165.98 (157.57)
19-ch EEG	0.02 (0.00)	0.02 (0.01)	148.88 (181.26)

Evaluation metrics which reflect similarities between the current density pattern for each combination and that for the reference (i.e., the combination of 64-ch EEG and fMRI prior) are compared. The evaluation metrics were averaged over two subjects. SDs are listed in parentheses.



**Fig. 11.** Results of VBMG with NIRS prior under different numbers of EEG sensors. Numbers of EEG sensors of 64, 31 and 19 (in reality, 62, 29 and 17 sensors, because of the rejection of extraordinary sensors) are shown from left to right. (A) Spatial patterns for estimated currents averaged over the 200 ms period just before movement onset. (B) Temporal patterns for estimated currents averaged within each area. PMA: premotor area, S1: primary somatosensory area, M1: primary motor area, SMA: supplementary motor area.

the estimation accuracy increased with the spatial resolution of prior information.

Next, we describe the effect of the number of EEG sensors (Table 3; see Fig. 11 for the qualitative evaluation) in combination with the interaction of the types of prior information and the number of EEG sensors. When the fMRI/NIRS prior was used, all evaluation metrics worsened as the number of EEG sensors decreased. These findings are consistent with the second step simulation results that estimation accuracy decreased with the number of EEG sensors when the prior is available. In addition, the combination of 19-ch EEG with the NIRS prior performed higher than that of 64-ch EEG with no prior, consistent with the simulation result that the prior information improves estimation performance even if the spatial resolution is relatively low and the number of EEG sensors is few.

## Discussion

The current study evaluated the performance of VBMG for estimating cortical currents from EEG data, and investigated the effects of prior information and the number of EEG sensors using both computer simulations and real experiments. A summary of the results is presented below.

The data from the first step simulation suggested the following: first, as long as the magnification parameter  $m_0$  was not far from the theoretical value, the estimation performance was very high when correct prior information with high spatial resolution was used. It should be noted that the SNR used in the simulations was similar to that observed in the present experiment. Second, two incorrect priors showed different effects on estimation accuracy; the

false positive prior had little harmful effect, but the missing prior worsened the detection of the source, especially that with false inactive prior information. Third, estimation performance was relatively high for the most types of prior information when the moderate value of the magnification parameter ( $m_0 = 100$ ) was used.

In the second step of the simulations, we observed that the performance of VBMG increased with both the spatial resolution of prior information and the number of EEG sensors, but the effect of the resolution was larger. In addition, estimation accuracy was better for the combination of the lowest density (i.e., 19-ch) EEG and prior with the lowest spatial resolution (i.e.,  $R = 18$ ) than for the combination of the highest density (i.e., 64-ch) EEG and no prior.

Experimental data confirmed the following. First, both the spatial and temporal patterns of the estimated current produced physiologically plausible results if fMRI information was incorporated as a hierarchical prior. This result is consistent with the simulation findings revealing that VBMG applied to EEG data with prior information of high spatial resolution correctly estimated the cortical current patterns, even when the noise level was similar to that in real situations. Second, the estimated current with NIRS prior information resembled that with fMRI prior information more than it resembled the estimation with no prior. This finding is also consistent with the simulation results indicating that the estimation accuracy was increased by the use of prior information (even when prior information is not very accurate), compared with the estimation accuracy with no prior information. Third, decreasing the number of EEG sensors worsened the accuracy in the fMRI/NIRS case, consistent with the second step simulation result that decreasing the number of EEG sensors slightly worsened the estimation when the prior was available. Fourth, the

current pattern for 19-ch EEG with NIRS prior was more similar to that for 64-ch EEG with fMRI prior than it was to that for 64-ch EEG with no prior information. This finding is also consistent with the simulation results.

The most important result of the present study is the demonstration that, by combining EEG and NIRS data under the framework of hierarchical Bayesian estimation, it is possible to obtain relatively high spatio-temporal, physiologically plausible, brain activity, which cannot be obtained using the individual data. Because both EEG and NIRS are associated with the merits of high portability and low sensitivity to body motion artifacts, VBMG with combined EEG–NIRS potentially represents a powerful tool for neuroscience research. This combined method is suitable for measuring brain activity in infants or patients, for investigating brain activity related to human motor control in realistic situations (e.g., in sitting or standing positions), and for daily use, e.g., monitoring day-to-day changes in brain activity during stroke recovery or motor learning. In addition to the benefits of combining EEG and NIRS data using VBMG, it is important to consider the potential weaknesses of VBMG. As seen in the first step of the simulation, (1) estimation accuracy deteriorates if the wrong combination of hyper-parameters is chosen, and (2) the source with false inactive prior information was hard to be detected. These methodological weaknesses should be taken into account when interpreting the results of estimation with VBMG.

In this study, task protocols used in EEG, NIRS and fMRI experiments were not consistent; in EEG experiments the subjects were instructed to perform a repetitive right finger movement slowly (less than 0.3 Hz), whereas in NIRS and fMRI experiments they were instructed to perform the motor task rapidly. According to previous studies, the SM1 and SMA are activated during slow-rate movements, reflecting not only the execution and somatosensory information-processing stages, but also the preparation stage, whereas SM1 is predominantly activated during fast-rate movements, mainly reflecting the late preparation, execution and somatosensory information-processing stages (Kunieda et al., 2000; Toma et al., 1999). In fact, results of our fMRI/NIRS experiments supported this. Considering our simulation result that VBMG with missing prior exhibited difficulty in detecting the source with false inactive prior, it may be problematic to use fMRI/NIRS activity directly as prior because it lacks the SMA activity. For this reason, as described in [EEG experiments and preprocessing](#), we eliminated the EEG component reflecting the preparation stage, the BP component, using a high-pass filter of 0.2 Hz. By this preprocessing, we can focus on the brain activity common for both fMRI/NIRS and EEG experiments, SM1 activity. We therefore consider that the fMRI/NIRS prior used in the present study is almost adequate. However, using the same task in the experiments for both prior information and EEG/MEG appears to be desirable. Moreover it will be better to adopt simultaneous measurement, because when EEG and fMRI/NIRS data are obtained in different experiments, the possibility that the brain activity measured in the fMRI/NIRS experiment may not correspond precisely to that in the EEG experiment may arise even if the task in both experiments is the same. Simultaneous measurement enhances the reliability of prior information, which may improve estimation accuracy. Another merit of simultaneous measurement is that it is suitable for monitoring short-term changes in brain activity, for example, during motor adaptation. Because such changes may have little reproducibility, it is impossible to monitor them by measuring EEG/MEG and NIRS/fMRI separately. It should be noted that simultaneous measurement is technically difficult for the combination of MEG–fMRI, but is possible for EEG–NIRS. Shimadzu Corporation has already developed a simultaneous measurement system for EEG–NIRS. The present study will therefore lay the foundations for future studies using simultaneous EEG and NIRS measurement.

When the mismatch between brain activity measured by NIRS and that by EEG is caused by the task difference, simultaneous measurement

will be the ideal solution. However, since the physiological sources of both modalities are different (i.e., EEG records the electrical activity and NIRS detects the hemodynamic changes), spatial characteristics of NIRS responses to evoked neural activity may not be consistent with those of EEG responses. Simulation results suggest that the reliability of estimation with VBMG depends on the type of such physiologically mismatch; only when the active areas in EEG responses are subsets of those in NIRS responses, the estimation is reliable.

In the simulation study, the estimation accuracy of VBMG decreased with the number of EEG sensors, though the effect of the number of EEG sensors was small as compared to that of the spatial resolution of prior. In the real experiments, the number of EEG sensors also affected the estimation accuracy when the fMRI/NIRS prior was available. This is not desirable, because it is better to reduce the number of EEG sensors to shorten the experimental time, especially in infants or patients, thus reducing the burden on the subject. However, both simulation and experimental data suggest that the combination of low-density (i.e., 19-ch) EEG and low-resolution (i.e., NIRS) prior yielded better estimation accuracy than high-density (i.e., 64-ch) EEG alone. We therefore propose that both EEG and NIRS data should be measured, at the cost of the number of EEG sensors.

In the present study, the estimated current amplitude in the experimental studies was in the order of  $10^0$ – $10^2$  [pAm/mm<sup>2</sup>], similar to that obtained in previous studies investigating MEG source currents in sensorimotor (Huang et al., 2004; Toda et al., 2011) and visual (Yoshioka et al., 2008) regions, as well as in theoretical studies (Hamalainen et al., 1993). However, this does not indicate that the estimated amplitude is true, because the range is wide and, more importantly, the amplitude in simulation data was underestimated by approximately 10%. Theoretically, the estimation gain decreases as the estimated noise level increases. In addition, the estimation gain should be underestimated if  $m_0$  used in the estimation is smaller than the theoretical value. The underestimation in the simulation study was within the range that can be explained by the effect of the noise level and the value of  $m_0$  used in the estimation (note that we used  $m_0 = 100$  though the theoretical value was 674). Because we set the noise level in the simulation study to be similar to that in real EEG data, the estimated amplitude in the real data may be underestimated.

As described above, the simulation results suggest that the spatial resolution of the prior information affects the accuracy of VBMG. It is therefore important to improve the spatial resolution of NIRS measurement. To this end, the skin blood flow artifact in this study was removed by using short-distance source-detector channels, which improved the accuracy of VBMG; estimation accuracy were low [ $R_{st} = 0.256$  (0.219),  $R_s = 0.288$  (0.273), RMSE = 16.2 (16.6) pAm/mm<sup>2</sup>] without removing the artifacts. However, even if the skin blood artifact is completely rejected, spatial resolution and localization accuracy are limited by measurement geometry to approximately 3 cm, comparable with the distance between emitters and detectors, in the plane parallel to the scalp. The resolution limitation can be overcome only by performing diffuse optical tomography (DOT), which uses a model of photon migration through the head to obtain the activated source (Boas and Dale, 2005). The use of the NIRS prior obtained with this technique will be a matter for future research.

In the present study, we used cortical and head models constructed from structural MRI, for the following reasons. The lead field of EEG is affected by the conductivity of the skull and scalp much more than that of MEG. Therefore, interpreting EEG signals requires more precise knowledge of the thickness and conductivity of the tissues in the head. In the spherical model, concentric inhomogeneities do not affect the magnetic field at all, whereas they have to be taken into account in the analysis of EEG data (Hamalainen et al., 1993). In addition, as described in a previous study (Sato et al., 2004), the introduction of structural MRI information significantly

increased the spatial resolution of VBMG in MEG analysis. Taken together, these findings indicate that structural MRI should be included as prior information, particularly in EEG analysis. Fortunately, once MRI data is obtained, it can be used repeatedly. In cases where MRI is unavailable, computed tomography (CT) images may provide a suitable alternative. Henson et al. (2009) compared the effect of head-models on estimation accuracy, within the context of two types of spatial prior on the sources: a single prior corresponding to a standard L2-minimum-norm (MNM) inversion, or multiple sparse priors (MSP). The study reported (1) no evidence that a cortical model derived from an individual's MRI was superior to a cortical model inverse-normalized from the Montreal Neurological Institute (MNI) template, but (2) clear evidence that a Boundary Element Model (BEM) head-model was superior to spherical head-models, provided individually-defined inner skull and scalp models were used. Because skull and scalp models can be derived from an individual's CT image, sufficient estimation accuracy may be achieved by using a combination of individual's CT images and template MRI.

In the analysis of experimental data, we used different statistics to impose spatial priors, when derived from either fMRI or NIRS; the *t*-statistic was used in the case of fMRI prior, but peak amplitude was used in the case of NIRS prior. We did not use *t*-values in the NIRS case for two reasons. First, *t*-statistics are not commonly used in the field of NIRS but they are commonly used in fMRI studies. Second, estimation with 64-ch EEG was better for the NIRS prior based on the peak values [averages (SDs) for evaluation metrics are as follows;  $R_{st} = 0.40$  (0.14),  $R_s = 0.45$  (0.13), RMSE = 5.34 (2.32)] than for the NIRS prior based on the *t*-value [ $R_{st} = 0.36$  (0.11),  $R_s = 0.42$  (0.09), RMSE = 5.79 (2.67)].

VBMG with fMRI or NIRS priors revealed that current density in M1 started increasing around 500 ms before movement onset, peaking just before movement onset, and exhibiting lateralization towards the hemisphere contralateral to the movement. These results are consistent with the knowledge that the NS and MP components of MRPs arise from the contralateral M1 and are related to late preparation and execution of movements. The estimation with VBMG from EEG with NIRS prior is therefore a useful tool for investigating motor-related cortical activity in both normal individuals and stroke patients. Activation of the SMA was weak in the present study. This may have been because we eliminated the BP component, arising predominantly from the SMA, using a high-pass filter, and/or because both fMRI and NIRS information as hierarchical priors of VBMG exhibited little SMA activity. Future studies using simultaneous measurements of EEG and NIRS should make it possible to focus on SMA activity related to the early preparation stage, by using slow-rate movement tasks with an event-related design.

## Acknowledgments

We thank Dr. Satoshi Tanaka for helping us to carry out fMRI experiments. We also thank Dr. Ken-ichi Morishige at Toyama Prefectural University and Dr. Taku Yoshioka at ATR Neural Information Analysis Laboratories for helpful comments. This research was supported by the Strategic Research Program for Brain Sciences (SRPBS), and partially supported by the National Institute of Information and Communications Technology (NICT) and Funding Program for Next Generation World-Leading Researchers.

## Appendix A. Variational Bayesian Multimodal EncephaloGraphy (VBMG)

### EEG forward model

VBMG (Sato et al., 2004; Yoshioka et al., 2008) is one of the distributed source methods, in which the cortical current is modeled by a number of current dipoles with fixed position and orientation.

The cortical dipole current directions were assumed to be perpendicular to the cortical surface. Because EEG data contains not only brain activity but also the effects of artifacts, we also placed dipoles on artifactual sources, eyes, according to previous studies (Fujiwara et al., 2009; Morishige et al., 2009). The artifact current dipoles were located in the center of the right and left eyeballs.

According to the previous study (Fujiwara et al., 2009), the EEG forward model, that is, the relationship between the amplitude of current dipoles and observed electric potential at time point *t*, is given by

$$\mathbf{E}(\mathbf{t}) = \mathbf{G}_{\text{brain}} \cdot \mathbf{J}_{\text{brain}}(\mathbf{t}) + \mathbf{G}_{\text{eye}} \cdot \mathbf{J}_{\text{eye}}(\mathbf{t}) + \boldsymbol{\varepsilon}(\mathbf{t}), \quad (1)$$

where  $\mathbf{E}(\mathbf{t})$  is an  $N \times 1$  vector for the observed electric potential (i.e., EEG),  $\mathbf{J}_{\text{brain}}(\mathbf{t})$  is an  $L \times 1$  vector for the cortical current, and  $\mathbf{J}_{\text{eye}}(\mathbf{t})$  is an  $K \times 1$  vector for the eye current. Constants *N*, *L*, and *K* denote the number of sensors, brain current sources, and eye current sources, respectively.  $\mathbf{G}_{\text{brain}}$  is an  $N \times L$  matrix, and  $\mathbf{G}_{\text{eye}}$  is an  $N \times K$  matrix.  $\mathbf{G}_{\text{brain}}$  and  $\mathbf{G}_{\text{eye}}$  are called the lead field matrices, whose *n*, *l*-th and *n*, *k*-th elements describe the sensitivity of the *n*-th sensor when a unit dipole is set on the *l*-th and *k*-th locations. The lead field matrix for brain current sources  $\mathbf{G}_{\text{brain}}$  is calculated using a three-shell BEM head model (brain, skull, and scalp). The lead field matrix for eye current sources  $\mathbf{G}_{\text{eye}}$  is calculated by the Biot-Savart equation. Observation noise  $\boldsymbol{\varepsilon}(\mathbf{t})$  is assumed to obey a Gaussian distribution with zero mean. Note that, in the simulation study, we did not assume currents induced by eyes and therefore the forward model did not have the right hand second term.

### Estimating current variance with VBMG

We use the hierarchical Bayesian estimation for current estimation (Sato et al., 2004). Assuming that the EEG observation noise follows Gaussian distribution with a spherical covariance, the EEG forward model (Eq. (1)) leads to the likelihood function:

$$P(E|\mathbf{J}_{\text{brain}}, \mathbf{J}_{\text{eye}}) \propto \exp \left[ -\frac{\beta}{2} \|E - \mathbf{G}_{\text{brain}} \cdot \mathbf{J}_{\text{brain}} - \mathbf{G}_{\text{eye}} \cdot \mathbf{J}_{\text{eye}}\|^2 \right], \quad (2)$$

where  $\beta$  denotes the inverse of the unknown variance of EEG observation noise. As for the prior probability distribution  $P_0(\mathbf{J}_{\text{brain}}, \mathbf{J}_{\text{eye}}|\boldsymbol{\alpha}_{\text{brain}}, \boldsymbol{\alpha}_{\text{eye}})$ , we assume a Normal prior:

$$P_0(\mathbf{J}|\boldsymbol{\alpha}_{\mathbf{J}}) \propto \exp \left[ -\frac{1}{2} \sum_{t=1}^T \mathbf{J}'(t) \cdot \mathbf{A}_{\mathbf{J}} \cdot \mathbf{J}(t) \right], \quad (3)$$

where  $\mathbf{J}' = [\mathbf{J}_{\text{brain}}' \mathbf{J}_{\text{eye}}']$ ,  $\mathbf{A}_{\mathbf{J}} = \text{diag}(\boldsymbol{\alpha}_{\mathbf{J}})$ , and  $\boldsymbol{\alpha}_{\mathbf{J}} = [\boldsymbol{\alpha}_{\text{brain}}' \boldsymbol{\alpha}_{\text{eye}}']$ . Vectors  $\boldsymbol{\alpha}_{\text{brain}}$  and  $\boldsymbol{\alpha}_{\text{eye}}$  are the inverse variances of the brain and eye current sources, respectively.

The current inverse variance parameters  $\boldsymbol{\alpha}_{\mathbf{J}}$  are estimated by introducing an Automatic Relevance Determination (ARD) hierarchical prior (Neal, 1996):

$$P_0(\boldsymbol{\alpha}_{\mathbf{J}}) = \prod_i \Gamma(\alpha_{J(i)} | \bar{\alpha}_{J0(i)}, \gamma_{J0(i)}), \quad (4)$$

where  $\Gamma(\alpha | \bar{\alpha}, \gamma)$  represents the Gamma distribution with mean  $\bar{\alpha}$  and degree of freedom  $\gamma$ .  $\bar{\alpha}_{J0(i)}$  is a mean prior of an inverse current variance and  $\gamma_{J0(i)}$  controls the spread of the distribution of corresponding  $\alpha_{J(i)}$ . A prior current variance  $\bar{V}_{J0} \equiv \bar{\alpha}_{J0}^{-1}$  represents the prior information about a current intensity. For large  $\bar{V}_{J0}$ , estimated current  $\bar{J}$  could be large. For small  $\bar{V}_{J0}$ , estimated current  $\bar{J}$  tends to be small. The parameter  $\gamma_{J0}$  reflects the reliability (confidence) of corresponding  $\bar{\alpha}_{J0}$ . For very small  $\gamma_{J0}$ , the distribution spreads uniformly, and prior information  $\bar{V}_{J0}$  does not affect the current estimation (non-informative prior). In contrast, for large  $\gamma_{J0}$ , since the distribution is concentrated around prior mean  $\bar{V}_{J0}$ , prior information  $\bar{V}_{J0}$  influences the current estimation more strongly.

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 (for VB method, see Sato et al., 2004).

The  $\bar{\nu}_{j0}$  and  $\gamma_{j0}$  for cortical dipoles (i.e.,  $\bar{\nu}_{brain0}$  and  $\gamma_{brain0}$ ) can be determined depending on whether fMRI/NIRS data are available. Here, we will explain how these were determined in the experimental data analysis.

When fMRI/NIRS data are available, the  $\bar{\nu}_{brain0}$  and  $\gamma_{brain0}$  were determined in the following way. We imposed fMRI/NIRS information on the prior variance parameter as:  $\bar{\nu}_{brain0(l)} = \nu_0 + (m_0 - 1) \cdot \nu_0 \cdot a_{(l)}^2$ , where  $\nu_0$  was the baseline of the prior current variance and estimated from the baseline interval of the EEG data by the minimum norm estimation (Wang et al., 1992),  $a_{(l)}$  was normalized activity data (*t*-value for the fMRI case and peak value of the artifact-free  $\Delta[\text{OxyHb}]$  data for the NIRS case) on the *l*-th vertex, and  $m_0$  was a variance magnification parameter and set to 100 for all dipoles. We set the confidence parameter  $\gamma_{brain0}$  to 10 for all dipoles.

When fMRI/NIRS data are not available, we used uniform spatial prior (i.e.,  $\bar{\nu}_{brain0} = m_0 \nu_0$  for all cortical dipoles) where the magnification parameter  $m_0$  was set to 100, and the confidence parameter  $\gamma_{brain0}$  was set to 10 for all dipoles (i.e., non-informative prior).

The  $\bar{\nu}_{j0}$  and  $\gamma_{j0}$  for eye dipoles (i.e.,  $\bar{\nu}_{eye0}$  and  $\gamma_{eye0}$ ) were set to  $10^2$  [nAm]<sup>2</sup> and  $10^{4.5}$ , respectively, according to a previous study (Morishige et al., 2009).

#### Spatial smoothness constraint

We also assumed a spatial smoothness constraint on the current distribution along with the cortical surface (Shibata et al., 2008). For this purpose, we used a smoothing filter matrix  $W_{ij} \propto \exp(-d_{ij}^2/R^2)$ , where  $d_{ij}$  is the distance between *i*-th and *j*-th current dipoles. The smoothing radius parameter, *R*, was set to 6 mm. By introducing an auxiliary variable  $Z(t)$  and letting

$$\mathbf{J}_{\text{brain}}(t) = W\mathbf{Z}_{\text{brain}}(t), \quad (5)$$

Eq. (1) can be replaced by

$$\mathbf{E}(t) = \tilde{\mathbf{G}}_{\text{brain}} \cdot \mathbf{Z}_{\text{brain}}(t) + \mathbf{G}_{\text{eye}} \cdot \mathbf{J}_{\text{eye}}(t) + \boldsymbol{\varepsilon}(t) \quad (6)$$

where  $\tilde{\mathbf{G}} \equiv G\mathbf{W}$  is a smoothed lead field matrix. Therefore, the EEG inverse problem is changed to estimate  $Z(t)$  instead of  $J(t)$ , with the smoothed lead field matrix  $\tilde{G}$ . After estimating  $Z(t)$ , the actual current amplitude  $J(t)$  was calculated using Eq. (5).

#### Appendix B. Unbiased estimation of the AUC

For each threshold  $\beta$ , the number of estimated or simulated inactive dipoles was much greater than the number of active dipoles. To interpret the AUC as a detection accuracy index, the same number of active and inactive dipoles should be provided in the ROC analysis (Grova et al., 2006).

Let  $\Theta$  define the set of all cortical dipoles ( $\text{card}(\Theta) = p$ ) and  $\Theta_a$  define the set of simulated active dipoles ( $\text{card}(\Theta_a) = p_a$ ). To provide less biased estimation of ROC parameters, we randomly chose a set  $\Theta_f$  of  $p_a$  fictive dipoles among the  $p - p_a$  remaining inactive dipoles, i.e.,  $\Theta_f \subseteq \Theta \setminus \Theta_a$ . Less biased ROC curves and AUC were then estimated for those sets of  $p_a$  pairs of active  $\Theta_a$  and fictive  $\Theta_f$  dipoles.

However, the false positive rate may now be greatly underestimated because many spurious active dipoles are missed by the random drawing of  $\Theta_f$  in  $\Theta \setminus \Theta_a$ . We thus adopted two strategies to choose  $\Theta_f$ , optimizing the false positive detection. We split the estimation of detection accuracy into two components, one dedicated to the focalization ability of the method and the second dedicated to false positive detection far from the simulated active dipoles.

For the localization ability, a criterion  $AUC_{\text{close}}$  was estimated by choosing the fictive dipoles within a 5-cm radius sphere<sup>5</sup> centered on the simulated active dipole, i.e.,  $\Theta_f \subseteq \Theta_{\text{sphere}}^5(\Theta_a) \setminus \Theta_a$ , where  $\Theta_{\text{sphere}}^5(\Theta_a)$  denotes the set of dipoles within a 5-cm radius sphere centered on the simulated active dipole. For the detection of false positive far from the simulated source, a criterion  $AUC_{\text{far}}$  was estimated by choosing the fictive dipoles among dipoles far from the simulated active source, i.e., within the complementary set of  $\Theta_{\text{sphere}}^5(\Theta_a)$ :  $\Theta_f \subseteq \Theta \setminus \Theta_{\text{sphere}}^5(\Theta_a)$ .

The proposed index of detection accuracy AUC was then defined as the mean of the previous criteria:

$$AUC = \frac{1}{2}(AUC_{\text{close}} + AUC_{\text{far}}).$$

To obtain consistent measurements not sensitive to one particular choice of  $\Theta_f$ , AUC was estimated over 100 independent drawings of  $\Theta_f$ . The mean AUC over those 100 trials will be presented here as an index of detection accuracy.

#### Appendix C. How to determine the cortical projection point of each NIRS channel

When the NIRS channel is on a gyrus, the cortical vertex that is the closest from the channel should be found as a single point (see Fig. 2F of Okamoto et al., 2004). In this case, we defined the vertex as a cortical projection point of the channel. When the NIRS channel is on a sulcus, we selected two to four cortical vertices on the gyri that sandwich or surround the sulcus and represent local minimum distances (see Figs. 2G–I of Okamoto et al., 2004). To the line defined by the two minimum-distance vertices or to the plane defined by the three minimum-distance vertices, we drew a normal line from the channel. We defined a cortical projection point as the point at which the normal line intersects the cortical surface (see Figs. 2F–I of Okamoto et al., 2004). In some extreme cases where we found four minimum-distance vertices, we drew two lines defined by two neighboring, non-overlapping vertices. We next drew normal lines from the channel to the two lines and obtained two separate intersections. By reducing four vertices to two, we could apply the procedure above. Similarly, we tested the other combination of two lines, and applied the average value of the two combinations.

#### Appendix D. Interpolation method of the NIRS data

The inverse distance weighting (IDW) method (Shepard, 1968) was used to interpolate a NIRS value ( $a_{(l)}$ ) at the *l*-th cortical vertex (interpolation point) on the brain using known normalized NIRS values at scattered known neighborhood cortical projection points (Takeuchi et al., 2009). The value  $a_{(l)}$  is obtained based on a following interpolating function:

$$a_{(l)} = \frac{\sum_{i=1}^n w_i f_i}{\sum_{i=1}^n w_i}$$

where *n* is the number of the neighborhood cortical projection points,  $w_i$  is a weight function assigned to each neighborhood cortical projection point, and  $f_i$  is a known normalized NIRS value at a known cortical projection point of the corresponding NIRS channel.

<sup>5</sup> Grova et al. (2006) did not adopt the 5-cm radius sphere centered on the source, but adopted the 10-th neighborhood of the source to calculate  $AUC_{\text{close}}$  and  $AUC_{\text{far}}$ . However, the region including the 10-th neighborhood will be different between our case and the case of Grova et al. (2006), because the density of cortical dipoles is different. Table 1 of Grova et al. (2006) suggests that the 5-cm radius sphere is comparable to the 10-th neighborhood. Therefore, we adopted the 5-cm radius sphere.

The weight function is obtained from the following equation:

$$w_i = \frac{1}{d_i^2}$$

where  $d_i$  is distance between the known cortical projection point and the interpolation point. The number of neighborhood cortical projection points determines how many NIRS channel points with the known NIRS values are included in the IDW. This number can be specified in terms of a radius (2 cm in this paper), where a center of the circle is the given interpolation point with  $a_{(l)}$ . Thus, the vertices more than 2 cm away from all the projection points had no NIRS values.

## References

- Boas, D.A., Dale, A.M., 2005. Simulation study of magnetic resonance imaging-guided cortically constrained diffuse optical tomography of human brain function. *Appl. Opt.* 44, 1957–1968.
- Cheyne, D., Bakhtazad, L., Gaetz, W., 2006. Spatiotemporal mapping of cortical activity accompanying voluntary movements using an event-related beamforming approach. *Hum. Brain Mapp.* 27, 213–229.
- Cunnington, R., Iansek, R., Bradshaw, J.L., Phillips, J.G., 1996. Movement-related potentials associated with movement preparation and motor imagery. *Exp. Brain Res.* 111, 429–436.
- Dale, A.M., Fischl, B.R., Sereno, M.I., 1999. Cortical surface-based analysis. I: segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, 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.
- Daunizeau, J., Grova, C., Mattout, J., Marrelec, G., Clonda, D., Goulard, B., Pelegri-Issac, M., Lina, J.-M., Benali, H., 2005. Assessing the relevance of fMRI-based prior in the EEG inverse problem: a Bayesian model comparison approach. *IEEE Trans. Signal Process.* 53, 3461–3472.
- Fujiwara, Y., Yamashita, O., Kawakami, D., Doya, K., Kawato, M., Toyama, K., Sato, M.A., 2009. A hierarchical Bayesian method to resolve an inverse problem of MEG contaminated with eye movement artifacts. *Neuroimage* 45, 393–409.
- Grova, C., Daunizeau, J., Lina, J.-M., Benar, C.G., Benali, H., Gotman, J., 2006. Evaluation of EEG localization methods using realistic simulations of interictal spikes. *Neuroimage* 29, 734–753.
- Grova, C., Daunizeau, J., Kobayashi, E., Bagshaw, A.P., Lina, J.-M., Dubeau, F., Gotman, J., 2008. Concordance between distributed EEG source localization and simultaneous EEG-fMRI studies of epileptic spikes. *Neuroimage* 39, 755–774.
- Hamalainen, M., Hari, R., Ilmoniemi, R.J., Knuttila, J., Lounasmaa, O.V., 1993. Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 65, 413–497.
- Hari, R., 1991. On brain's magnetic responses to sensory stimuli. *J. Clin. Neurophysiol.* 8, 157–169.
- Henson, R.N., Mattout, J., Phillips, C., Friston, K.J., 2009. Selecting forward models for MEG source-reconstruction using model-evidence. *Neuroimage* 46, 168–176.
- Henson, R.N., Flandin, G., Friston, K.J., Mattout, J., 2010. A parametric empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction. *Hum. Brain Mapp.* 31, 1512–1531.
- Hoshi, Y., Kobayashi, N., Tamura, M., 2001. Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. *J. Appl. Physiol.* 90, 1657–1662.
- Huang, M.X., Harrington, D.L., Paulson, K.M., Weisend, M.P., Lee, R.R., 2004. Temporal dynamics of ipsilateral and contralateral motor activity during voluntary finger movement. *Hum. Brain Mapp.* 23, 26–39.
- Kajihara, S., Ohtani, Y., Goda, N., Tanigawa, M., Ejima, Y., Toyama, K., 2004. Wiener filter-magnetoencephalography of visual cortical activity. *Brain Topogr.* 17, 13–25.
- Kunieda, T., Ikeda, A., Ohara, S., Yazawa, S., Nagamine, T., Taki, W., Hashimoto, N., Shibusaki, H., 2000. Different activation of presupplementary motor area, supplementary motor area proper, and primary sensorimotor area, depending on the movement repetition rate in humans. *Exp. Brain Res.* 135, 163–172.
- Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239.
- Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Gravé de Peralta, R., 2004. EEG source imaging. *Clin. Neurophysiol.* 115, 2195–2222.
- Morishige, K., Kawakami, D., Yoshioka, T., Sato, M.A., Kawato, M., 2009. Artifact removal using simultaneous current estimation of noise and cortical sources. *Lecture notes in computer science: ICONIP 2008*. Springer-Verlag, Berlin, pp. 335–342.
- Mosher, J.C., Lewis, P.S., Leahy, R.M., 1992. Multiple dipole modeling 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.
- Okada, E., Firbank, M., Schweiger, M., Arridge, S.R., Cope, M., Delpy, D.T., 1997. Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head. *Appl. Opt.* 36, 21–31.
- Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., Oda, I., Isobe, S., Suzuki, T., Kohyama, K., Dan, I., 2004. Three-dimensional probabilistic anatomical crano-cerebral correlation via the international 10–20 system oriented for transcranial functional brain mapping. *Neuroimage* 21, 99–111.
- Ou, W., Nummenmaa, A., Ahneninen, J., Belliveau, J.W., Hamalainen, M.S., Golland, P., 2010. Multimodal functional imaging using fMRI-informed regional EEG/MEG source estimation. *Neuroimage* 52, 97–108.
- 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.
- Rosa, M.J., Daunizeau, J., Friston, K.J., 2010. EEG-fMRI integration: a critical review of biophysical modeling and data analysis approaches. *J. Integr. Neurosci.* 9, 453–476.
- Sato, M.A., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. *Neuroimage* 23, 806–826.
- Schmidt, D.M., George, J.S., Wood, C.C., 1999. Bayesian inference applied to the electromagnetic inverse problem. *Hum. Brain Mapp.* 7, 195–212.
- Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 23rd National Conference ACM*. ACM Press, New York, pp. 517–524.
- Shibata, K., Yamagishi, N., Goda, N., Yoshioka, T., Yamashita, O., Sato, M.A., Kawato, M., 2008. The effects of feature attention on prestimulus cortical activity in the human visual system. *Cereb. Cortex* 18, 1664–1675.
- Strangman, G., Culver, J.P., Thompson, J.H., Boas, D.A., 2002. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage* 17, 719–731.
- Takeuchi, M., Hori, E., Takamoto, K., Tran, A.H., Satoru, K., Ishikawa, A., Ono, T., Endo, S., Nishijo, H., 2009. Brain cortical mapping by simultaneous recording of functional near infrared spectroscopy and electroencephalograms from the whole brain during right median nerve stimulation. *Brain Topogr.* 22, 197–214.
- Toda, A., Imamizu, H., Kawato, M., Sato, M.A., 2011. Reconstruction of two-dimensional movement trajectories from selected magnetoencephalography cortical currents by combined sparse Bayesian methods. *Neuroimage* 54, 892–905.
- Toma, K., Honda, M., Hanakawa, T., Okada, T., Fukuyama, H., Ikeda, A., Nishizawa, S., Konishi, J., Shibusaki, H., 1999. Activities of the primary and supplementary motor areas increase in preparation and execution of voluntary muscle relaxation: an event-related fMRI study. *J. Neurosci.* 19, 3527–3534.
- Waberski, T.D., Buchner, H., Lehnertz, K., Hufnagel, A., Fuchs, M., Beckmann, R., Rienacker, A., 1998. The properties of source localization of epileptiform activity using advanced head modelling and source reconstruction. *Brain Topogr.* 10, 283–290.
- Wang, J.Z., Williamson, S.J., Kaufman, L., 1992. Magnetic source images determined by a lead-field analysis: the unique minimum-norm least-squares estimation. *IEEE Trans. Biomed. Eng.* 39, 665–675.
- Yoshioka, T., Toyama, K., Kawato, M., Yamashita, O., Nishina, S., Yamagishi, N., Sato, M.A., 2008. Evaluation of hierarchical Bayesian method through retinotopic brain activities reconstruction from fMRI and MEG signals. *Neuroimage* 42, 1397–1413.