Journal of Signal Processing, Vol. 8, No. 6, pp. 461-472, November 2004

PAPER

Evaluation of Brain Source Separation for MEG Data Applying JADE, Fast-ICA and Natural Gradient-Based Algorithm with Robust Pre-Whitening Technique

Yoshio Konno[1], Jianting Cao[2,3], Tsunehiro Takeda[4], Hiroshi Endo[5], Takayuki Arai[1] and Mamoru Tanaka[1]

[1] Department of Electrical and Electronics Engineering, Sophia University
7-1 Kioicho, Chiyoda-ku, Tokyo 102-8554, Japan
[2] Department of Electronics Engineering, Saitama Institute of Technology
1690 Fusaiji, Okabe, Saitama 369-0293, Japan
[3] Lab. for Advanced Brain Signal Processing, RIKEN
2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
[4] Department of Complexity of Science and Engineering, Graduate School of Tokyo University
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[5] National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Tsukuba-shi-higashi, Ibaraki 305-8566, Japan
Phone: +81-3-3238-3878, Fax: +81-3-3238-3321, E-mail : yo-konno@sophia.ac.jp

Abstract **Independent component analysis (ICA) has been applied to magnetoencephalographic (MEG) data to determine the behavior and localization of brain sources. In this work, both nonaveraged single-trial data and averaged multiple-trial data are analyzed, in order to study the relationship between the performance of decomposition and the number of averages across data trials. To evaluate the performance of source decomposition, 1) the ratio of source decomposition, which indicates whether source components will be decomposed or not, 2) the power of decomposed components, which is calculated as the covariance of decomposed components, and 3) the accuracy of source estimation, are demonstrated. In addition, a number of existing ICA algorithms such as JADE, Fast-ICA, and the natural gradient-based algorithm with a robust pre-whitening technique are used to decompose MEG data. Our results show the relationship between the accuracy of source decomposition and the number of averages and, by applying the ICA approach, the number of averages can be reduced.**

Keywords: magnetoencephalography (MEG), independent component analysis (ICA), JADE, fast-ICA, natural gradient-based algorithm, robust pre-whitening technique, source localization

1. Introduction

Electroencephalography (EEG) and MEG are powerful and noninvasive techniques for measuring human brain activity with a high temporal resolution. The motivation for studying EEG/MEG data analysis is to extract the essential features of measured data and represent them as corresponding human brain functions. ICA or blind source separation (BSS) [1]-[7] has been applied to MEG data to determine the behavior and localization of brain sources, by many researchers focused on neural networks and statistical signal processing [8]-[20]. When detecting a MEG signal by applying ICA, spontaneous and environmental noises may seriously affect the recorded data because the magnetic fields of brain signals are weak, particularly in the case of nonaveraged data.

The most widely used technique for reducing instrumental and environmental noises is to take an average across many stimulation trials. In fact, when applying ICA to MEG data, most researchers have treated averaged data. Makeig et al. [8] first applied the Informax algorithm [7] to study averaged EEG data with the event-related potentials (ERP) task. Vigário and coworkers [9], [15] applied FastICA [6] to study averaged EEG and MEG data with auditory evoked fields (AEF) and somatosensory evoked fields (SEF), and further localized the sources by us-

ing magnetic field patterns. Ikeda and Toyama [13] and Ikeda [14] developed a factor analysis algorithm with applying JADE [4] to study averaged MEG data with phantom and visual tasks, and further visualized the decomposed components by using the spatial filtering technique. Other researchers applied various kinds of ICA algorithms [1]-[7] to extract neural activities or/and artifacts from EEG or MEG data [10], [11], [16], [18], [20].

However, by taking an average, important information is lost, making it advantageous to decrease the number of averages taken across data trials. The disadvantage of having fewer averages is that because the SNR is very poor, the decomposition of a low-power source signal from recorded data is still influenced by noise. Recently, some researchers have studied non-averaged, single-trial data with the aim of retaining the information lost by averaging [15], [17], [19], [20]. Jung et al. have studied single-trial EEG records from normal and autistic subjects performing visual selective attention tasks [12].

In this work, both nonaveraged single-trial data and averaged multiple-trial data are analyzed, in order to study the relationship between the accuracy of decomposed brain sources and the number of averages taken. To evaluate the decomposition performance, 1) the ratio of source decomposition, 2) the power of decomposed components, and 3) the accuracy of source estimation, are demonstrated.

When applying ICA to physiological data, most researchers have used real, measured, physiological data and evaluated the decomposed components from a neuroscience perspective [8]-[17], [19]. In this study, a synthesized MEG data set is used, which includes an artificial evoked field and real, measured brain data. The behavior of our data set is similar to AEF. The main advantage of our data set is that the dipole locations and the dynamics of evoked responses are known in advance, which facilitates the evaluation of the decomposed components.

2. Data Analysis Model

In this section, the model for applying ICA to MEG data is described. Since neurons in different physical regions of the brain are simultaneously active during experimentation, currents in a group of neurons situated close together can be equivalently represented by a single current dipole called a neural source. Multiple neural sources, including evoked responses, are superimposed on each other and are detected by sensors arranged on the scalp. During measurement, some undesirable components such as environmental interferences and instrumental noises are recorded at the same time as the neural sources.

The particular ICA model considered in this paper is

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{e}(t) \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), \cdots, x_m(t)]^T$ represents the transposition of m observations at time t. $\mathbf{s}(t) = [s_1(t), \cdots, s_n(t)]^T$ represents n unknown source components which contain the neural sources and the environmental interference sources. Neural sources include the evoked response derived by the auditory or visual stimulus and include spontaneous noises such as spontaneous α-wave components. Environmental interference sources include electrical power interference and artifacts such as eye blinks. Since the environmental interferences contribute to at least two sensors, they are regarded as common components and are defined as "source noise". Another kind of component, the additive noise represented by $\mathbf{e}(t) = [e_1(t), \cdots, e_m(t)]^T$ is called a unique component. Since it only contributes to one sensor, it is defined as "sensor noise". Since human tissue and skull do not attenuate magnetic fields in MEG, $\mathbf{A} \in \mathbf{R}^{m \times n} = (a_{ij})$ can be represented by a numerical matrix whose element a_{ij} is simply a quantity related to the physical distance between the i-th sensor and the j-th source.

In the model, the sources $\mathbf{s}$ and their number n, additive noise $\mathbf{e}$ and matrix $\mathbf{A}$ are unknown but the sensor signals $\mathbf{x}$ are accessible. It is assumed that the components of $\mathbf{s}$ are mutually statistically independent, as well as being statistically independent of the noise components $\mathbf{e}$. Moreover, the noise components $\mathbf{e}$ themselves are assumed to be mutually independent.

Summarizing this model, there are two kinds of noises that must be reduced in the data analysis. The first kind of noise is sensor noise (additive noise). It is usually generated from instruments such as individual sensors, and contaminates the observed signals. Their power will be reduced by using the robust pre-whitening technique in the pre-processing stage. The second kind of noise is source noise. It will be discarded after the source decomposition by means of the ICA approach. After removing these noises from the data, estimated components such as the evoked responses will be obtained.

3. Pre-Whitening Technique

3.1 Robust pre-whitening technique

In this subsection, the robust pre-whitening technique [19] is described. This technique is very capable of reducing the power of sensor noise (additive noise). The aim of this task is to find the estimate lower dimension orthogonal matrix and covariance matrix of the noise in the orthogonalization procedure.

Let us rewrite Eq. (1) in a data matrix form as

$$\mathbf{X}_{(m \times N)} = \mathbf{A}_{(m \times n)}\mathbf{S}_{(n \times N)} + \mathbf{E}_{(m \times N)} \tag{2}$$

where N denotes the number of data samples. When the sample size is sufficiently large, the covariance matrix of the observed data in the mixing model $\mathbf{\Sigma}$ can be written as $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T + \mathbf{\Psi}$, where $\mathbf{\Psi}$ is a diagonal matrix of the additive noise $\mathbf{E}$. Also the covariance matrix of the observed data recorded by sensors can be given by $\mathbf{C} = \mathbf{X}\mathbf{X}^T$.

For the robust pre-whitening technique, $\mathbf{A}$ can be estimated as

$$\widehat{\mathbf{A}} = \mathbf{U}_{\hat{n}}\mathbf{\Lambda}_{\hat{n}}^{\frac{1}{2}} \tag{3}$$

by applying the standard principal component analysis (PCA) approach, where $\mathbf{\Lambda}_{\hat{n}}$ is a diagonal matrix whose elements are eigenvalues of $\mathbf{C}$, the columns of $\mathbf{U}_{\hat{n}}$ are the corresponding eigenvectors, and $\hat{n}$ is the estimated number of sources.

To estimate $\mathbf{\Psi}$, we fit $\mathbf{\Sigma}$ to $\mathbf{C}$ using the eigenvalue decomposition method. In this case, the cost function is obtained as $L(\mathbf{A}, \mathbf{\Psi}) = \mathrm{tr}[\mathbf{\Sigma} - \mathbf{C}]^2$. It is minimized by $\partial L(\mathbf{A}, \mathbf{\Psi})/\partial \mathbf{\Psi} = 0$, whereby the estimate noise variance $\mathbf{\Psi}$ is obtained as

$$\widehat{\mathbf{\Psi}} = \mathrm{diag}(\mathbf{C} - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^{\mathrm{T}}) \tag{4}$$

where the estimate $\widehat{\mathbf{A}}$ is obtained in the same manner as shown in Eq. (3). Using these estimates, $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{\Psi}}$, the transform matrix can be obtained as

$$\mathbf{Q} = [\widehat{\mathbf{A}}^T\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{A}}]^{-1}\widehat{\mathbf{A}}^T\widehat{\mathbf{\Psi}}^{-1} \tag{5}$$

Using this, the new set of data, the orthogonal observation, can be obtained by

$$\mathbf{z}(t) = \mathbf{Q}\mathbf{x}(t) \tag{6}$$

Note that the covariance matrix is $\mathrm{E}[\mathbf{z}\mathbf{z}^{\mathrm{T}}] = \mathbf{I}_{\hat{n}} + \mathbf{Q}\mathbf{\Psi}\mathbf{Q}^{\mathrm{T}}$, which implies that the source signals in a subspace are de-correlated.

A similar noise reduction approach that involves the application of factor analysis (FA) to the decomposition of MEG data has been reported in [13]. Both this method and ours take additive noises into account, but with our robust pre-whitening technique, the distribution of additive noises is not restricted. Therefore, our technique is more robust and effective for data with non-Gaussian noise such as the outlier.

3.2 Criterion for choosing the number of sources

In this subsection, the criterion for choosing the number of sources is explained. The total variance accounted for by all the orthogonal components is given by $\mathrm{tr}\mathbf{\Lambda}_{\mathrm{k}}$, that is, the sum of the variances of all the linear combinations:

$$\mathrm{tr}\mathbf{\Lambda}_{\mathrm{k}} = \lambda_1 + \lambda_2 + \cdots + \lambda_{\mathrm{k}} \tag{7}$$

where λ_k denotes the k-th largest eigenvalue of $\mathbf{C}$. That will be the same as the total variance of all the variables if $k = m$, because $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ and $\mathrm{tr}\mathbf{C} = \mathrm{tr}(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}) = \mathrm{tr}\mathbf{\Lambda}$. The variance left unexplained by the components, that is, the total residual variance, is therefore $\lambda_{k+1} + \lambda_{k+2} + \cdots + \lambda_m$. Consequently, the total variance of all the variables is the sum of the variance explained by the k components and the residual variance.

The criterion for choosing the number of sources is defined in such a manner as to account for a satisfactory amount of total variation. The way to start is to compute the cumulative percentage variance contribution obtained for successive values of $\hat{n} = 1, 2, \cdots$, and to stop when this is sufficiently large, for example, larger than 75 %, 90 %, or 95 %. The cumulative percentage variance contribution for a given value of $\hat{n}$ is computed as

$$c_k = \frac{\sum_{k=1}^{\hat{n}} \lambda_k}{\mathrm{tr}\mathbf{C}} \times 100 \tag{8}$$

4. Independent Component Analysis

It should be noted that the robust pre-whitening technique is needed to reduce the power of sensor noises and the number of parameters, but it is insufficient to obtain the independent components since an orthogonal matrix in general contains additional degrees of freedom. Therefore, the remaining parameters must be further estimated by using an ICA algorithm. In other words, the power of the sensor noise has been reduced by using the robust pre-whitening technique, but some source noises are still overlapped, and those will need to be eliminated by means of the ICA procedure. After removing these noises from the data, we will obtain the estimated source components, such as the evoked responses.

After pre-processing and ICA, the decomposed independent sources $\mathbf{y} \in \mathbf{R}^n$ can be obtained from a linear transformation as

$$\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t) = \mathbf{W}\mathbf{Q}\mathbf{x}(t) \tag{9}$$

where $\mathbf{W} \in \mathbf{R}^{n \times n}$ is termed the demixing matrix which can be computed by using an ICA algorithm such as JADE [4], Fast-ICA [6], the natural gradient-based algorithm [3], or Info-MAX [2].

4.1 JADE algorithm

The joint approximate diagonalization of eigenmatrices (JADE) has been proposed in [4]. The JADE algorithm has two procedures termed orthogonalization in PCA and rotation. We apply the rotation procedure in the JADE algorithm, described below, but instead of the orthogonalization in PCA, we apply the robust pre-whitening technique described in Section **3.1**.

The main advantage of the JADE algorithm is its efficiency. However, it is memory intensive for large dimension matrix calculations. We solve this problem by optimizing the dimensionality in the robust pre-whitening technique.

The rotation procedure in JADE uses matrices $\mathbf{F}(\mathbf{M})$ formulated by a fourth-order cumulant tensor of the outputs with an arbitrary matrix $\mathbf{M}$ as

$$\mathbf{F}(\mathbf{M}) = \sum_{k=1}^{K}\sum_{l=1}^{L} \mathrm{Cum}(z_i, z_j, z_k, z_l) m_{lk} \qquad (10)$$

where $\mathrm{Cum}(\cdot)$ denotes a standard cumulant and m_{lk} is the (l,k)-th element of matrix $\mathbf{M}$. The correct rotation matrix $\mathbf{W}$ can be obtained by diagonalizing the matrix $\mathbf{F}(\mathbf{M})$, namely, $\mathbf{W}\mathbf{F}(\mathbf{M})\mathbf{W}^T$ approximates a diagonal matrix.

4.2 Fast-ICA algorithm

The Fast-ICA algorithm has been proposed in [6]. This algorithm is based on a fixed-point method and is represented by

$$\mathbf{w}^+ = \mathbf{w}(t) - \eta \frac{E[xg(\mathbf{w}(t)^T x)] - \beta \mathbf{w}(t)}{E[g'(\mathbf{w}(t)^T x)] - \beta} \qquad (11)$$

$$\mathbf{w}(t+1) = \frac{\mathbf{w}^+}{||\mathbf{w}^+||} \qquad (12)$$

where $g(y) = y^3$, or $g(y) = \tanh(\mathrm{y})$.

Comparing the algorithms based on the gradient descent methods, the fixed-point algorithm has a higher speed convergence property since the Newton method in block mode is applied. It is easy to apply in data analysis since there is no learning rate parameter that must be adjusted. Furthermore, we can extract independent sources one by one. This means the condition of the prior knowledge of the source number will be more relaxed. In this study, the process is iterated until the weight vector $\mathbf{w}$ converges to a stable value, $\mathbf{w}(t)^T\mathbf{w}(t-1) \leq 0.0001$.

When the power of noise is strong in the data, particularly in the case of nonaveraged single-trial data, the procedure of decorrelation will fail. Therefore, it should be noted that the fixed-point algorithm requires a preliminary sphering of the data. In this study, instead of sphering in PCA, we apply the robust pre-whitening technique described in Section **3.1**.

4.3 Natural gradient-based algorithm

The Kullback-Leibler divergence is the basic ICA tool [3] which measures the mutual stochastic independence of the output signals $y_i(t)$ between the joint probability density function $p_y(\mathbf{y})$ and the marginal probability density function $p_i(y_i)$ as

$$D(\mathbf{y}|\mathbf{W}) = \int p_y(\mathbf{y}) \log \frac{p_y(\mathbf{y})}{\prod_{i=1}^{n} p_i(y_i)} d\mathbf{y} \qquad (13)$$

In Eq. (13), the Kullback-Leibler divergence $D(\mathbf{y}|\mathbf{W}) = 0$, if and only if the independence condition $p_y(\mathbf{y}) = \prod_{i=1}^{n} p_i(y_i)$ is satisfied.

By applying the natural gradient to minimize the Kullback-Leibler divergence, Eq. (13), the general learning rule for updating $\mathbf{W}$ can be developed as

$$\Delta\mathbf{W}(t) = \eta \left[\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(t))\mathbf{y}^T(t)\right] \mathbf{W}(t) \qquad (14)$$

In Eq. (14), $\eta > 0$ is a learning rate, and $\boldsymbol{\varphi}(\cdot)$ is the vector of activation functions, whose optimal components are described as

$$\varphi_i(y_i) = -\frac{d}{dy_i} \log p_i(y_i) = -\frac{\dot{p}_i(y_i)}{p_i(y_i)} \qquad (15)$$

where $\dot{p}_i(y_i) = dp_i(y_i)/dy_i$.

It is noted that some other independently developed algorithms, for example, the extended Info-MAX algorithm [2], can be classified in the same category as the above algorithm, since they have a similar form.

The key point for applying this type of algorithm is to employ a suitable nonlinear function that enables us to separate the mixtures of sub- and super-Gaussian sources. It is desirable that the nonlinear function be robust to the underinfluence of the outlier. In this work, we apply the algorithm with a robust nonlinear function based on the **e**-distribution (exponential family of distributions) and **t**-distribution models [19] as

$$\varphi_i(y_i) = \alpha\lambda_\alpha \,\mathrm{sgn}(y_i)|\lambda_\alpha y_i|^{\alpha-1},\ \kappa_\alpha = \hat{\kappa}_i \leq 0 \qquad (16)$$

$$\varphi_i(y_i) = \frac{(1+\beta)y_i}{y_i^2 + \frac{\beta}{\lambda_\beta^2}},\ \kappa_\beta = \hat{\kappa}_i > 0 \qquad (17)$$

In Eqs. (16) and (17), α and β are the parameters which can be obtained by estimating the kurtosis $\hat{\kappa}_i$ of output signal y_i, and λ_α and λ_β are the scaling constants for normalizing the variance.

Here the procedure of this algorithm is described. First, the kurtosis is calculated as $\hat{\kappa}_i = \hat{m}_4/\hat{m}_2^2 - 3$, where the 2nd- and 4th-order moments are estimated by using $\hat{m}_j(t) = [1-\eta(t)]\hat{m}_j(t-1) + \eta(t)y_i^j(t), (j = 2, 4)$. Here, $0 < \eta < 1$ is a learning rate. Two look-up tables for

$$\kappa_\alpha = \frac{\Gamma(\frac{5}{\alpha})\Gamma(\frac{1}{\alpha})}{\Gamma^2(\frac{3}{\alpha})} - 3 \text{ and } \kappa_\beta = \frac{3\Gamma(\frac{\beta-4}{2})\Gamma(\frac{\beta}{2})}{\Gamma^2(\frac{\beta-2}{2})} - 3 \qquad (18)$$

are established in advance, and α or β is sought from the table according to the value of $\hat{\kappa}_i$. Next, the scaling constant is calculated as

$$\lambda_\alpha = \left[\frac{\Gamma(\frac{3}{\alpha})}{m_2\Gamma(\frac{1}{\alpha})}\right]^{\frac{1}{2}} \text{ or } \lambda_\beta = \left[\frac{\beta\Gamma(\frac{\beta-2}{2})}{2m_2\Gamma(\frac{\beta}{2})}\right]^{\frac{1}{2}} \qquad (19)$$

using the estimate $\hat{m}_2$ and α or β. Finally, the nonlinear function is calculated by using Eq. (16) or (17) and $\mathbf{W}$ is updated by using Eq. (14). In this study, $\eta = 0.01$ is used to separate the overlapping sub- and super-Gaussian components.

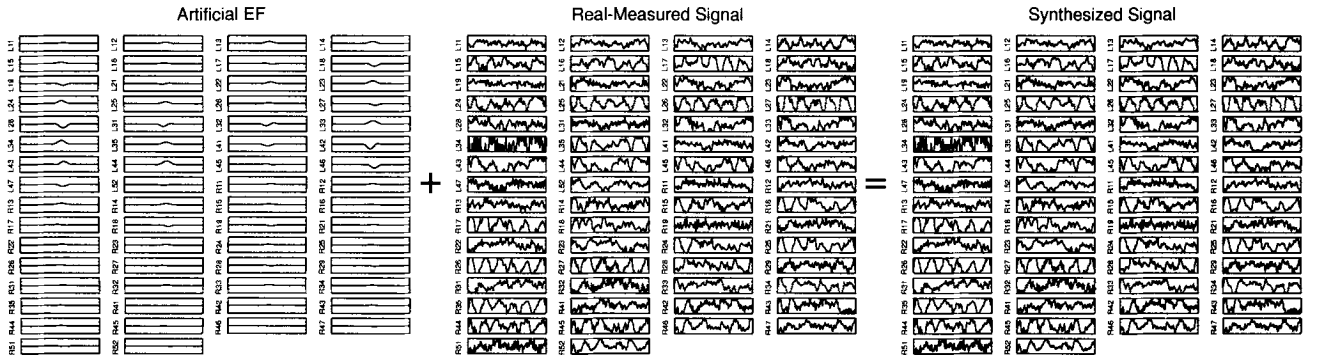


Fig. 1 Example of data synthesis at 1st trial: artificial EF signals (left), real measured 1st trial MEG signals (middle), and synthesized signals (right): The horizontal axis expresses time from 0 to 0.5 sec. and the vertical axis expresses amplitude from -0.4 to 0.4 pT.

4.4 Projection of the decomposed components

The robust pre-whitening and ICA techniques serve to filter the raw data, decreasing the power of the sensor noises and decomposing the sources. The estimated behavior of the individual sources can be represented as Eq. (9). To better visualize the information, we projected the decomposed components into the sensor space.

The virtual sensor signals coming from multiple components are obtained as

$$\widehat{\mathbf{x}}(t) = \widehat{\mathbf{A}}\mathbf{W}^{-1}\mathbf{y}(t) \tag{20}$$

To determine the information of the k-th individual components, we forced every element to be zero except the k-th ($k = 1, \cdots, \widehat{n}$) of $\mathbf{y}(t)$ in Eq. (20). The virtual sensor signals coming from the k-th individual components are obtained as

$$\widehat{\mathbf{x}}_k(t) = \widehat{\mathbf{A}}\mathbf{W}^{-1}[0 \cdots \mathbf{y}_k(t) \cdots 0]^T \tag{21}$$

The relationship between the virtual sensor signals from the multiple and the k-th individual components $\widehat{\mathbf{x}}_k$ is

$$\widehat{\mathbf{x}}(t) = \sum_{k=1}^{\hat{n}} \widehat{\mathbf{x}}_k(t) \tag{22}$$

Note that the sensor noises and some source noises have been reduced in the estimated observation $\widehat{\mathbf{x}}_k(t)$.

5. Simulation Method

5.1 MEG data set

In this subsection, our synthesized MEG data set is described, which is similar to AEF. This data is synthesized from an artificial signal and a real measured MEG signal (see Fig. 1), and is used to accurately evaluate the performance of source decomposition. The main advantage of our data set is that the dipole locations and the dynamics of evoked responses are known in advance, which facilitates the evaluation of the decomposed components.

When analyzing the synthesized data, the data must be realistic or adequate for real brain source separation. Therefore, not an artificial noise signal but a real measured brain signal is added together to the artificial evoked signal. This data set is based on a neuroscience perspective and is synthesized similarly to AEF. This means that the power and range of the evoked field response are based on experimental data.

The artificial evoked field signals are the virtual sensor signals, under the assumption that the evoked field source is evoked at the source of the magnetic field. The evoked signal is artificially evoked from 0.2 sec. to 0.3 sec. with a peak at 0.25 sec. (see Fig. 1). The location [x, y, z] (mm), direction vector [azimuth: az, declination: dec] (deg.), and dipole moment Q (nAm) are set as [x, y, z] = [10.0, 10.0, 60.0] mm, [az, dec] = [50.0, 103.0] deg. and $Q = 40$ nAm, respectively, where a head model presupposes a sphere with a radius of 75 mm.

The real measured signals are recorded by using an Omega-64 (CTF Systems Inc., Canada). During detection of the real measured MEG signal, a male subject is instructed to keep his eyes closed in a relaxed condition. We consider that there is almost no eye blink components, eye movement components, or muscle artifacts in this data set. The sensor arrays consist of 64 channels and the sampling rate is 250 Hz with a duration of 40 sec. The observed data is segmented into 80 trials, so the duration of each trial is 0.5 sec. and each trial has 125 samples.

The source signals in this data set include the evoked field response as well as the electrical power interference and the spontaneous α-wave component involved in the real measured MEG data.

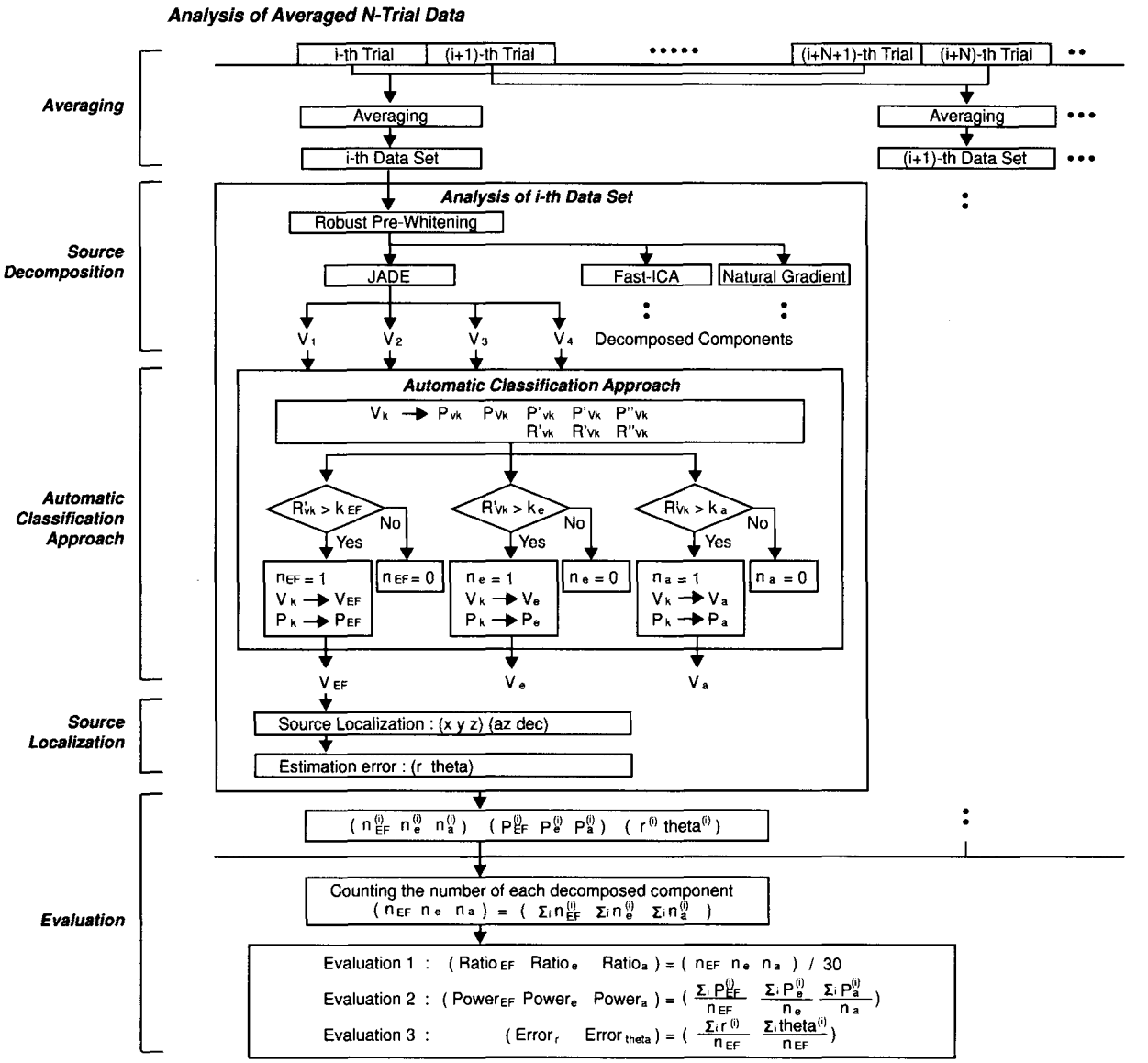


Fig. 2 Procedure for averaging, source decomposition, automatic classification, source localization, and evaluation

5.2 Power of the decomposed components

It is well known that the amplitudes of the signals decomposed by ICA are not preserved. In order to measure the power of the decomposed signals, the procedure of data projection as described in Section **4.4** is applied . The virtual sensor signal coming from the k-th ($k = 1, \cdots, \hat{n}$) individual component is represented as Eq. (21). In this work, the average of the virtual sensor signal from the k-th individual component is defined as:

$$\mathbf{v}_k(t) = \frac{1}{M}\sum_{i=1}^{M} \hat{\mathbf{x}}_{k,i}(t), \quad k = 1, \cdots, \hat{n} \qquad (23)$$

where t represents sample number, $t = 0$ (0 sec.), $\cdots$, 125 (0.5 sec.), $M = 64$ denotes the number of sensors and $\hat{\mathbf{x}}_{k,i}$ denotes the virtual sensor signal at channel i ($i = 1, \cdots, 64$) coming from the k-th decomposed components. Therefore, $\mathbf{v}_k$ represents the averaged virtual sensor signal, and its amplitude is not ambiguous.

Using the above results, the power of the k-th output $\mathbf{v}_k$ is defined as

$$P_{\mathbf{v}k} = \sum_{t=0}^{N} \mathbf{v}_k(t)\mathbf{v}_k^T(t) \qquad (24)$$

where $N = 125$ denotes the number of data samples. Using $P_{\mathbf{v}k}$ as the power of the k-th decomposed components, the power of individual components can be compared.

5.3 Simulation and evaluation methods

In this subsection, the simulation and evaluation methods are explained. To investigate the relationship between the performance of source decomposition and the number of averages, 1) the ratio of source decomposition, 2) the power of decomposed components, and 3) the accuracy of source estimation are demonstrated.

The procedure of analysis of the averaged N-trial data is shown in Fig. 2. In this study, we take averages across different numbers of trials from 1 to 50 ($N = 1, \cdots, 50$ in Fig. 2). In each averaging process, 30 different data sets are used as the input data ($i = 1, \cdots, 30$ in Fig. 2).

5.3.1 Automatic classification approach

In this study, the decomposed evoked field signal, the electrical power interference and the spontaneous α-wave are investigated. Here, the procedure of automatic classification of the decomposed components is described.

First, the power of each decomposed component in the time domain from 0 sec. ($t = 0$) to 0.5 sec. ($t = 125$) is calculated using Eq. (24). To define the criterion for classifying the decomposed components into the evoked field signal, the power of each decomposed component in the duration from 0.2 sec. ($t = 50$) to 0.3 sec. ($t = 75$) is calculated as

$$P'_{\mathbf{v}k} = \sum_{t=50}^{75} \mathbf{v}_k(t)\mathbf{v}_k^T(t) \tag{25}$$

Using the above results, the ratio

$$R'_{\mathbf{v}k} = \frac{P'_{\mathbf{v}k}}{P_{\mathbf{v}k}} \tag{26}$$

is defined. When $R'_{\mathbf{v}k} \geq k_{EF}$ (positive constant), the decomposed component $\mathbf{v}_k$ is the evoked field signal, since the evoked field signal is artificially evoked from 0.2 to 0.3 sec.

Similarly, to define the criterion for classifying the decomposed components into the electrical power interference and spontaneous α-wave, the power of each decomposed component in the frequency domain from 0 Hz ($f = 0$) to 124 Hz ($f = 62$) is calculated as

$$P_{\mathbf{V}k} = \sum_{f=0}^{62} \mathbf{V}_k(f)\mathbf{V}_k^T(f) \tag{27}$$

where $\mathbf{V}_k(f)$ denotes the power spectrum of $\mathbf{v}_k(t)$ and f denotes the sample number, $f = 0$ (0 Hz), $\cdots$ 62 (124 Hz). Next, that from 48 Hz ($f = 24$) to 52 Hz ($f = 26$) and that from 8 Hz ($f = 4$) to 12 Hz ($f = 6$) are calculated as

$$P'_{\mathbf{V}k} = \sum_{f=24}^{26} \mathbf{V}_k(f)\mathbf{V}_k^T(f) \tag{28}$$

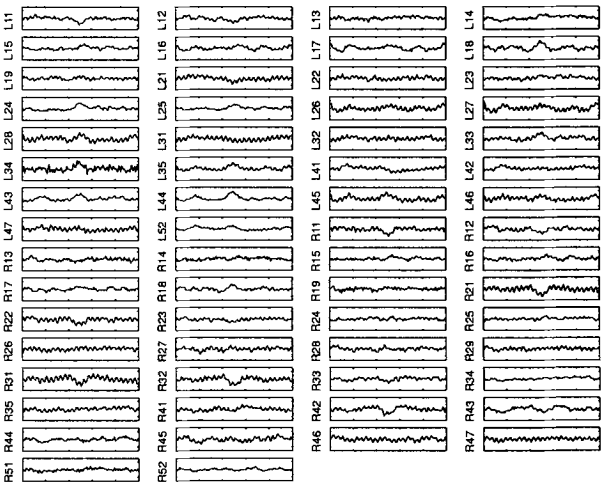


Fig. 3 Averaged waveforms of the 1st to 30th trial data: horizontal axis and vertical axis express time from 0 to 0.5 sec. and amplitude from -0.3 to 0.3 pT, respectively.

$$P''_{\mathbf{V}k} = \sum_{f=4}^{6} \mathbf{V}_k(f)\mathbf{V}_k^T(f) \tag{29}$$

respectively, and then the ratios

$$R'_{\mathbf{V}k} = \frac{P'_{\mathbf{V}k}}{P_{\mathbf{V}k}}, \qquad R''_{\mathbf{V}k} = \frac{P''_{\mathbf{V}k}}{P_{\mathbf{V}k}} \tag{30}$$

are defined. When $R'_{\mathbf{V}k} \geq k_e$ (positive constant) or $R''_{\mathbf{V}k} \geq k_\alpha$ (positive constant), the decomposed component $\mathbf{v}_k$ is the electrical power interference or α-wave component, respectively.

Based on earlier experiments, for this experiment, these parameters are set as $k_{EF} = 0.6$, $k_e = 0.4$ and $k_\alpha = 0.35$, respectively. The parameters k are determined by the following conditions. a) Only one component (evoked field signal, electrical interference, α-wave, or other components) can be decided from one decomposed component ($\mathbf{v}_k$). b) In case more than two decomposed components are decided as the same kind of component, these component are regarded as the same component. c) The number of definite components from all decomposed components will become the maximum.

5.3.2 Evaluation methods

In order to study the ratio of source decomposition, we count the number of decomposed evoked field signals (n_{EF}), electrical power interferences (n_e) and spontaneous α-waves (n_α) in 30 simulations ($i = 1, \cdots, 30$), as shown in Fig. 2, and then calculate their ratios:

$$Ratio_{EF} = \frac{n_{EF}}{30}, \; Ratio_e = \frac{n_e}{30} \text{ and } Ratio_\alpha = \frac{n_\alpha}{30} \tag{31}$$

Table 1 Estimated map and estimation error for averaged 30-trial data with application of the JADE algorithm

	dipole location (mm)				direction vector (deg.)		
	x	y	z	r	az	dec	Θ
true evoked field	10.0	10.0	60.0	–	50.0	103.0	–
averaged observation	−0.1	0.0	69.8	17.3	47.7	89.9	13.3
result of pre-whitening and JADE	1.0	8.4	70.3	13.6	46.4	95.5	8.3

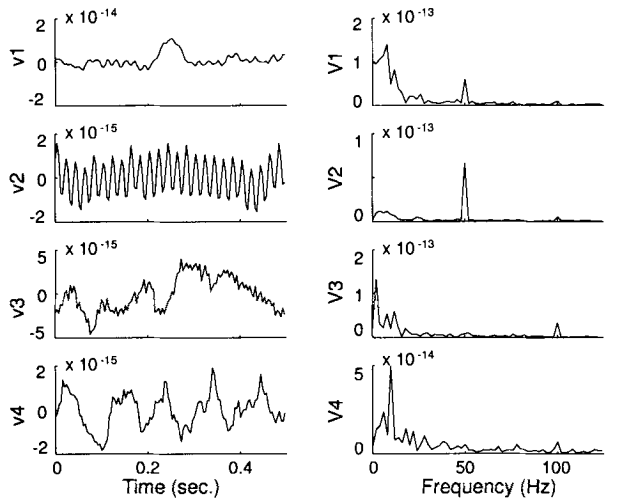


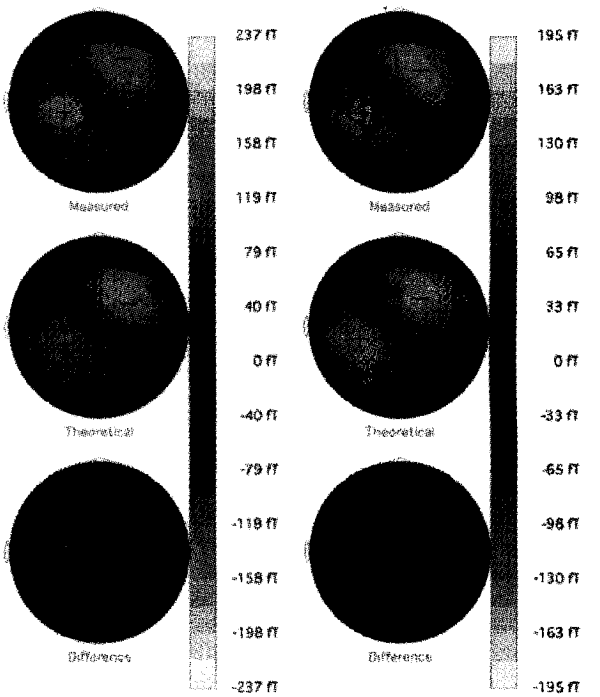
Fig. 4 Result of source decomposition and its frequency contents for averaged 30-trial data with application of the JADE algorithm

In order to investigate the relationship between the power of the decomposed components and the number of averages, we take the averages of the power of decomposed evoked field, electrical power interference, and spontaneous α-wave as

$$\begin{aligned} Power_{EF} &= \frac{1}{n_{EF}} \sum_i P_{EF}^{(i)} \\ Power_e &= \frac{1}{n_e} \sum_i P_e^{(i)} \\ Power_\alpha &= \frac{1}{n_\alpha} \sum_i P_\alpha^{(i)} \end{aligned} \tag{32}$$

respectively (See Fig. 2). Here, $P_{EF}^{(i)}$ denotes the power of the evoked field signal decomposed from the i-th data set, and $P_e^{(i)}$ and $P_\alpha^{(i)}$ denote that of the electrical power interference and the α-wave, respectively.

To investigate the accuracy of source estimation, we apply dipole estimation, focusing on the evoked response, to averaged observation and the decomposed evoked signal. Since the dipole locations and the direction vector of the evoked field are known in advance, the estimation error of the evoked field can be obtained, by comparing the true and estimated ones.



(a) (b)

Fig. 5 Estimated map focusing on evoked field: (a) averaged data, (b) decomposed evoked field

The standard spatio-temporal dipole fitting routine, MEG v3.3a (CTF System Inc., Canada), is used to locate the dipole.

The distance between the true dipole location ($[x,\ y,\ z]$ mm) and the estimated dipole location ($[\widehat{x},\ \widehat{y},\ \widehat{z}]$ mm) is obtained as

$$r = \sqrt{(x-\widehat{x})^2 + (y-\widehat{y})^2 + (z-\widehat{z})^2} \tag{33}$$

When combining the starting points of the two vectors, the angle Θ between the direction vectors of the true and estimated evoked fields is obtained as

$$\Theta = \cos^{-1}[\sin(d)\sin(\widehat{d})\cos(a-\widehat{a}) + \cos(d)\cos(\widehat{d})] \tag{34}$$

where a and d denote the azimuth and declination (deg.), respectively. To compare these errors, the averages are calculated as

$$Error_r = \frac{1}{n_{EF}} \sum_i r^{(i)}$$

$$Error_{\Theta} = \frac{1}{n_{EF}} \sum_{i} \Theta^{(i)} \quad (35)$$

Here, $r^{(i)}$ denotes the error r derived from the i-th data set, and $\Theta^{(i)}$ denotes that of Θ.

6. Simulation Results

6.1 Results of source decomposition using the JADE algorithm

In this subsection, the results of the source decomposition and localization for the averaged 30-trial data (Fig. 3), with application of the JADE algorithm, are demonstrated. In Fig. 3, the horizontal axis and vertical axis express time from 0 to 0.5 sec. and amplitude from -0.3 to 0.3 pT, respectively.

First, the number of sources is estimated under the criterion of choosing the number of sources as described in Section **3.2**. In this simulation, it is decided when the cumulative percentage variance contribution as shown in Eq. (8) becomes larger than 75 %. In applying this method, the cumulative percentage variance contributions for a given value of $\hat{n}$ become $c_k = 34.03,\ 59.81,\ 69.53,\ 76.22\ (k = 1, \cdots, 4)$, and the number of sources is assumed to be $\hat{n} = 4$.

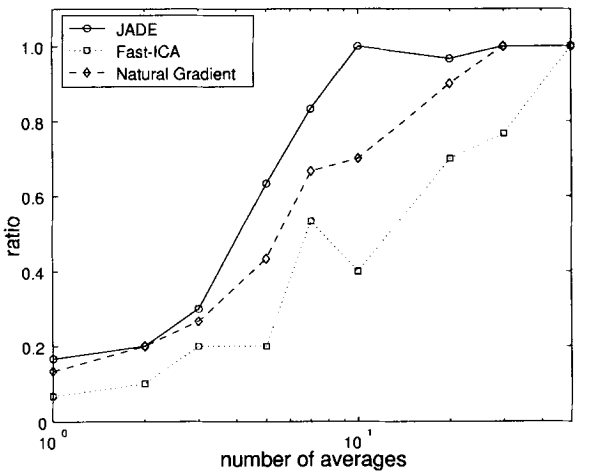
Next, the robust pre-whitening technique and JADE are applied. The results $\mathbf{v}$ and the power spectra $\mathbf{V}$ are shown in Fig. 4. After applying the automatic classifying approach to these decomposed components, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_4$ in Fig. 4 are regarded as evoked field response, electrical power interference, and spontaneous α-wave, respectively.

The estimated maps derived from the averaged observation and decomposed evoked response are shown in Fig. 5. In this figure, the top 'Measured' map is derived from a decomposed component. The middle 'Theoretical' map is computed from the moving dipoles. The bottom 'Difference' map represents the difference between the 'Measured' and 'Theoretical' maps. As for the result of using the evoked response, note that the evoked response appears on the left area of the brain. The estimated evoked field is shown in Table 1. Comparing the results of JADE and taking averages, both the source location and direction vector are rendered more accurate by applying the JADE algorithm.

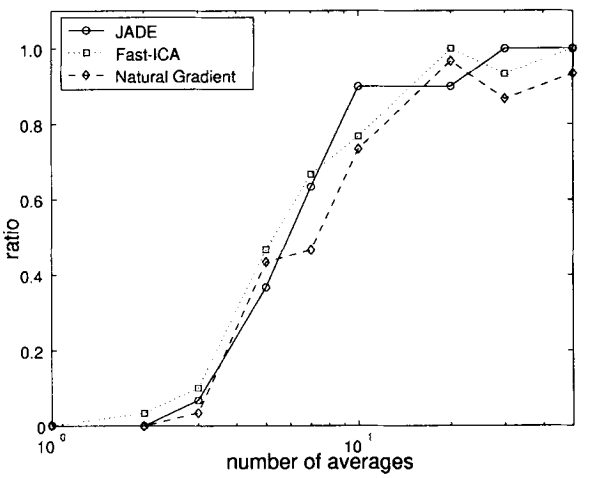
6.2 Relationship between the performance of decomposition and the number of averages

In this subsection, the relationships between the performance of decomposition and the number of averages using JADE, Fast-ICA and the natural gradient-based algorithm are demonstrated.

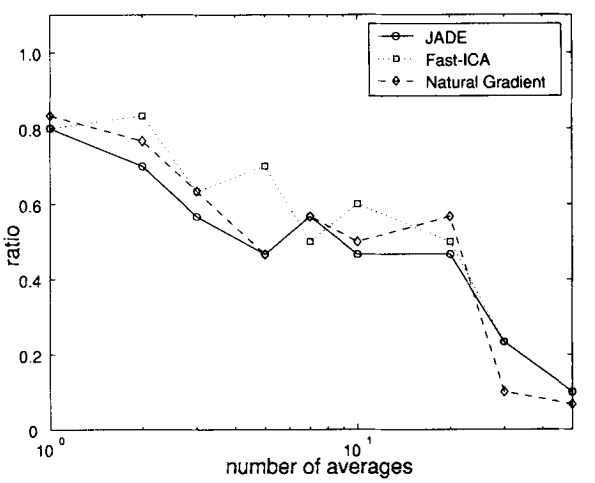
The relationship between the ratio of source decomposition, calculated by Eq. (31), and the number



(a)



(b)



(c)

Fig. 6 Decomposition ratios and the number of averages across data trials focusing on: (a) evoked fields response, (b) electrical power interference, and (c) spontaneous α-wave component

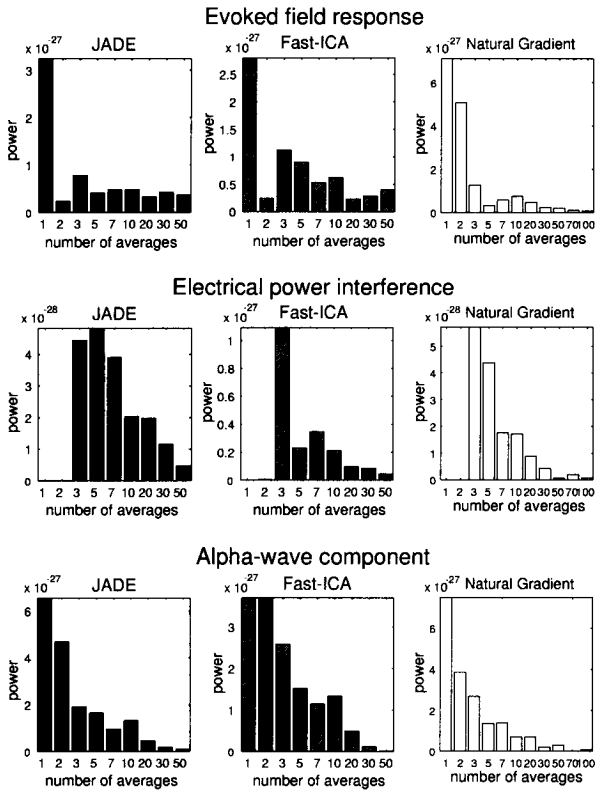


Fig. 7 Power of decomposed components with application of JADE, Fast-ICA and the natural gradient-based algorithm, focusing on: decomposed evoked response, electrical interference and α-wave component

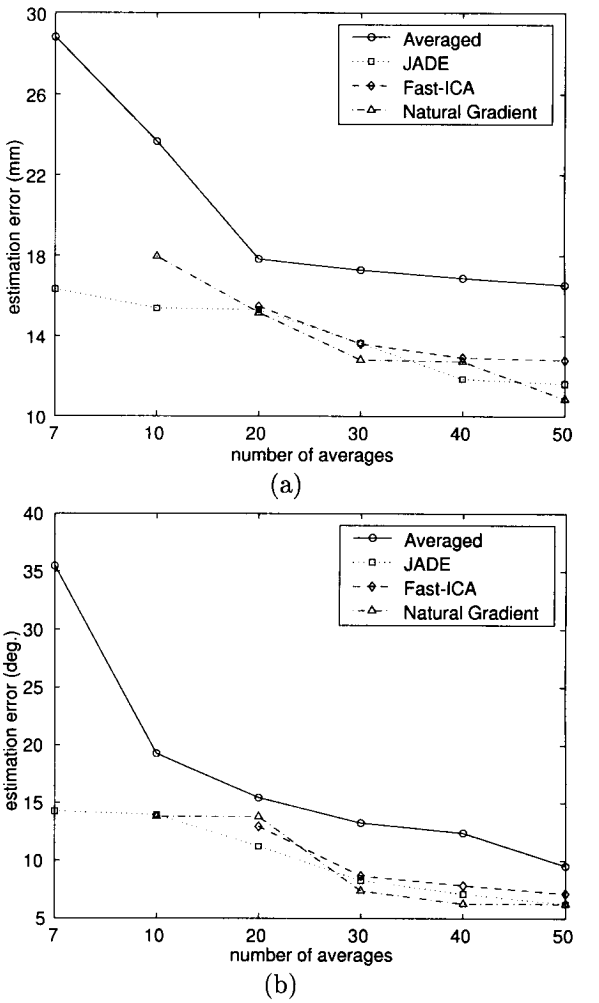


Fig. 8 Estimation error with application of JADE, Fast-ICA, and the natural gradient-based algorithm : (a) dipole location, and (b) direction vector

of averages is shown in Fig. 6. In these figures, the horizontal axis expresses the number of averages from 1 to 50 trials and the vertical axis expresses these ratios. There are several points to note in these figures. The ratios of the evoked field signal and electrical power interference are very low when the number of averages is small, but when the number of averages is large, the ratios are high. Note that, in this simulation, the electrical power interference is synchronous with the cycle of trials. In contrast, the ratio of the spontaneous α-wave component is high when the number of averages is small, but the larger the number of averages, the lower the ratio. Considering the characteristics of taking averages, these results appear to be valid. Comparing the results of applying JADE, Fast-ICA and the natural gradient-based algorithm, the highest ratio of the source decomposition of an evoked field signal can be obtained when applying the JADE algorithm. Regarding the ratios of the electrical power interference and α-wave component, there is not a significant difference between these results.

The relationship between the power of each decomposed component and the number of averages is shown in Fig. 7. Regarding the power of the evoked response, it is not influenced by taking an average. The results of applying the JADE algorithm appear to be constant against the number of averages. This means that the JADE algorithm is efficient for source decomposition and estimation. But with respect to the single-trial data analysis, the power of the evoked field becomes very large, because of the influence of additive noise. In contrast, regarding the power of electrical power interference and α-wave, the larger the number of averages, the lower these powers become. This means that the power of decomposed electrical power interference is influenced by an α-wave or other high power noise components. Here, in case the number of averages is one or two, the power of the electrical power interference becomes zero, because this signal could not be decomposed in such cases.

The accuracy of source estimation, the relationship between the estimation error of evoked field and

the number of averages, is shown in Fig. 8. In this experiment, we apply the dipole estimation of the evoked response, in case the ratio of source decomposition of the evoked field signal ($Ratio_{EF}$) becomes $Ratio_{EF} > 0.7$. Given these results, regardless of the analysis, the larger the number of averages, the lower the estimation error becomes. Comparing the results of averaged data with the decomposed evoked field, regardless of the number of averages, the evoked field signal is accurately decomposed by applying any ICA algorithm. This means that the number of averages can be reduced by applying the ICA approach.

Given the results of the three kinds of experiments conducted for the evaluation of source decomposition, the JADE algorithm is slightly effective for our MEG data analysis.

7. Conclusions

In this work, the relationship between the performance of source decomposition and the number of averages across data trials was investigated. Moreover, a number of existing ICA algorithms such as JADE, Fast-ICA, and the natural gradient-based algorithm with a robust pre-whitening technique were used to decompose the MEG data. Our results show the relationship between the accuracy of source decomposition and the number of averages, and by applying the ICA approach, the number of averages can be reduced. These results confirm the effectiveness of the developed methods of data analysis.

Acknowledgments

This research is partially supported by the Japan Society for the Promotion Science (JSPS) in the Japan-China Research Cooperative Program.

References

[1] L. Molgedey and H. G. Schuster: Separation of a mixtures of independent signals using time delayed correlations, Phys. Rev. Lett., Vol. 72, No. 23, pp. 3634-3637, 1994.

[2] A.T. Bell and T.J. Sejnowski: An information maximization approach to blind separation and blind deconvolution, Neural Computation, Vol. 7, No. 6, pp. 1004-1034, 1995. Matlab code in WWW : http://www.cnl.salk.edu/ tony/ica.html

[3] S. Amari, A. Cichocki and H. H. Yang: A new learning algorithm for blind signal separation, Advances in Neural Information Processing System 8, MIT Press, pp. 757-763, 1996.

[4] J. F. Cardoso and A. Souloumiac: Jacobi angles for simultaneous diagonalization, SIAM J. Mat. Anal. Appl., Vol. 17, No. 1, pp. 145-151, 1996. Matlab code in WWW : http://sig.enst.fr/ cardoso/jointdiag.html

[5] B. A. Pearlmutter and L. C. Parra: A context-sensitive generalization of ICA, International Conference on Neural Information Processing, pp. 151-157, September 1996.

[6] A. Hyvaerinen and E. Oja: A fast fixed-point algorithm for independent component analysis, Neural Computation, Vol. 9, No. 7, pp. 1483-1492, 1997.

[7] T-W. Lee, M. Girolami and T. Sejnowski: Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources, Neural Computation, Vol. 11, No. 2, pp. 417-441, 1998.

[8] S. Makeig, A. J. Bell, T. -P. Jung and T. J. Sejnowski: Independent component analysis of electroencephalographic data, Advances in Neural Information Processing System 8, MIT press, pp.145-151, 1996.

[9] R. Vigário, A. Hyvärinen and E. Oja: ICA fixed-point algorithm in extraction of artifacts from EEG, NORSIG-96, pp. 383-386, 1996.

[10] J. Cao and A. Cichocki: Blind source separation algorithm based on 4th-order self- and cross-cumulant criterion applied for EEG data analysis, Proc. of 1997 International Symposium on Nonlinear Theory and its Applications, Vol. 2, pp. 1005-1008, 1997.

[11] J. Cao, N. Murata, S. Amari, A. Cichocki and T. Takeda: MEG data analysis based on ICA approach with pre- & post-processing techniques, Proc. of 1998 International Symposium on Nonlinear Theory and its Applications, Vol. 1, pp. 287-290, 1998.

[12] T-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne and T.J. Sejnowski: Independent component analysis of single-trial event related potentials, Proc. ICA'99, pp. 173-179, 1999.

[13] S. Ikeda and K. Toyama: Independent component analysis for noisy data - MEG data analysis, Neural Networks 13, pp. 1063-1074, 2000.

[14] S. Ikeda: ICA on noisy data: A factor analysis approach, in Advances in Independent Component Analysis, Edt. M. Girolami, Springer, 2000.

[15] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen and E. Oja: Independent component approach to the analysis of EEG and MEG recordings, IEEE Trans. Biomed. Eng., Vol. 47, No. 5, pp. 589-593, 2000.

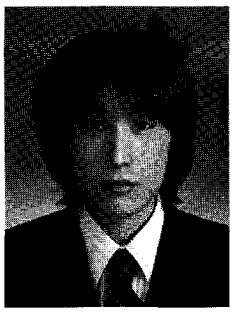
[16] A. Ziehe, K. -R. Müller, G. Nolte, B. -M. Mackert and G. Curio: Artifact reduction in magnetoneurography based on time-delayed second-order correlations, IEEE Trans. Biomed. Eng., Vol. 47, No. 1, pp. 75-87, 2000.

[17] J. Cao, N. Murata, S. Amari, A. Cichocki and T. Takeda: Independent component analysis for single-trial MEG data decomposition and single-dipole source localization, Neurocomputing, Vol. 49, pp. 255-277, 2002.

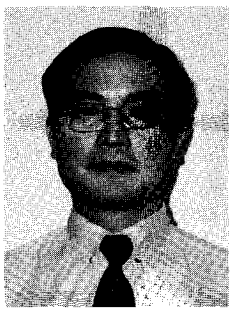
[18] Y. Konno, J. Cao and T. Takeda: Decomposition and localization of MEG brain sources, Journal of Signal Processing, Vol. 6, No, 6, pp. 391-400, Nov. 2002.

[19] J. Cao, N. Murata, S. Amari, A. Cichocki and T. Takeda: A robust approach to independent component analysis of signals with high-level noise measurements, IEEE Trans. on Neural Networks, Vol. 14, No. 3, pp. 631-645, June 2003.

[20] Y. Konno, J. Cao, T. Arai and T. Takeda: Visualization of brain activities of single-trial and averaged multiple-trials MEG data, IEICE Trans. on Fundamentals, Vol. E86-A, No. 9, pp. 2294-2302, Sep. 2003.



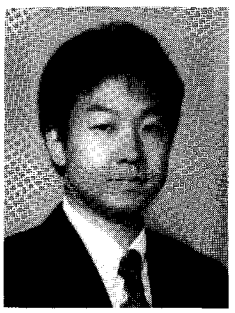
Yoshio Konno received the B.Eng. and M.Eng. degrees from Sophia University, Tokyo, Japan, in 2002 and 2004, respectively. He is currently a Ph.D. course student in the Department of Electrical and Electronics Engineering, Sophia University. He received the Student Paper Award from the Research Institute of Signal Processing Japan (RISP) in 2004. His research interests include signal processing and blind signal processing. He is a student member of IEICE, Japan Society for Medical and Biological Engineering and Japan Biomagnetism and Bioelectromagnetics Society.



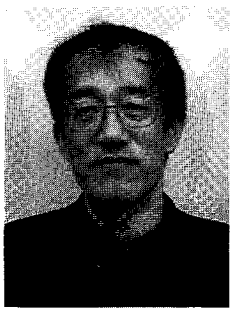
Jianting Cao received the M. Eng. and Ph.D. degrees from the Graduate School of Science and Technology, Chiba University, Japan, in 1993 and 1996, respectively. From 1983 to 1988, he worked as a Researcher at the Institute of Technology and Equipment in the Ministry of Geological and Mineral in China. From 1996 to 1998, he worked as a Researcher at the Brain Science Institute, RIKEN (The Institute of Physical and Chemical Research) in Japan. From 1998 to 2002, he worked as an Associate, and an Assistant Professor at the Sophia University in Japan. He is currently working as an Associate Professor at the Department of Electronic Engineering, Saitama Institute of Technology, and a Visiting Research Scientist at the Brain Science Institute, RIKEN in Japan. He received the Best Paper Award from the Telecommunications Advancement Foundation (Japan) in 1996. His research interests include blind signal processing, biomedical signal processing, neural networks and learning algorithms. Dr. Cao is a member of IEEE and IEICE (Japan).

Tsunehiro Takeda was born in 1948. He graduated from Department of Mathematical Engineering and Information Physics, Tokyo University in 1972. He received Master and Doctor degrees in 1974 and 1977 from the same University. He served Industrial Products Research Institute, MITI from 1977 to 1997. From 1997, he is a Professor of Tokyo University. At present, He is a professor of Department of Complexity of Science and Engineering, Graduate School of Frontier Sciences, Tokyo University. His main current research is Elucidation of Human Higher Brain Functions using MEG (Magnetoencephalography). He received Awards from Agency of Science and Technology (1991), Agency of Industrial Science and technology (1992), Tsukuba City (1993) for outstanding researches and Agency of Science and Technology for prominent patents (1988, 1991, 1994).

Hiroshi Endo was born in Japan in 1963. He received the B. Eng. and M. Eng. in Mechanical Engineering from the Keio University in 1985 and 1987, respectively. From 1987 to 1993, he was a researcher at the Industrial Products Research Institute, and he is currently a researcher at the National Institute of Advanced Industrial Science and Technology (AIST). His research interests include the human motion control system and functional imaging with MEG.



Takayuki Arai received the B.E., M.E. and Ph.D. degrees in electrical engineering from Sophia Univ., Tokyo, Japan, in 1989, 1991 and 1994, respectively. In 1992-1993 and 1995-1996, he was with Oregon Graduate Institute of Science and Technology (Portland, OR, USA). In 1997-1998, he was with International Computer Science Institute (Berkeley, CA, USA). He is currently Associate Professor of the Department of Electrical and Electronics Engineering, Sophia Univ. In 2003-2004, he was a visiting scientist at Massachusetts Institute of Technology. His research interests include signal processing, acoustics, speech and hearing sciences, and spoken language processing.



Mamoru Tanaka was born on September 20, 1948, in Kanagawa, Japan. He received the B.E., M.E. and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1972, 1974 and 1981, respectively. In 1974, he joined the Circuit Development Department, Computer Engineering Division, Nippon Electric Company (NEC) where he was engaged in design and development of NEC's LSI computers. He resigned from the company and entered in the graduate school of Keio University in 1978. In April 1981, he became an Associate Professor with the Department of Electrical and Electronics Engineering of Sophia University, Tokyo, Japan. He is now a Professor of Sophia University. He has done research in analysis of a large scale of networks and architectures of new LSI computers. He is now interested in analysis and synthesis of data mining by Cellular Neural Networks. He was an Associate Editor of the IEEE Transactions on Circuits and Systems.

(Received May 13, 2004: revised August 5, 2004)