IEICE TRANS. FUNDAMENTALS, VOL.E86–A, NO.9 SEPTEMBER 2003

PAPER *Special Section on Nonlinear Theory and its Applications*

Visualization of Brain Activities of Single-Trial and Averaged Multiple-Trials MEG Data

Yoshio KONNO[†a)], ***Student Member***, **Jianting CAO**[††], **Takayuki ARAI**[†], ***and*** **Tsunehiro TAKEDA**[†††], ***Regular Members***

SUMMARY Treating an averaged multiple-trials data or non-averaged single-trial data is a main approach in recent topics on applying independent component analysis (ICA) to neurobiological signal processing. By taking an average, the signal-to-noise ratio (SNR) is increased but some important information such as the strength of an evoked response and its dynamics will be lost. The single-trial data analysis, on the other hand, can avoid this problem but the SNR is very poor. In this study, we apply ICA to both non-averaged single-trial data and averaged multiple-trials data to determine the properties and advantages of both. Our results show that the analysis of averaged data is effective for seeking the response and dipole location of evoked fields. The non-averaged single-trial data analysis efficiently identifies the strength and dynamic component such as α-wave. For determining both the range of evoked strength and dipole location, an analysis of averaged limited-trials data is better option.

key words: *magnetoencephalography (MEG), robust prewhitening technique, independent component analysis (ICA), source localization*

1. Introduction

Many researchers have applied ICA to electroencephalographic (EEG) or magnetoencephalographic (MEG) data to determine the behavior and localization of brain sources [1]–[11]. When detecting a MEG signal, spontaneous and environmental noises may seriously effect recorded data because the magnetic field of brain signals is weak, particularly in the case of non-averaged single-trial data.

Taking an average across many stimulation trials is the most widely used technique for reducing instrumental and environmental noises and for identifying the behavior and location of activities of interest such as evoked field responses. In fact, when applying ICA to MEG data, most researchers have treated averaged data [1], [5], [6], [8]–[11]. However, by taking an average, many pieces of important information such as the trial-by-trial variation of the signals' amplitude or its dynamics will be lost.

Recently, some researchers have begun to study non-averaged, single-trial data to retain the information lost by averaging. Several ICA algorithms have been investigated for the single-trial MEG data [2], [3], [7]. The disadvantage here is that because the SNR of single-trial data is very poor, the decomposition of a low-power source signal from recorded data is still influenced by noise.

In this paper, we deal with both non-averaged single-trial data and averaged multiple-trials data and we demonstrate several properties for analyzing the data sets. We take note of several advantages with regards to amplitude, dipole location and dynamics particular to each of the data sets.

When applying ICA to physiological data, most researchers have used real, measured physiological data, with some individual responses evoked by stimuli, and their decomposed components are evaluated from a neuroscience perspective. In this study, we use a synthesized MEG data set, which includes an artificial evoked field and real, measured brain data. The behavior of our data set is similar to auditory evoked fields (AEFs). The main advantage of our data set is that dipole location of evoked responses and its dynamics are known in advance, which facilitates the evaluation of the decomposed components.

To decompose the source signal, in the preprocessing of raw data, we applied an additive noise reduction technique [2], [3]. At this stage, the non-averaged or averaged MEG data are first decomposed into de-correlated signals with the reduction of additive noise. In the stage of ICA, the de-correlated source signals are further decomposed into independent components by applying the joint approximate diagonalization of eigenmatrices (JADE) algorithm [12]. In the post-processing stage, we project these decomposed components onto the sensor space. At this stage, we propose a new criterion to measure the power of an individual component.

Our paper is organized as follows. A model for applying the ICA approach to MEG data analysis is described in Sect. 2. A description of the method for

Manuscript received December 31, 2002.
Manuscript revised March 30, 2003.
Final manuscript received May 13, 2003.

[†]The authors are with the Department of Electrical and Electronics Engineering, Sophia University, Tokyo, 102-8554 Japan.
[††]The author is with the Department of Electronics Engineering, Saitama Institute of Technology, Saitama-ken, 369-0293 Japan, also with the Lab for Advanced Brain Signal Processing, BSI, RIKEN, Wako-shi, 351-0198 Japan.
[†††]The author is with the Department of Complexity of Science and Engineering, Graduate School of Frontier Sciences, Tokyo University, Tokyo, 113-8656 Japan.
a) E-mail: yo-konno@sophia.ac.jp

source decomposition is presented in Sect. 3. The experimental results are presented in Sect. 4. The conclusions are drawn in Sect. 5.

2. Data Analysis Model

In this section, we describe the model for applying ICA to MEG data. Briefly, we assume there are two kinds of noises which must be reduced in the analysis: The first kind of noise is an additive or sensor noise, usually generated from the instruments (the individual sensors), which contaminates the observed signals. Their power will be reduced by using the robust pre-whitening technique in the pre-processing stage (c.f. Sect. 3.1). The second kind of noise is source noise. It will be discarded after the source decomposition by means of the ICA approach (c.f. Sect. 3.2). After removing these noises from the data, we will obtain the estimated components, such as the evoked responses.

Sensor and source noises come from the following process. We assume that 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.

Based on the principle of MEG measurement, this problem can be formulated as

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

where $\mathbf{x}(t) = [x_1(t), \cdots, x_m(t)]^T$ represents the transpose of m observations at time t. $\mathbf{s}(t) = [s_1(t), \cdots, s_n(t)]^T$ represents n unknown source components which contains 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 α-wave components. Environmental interference sources include 50 Hz electrical power interference and artifacts such as eye blinks. Since the environmental interferences contribute to at least two sensors, we regard them as common components and define them as "source noise." Another kind of component such as the additive noise represented by $\mathbf{e}(t) = [e_1(t), \cdots, e_m(t)]^T$ is called a unique component, and since it only contributes to one sensor we define it 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.

3. Data Analysis Method

3.1 Robust Pre-Whitening Technique

In this subsection, we describe our robust pre-whitening technique [2], [3]. This technique is very capable of reducing the power of sensor noises.

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 data samples. When the sample size N is sufficiently large, the covariance matrix of the observed data in the mixing model $\mathbf{\Sigma}$ can be written as

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T + \mathbf{\Psi}, \tag{3}$$

where $\mathbf{\Psi}$ is a diagonal matrix of an additive noise $\mathbf{E}$. Also the covariance matrix of the observed data recorded by sensors can be given by

$$\mathbf{C} = \mathbf{X}\mathbf{X}^T. \tag{4}$$

When the SNR is high, and the noise variance $\mathbf{\Psi}$ is small or close to zero, a cost function for fitting the model to the data can be employed to render $\mathbf{C} - \mathbf{A}\mathbf{A}^T$ as small as possible. Here, the mixing matrix $\mathbf{A}$ and its estimation $\widehat{\mathbf{A}}$ can be estimated as

$$\widehat{\mathbf{A}}\widehat{\mathbf{A}}^T = \mathbf{U}_{\hat{n}} \mathbf{\Lambda}_{\hat{n}} \mathbf{U}_{\hat{n}}^T, \tag{5}$$

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 $\widehat{n}$ is the estimated number of sources. In Eq. (5), let one possible solution for $\widehat{\mathbf{A}}$ is

$$\widehat{\mathbf{A}} = \mathbf{U}_{\hat{n}} \mathbf{\Lambda}_{\hat{n}}^{\frac{1}{2}}. \tag{6}$$

Note that $\widehat{\mathbf{A}}^T \widehat{\mathbf{A}} = \mathbf{\Lambda}_{\hat{n}}$, and the principal-component scores can be obtained from $\mathbf{x} = \widehat{\mathbf{A}}\mathbf{z}$, that is,

$$\mathbf{z}(t) = \mathbf{\Lambda}_{\hat{n}}^{-\frac{1}{2}} \mathbf{U}_{\hat{n}}^T \mathbf{x}(t). \tag{7}$$

Using this result, the covariance matrix is obtained as $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}_{\hat{n}}$, which means that $\mathbf{z}\mathbf{z}^T$ are identity and that the components are de-correlated in the new set of data.

This is the well-known, standard principal component analysis (PCA), but this technique does not take into account the diagonal elements of $\mathbf{\Psi}$. For a practical model such as ours, additive noises cannot be neglected. Therefore we needed to decompose the data

into de-correlated signals, reducing additive noise and optimizing dimensionality. In this study, we applied a robust pre-whitening technique to accomplish this task.

With the robust pre-whitening technique, instead of fitting $\mathbf{A}\mathbf{A}^T$ to $\mathbf{C}$ as is done for the standard PCA approach, we fit $\mathbf{A}\mathbf{A}^T$ to $\mathbf{C}-\mathbf{\Psi}$ using the eigenvalue decomposition method. In this case, the cost function is obtained as

$$L(\mathbf{A},\mathbf{\Psi}) = tr[\mathbf{A}\mathbf{A}^T - (\mathbf{C}-\mathbf{\Psi})]^2. \tag{8}$$

And we minimize it by $\frac{\partial L(\mathbf{A},\mathbf{\Psi})}{\partial \mathbf{\Psi}} = 0$, whereby the estimate noise variance $\mathbf{\Psi}$ is obtained as

$$\widehat{\mathbf{\Psi}} = diag(\mathbf{C} - \widehat{\mathbf{A}}\widehat{\mathbf{A}}^T), \tag{9}$$

where the estimate $\widehat{\mathbf{A}}$ is obtained in the same manner as shown in Eq. (6). Using these estimates $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{\Psi}}$, we can obtain the transform matrix for the robust pre-whitening technique as

$$\mathbf{Q} = [\widehat{\mathbf{A}}^T \widehat{\mathbf{\Psi}}^{-1} \widehat{\mathbf{A}}]^{-1} \widehat{\mathbf{A}}^T \widehat{\mathbf{\Psi}}^{-1}. \tag{10}$$

Using the above result, the new set of data transformed from the observations can be obtained by

$$\mathbf{z}(t) = \mathbf{Q}\mathbf{x}(t). \tag{11}$$

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

A similar noise reduction approach that applies factor analysis (FA) to decomposition of MEG data has been reported in [5]. 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 JADE Algorithm

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 still overlapped, and those will need to be discarded 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.

The JADE algorithm has two procedures termed orthogonalization in PCA and rotation. We did apply the rotation procedure in the JADE algorithm, described below, but instead of the orthogonalization in PCA, we applied the robust pre-whitening technique described in Sect. 3.1.

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

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

$$\mathbf{F}(M) = \sum_{k=1}^{K}\sum_{l=1}^{L} \mathrm{Cum}(z_i, z_j, z_k, z_l) m_{lk}, \tag{12}$$

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

After performing the robust pre-whitening technique and rotation in JADE, the de-mixture matrix can be obtained by $\mathbf{WQ}$. With these we can calculate the decomposed sources $\mathbf{y}\in\mathbf{R}^n$ as

$$\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t) = \mathbf{W}\mathbf{Q}\mathbf{x}(t). \tag{13}$$

3.3 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. (13). To better visualize the information, we projected the decomposed components onto the sensor space.

The virtual sensor signals coming from multiple components are obtained as

$$\widehat{\mathbf{x}}(t) = \widehat{\mathbf{A}}\mathbf{W}^{-1}\mathbf{y}(t). \tag{14}$$

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. (14). The virtual sensor signals coming from k-th individual components are obtained as

$$\widehat{\mathbf{x}}_k(t) = \widehat{\mathbf{A}}\mathbf{W}^{-1}[0 \cdots \mathbf{y}_k(t) \cdots 0]^T. \tag{15}$$

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

$$\widehat{\mathbf{x}}(t) = \sum_{k=1}^{\hat{n}} \widehat{\mathbf{x}}_k(t). \tag{16}$$

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

3.4 Power of the Decomposed Components

It is well known that the amplitudes of signals decomposed by ICA are irregular. In order to measure the power of decomposed signals, we apply the procedure of data projection as described in Sect. 3.3. The virtual sensor signals from the k-th individual components are obtained as Eq. (15). In this paper, we define the total sum of each virtual sensor signal from the k-th individual component as:

$$\mathbf{v}_k(t) = \frac{1}{M}\sum_{i=1}^{M}\widehat{\mathbf{x}}_{k,i}(t) \tag{17}$$

to compare the power of decomposed components. In Eq. (17), M denotes the number of sensors and $\widehat{\mathbf{x}}_{k,i}$ denotes the k-th decomposed components of $\mathbf{y}(t)$ into the i-th sensor. Here $\mathbf{v}_k$ represents the total observation signals derived by the k-th decomposed signal, so that its amplitude is not ambiguous.

Using the above results, we define the power of the k-th decomposed components $\mathbf{v}_k$ as

$$P_{\mathbf{v}k} = \sum_{t=1}^{N}\mathbf{v}_k(t)\mathbf{v}_k^T(t), \tag{18}$$

where, N denotes the number of data samples. Applying $P_{\mathbf{v}k}$ as the power of the k-th decomposed components, we can compare the power of individual components decomposed by ICA.

4. Experimental Results

4.1 Synthesized MEG Data

In this subsection, we describe the synthesized MEG data set, used for simulation, which is similar to the Auditory Evoked Fields (AEFs). We synthesized an artificial signal (Fig. 1(left)) and a real measured MEG signal (Fig. 1(middle)) which is recorded by using an Omega-64 (CTF Systems Inc., Canada). The sensor arrays consist of 64 channels and the sensor distribution is shown in Fig. 2. The sampling rate was 250 Hz with a duration of 50 sec. for 12500 samples. The observed data $\mathbf{X}_{(64\times12500)}$ was segmented into 100 trials, so the duration of each trial $\mathbf{X}_{i(64\times125)}(i = 1, \cdots, 100)$ is 0.5 sec. and each trial has 125 samples, where i denotes the trial number.

The source signals in this data set include two different evoked fields responses, EF1 and EF2, as shown in Fig. 1(left) and include the 50 Hz electrical power interference and the α-wave component involved in the real measured MEG data shown in Fig. 1(middle). The signal EF1 was artificially evoked from 0.2 sec. to 0.3 sec. with a peak at 0.25 sec. and its strength was 20 nAm (see Fig. 1(a)). The signal EF2 was artificially evoked from 0.22 sec. to 0.32 sec. with a peak at 0.27 sec. and its strength was 30 nAm (see Fig. 1(b)). The EF1 was located at $[x,\ y,\ z] = [10,\ 50,\ 50]$ mm and the EF2 was located at $[x,\ y,\ z] = [-40,\ 40,\ 40]$ mm (see Table 1), where a head model presupposes a sphere with a radius of 75 mm and x, y, z axis are set according to Fig. 3.

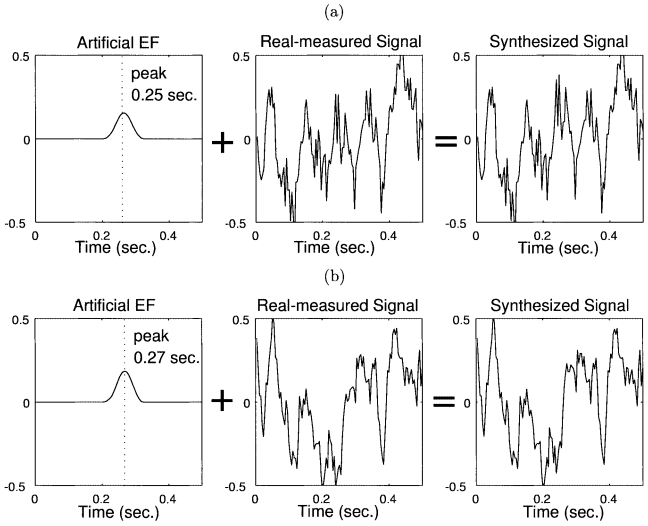

Fig. 1 An example for data synthesizing: (a) Data synthesizing at sensor-L24, which detects signal EF1. (b) Data synthesizing at sensor-L44, which detects signal EF2. In each example, artificial EF signals (left), real measured MEG signals (middle), synthesized signals (right) are represented. The horizontal axis expresses time (sec.) and the vertical axis expresses amplitude (pT).

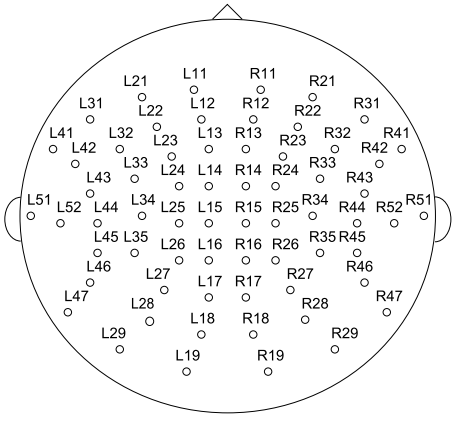

Fig. 2 Sensor distribution.

Table 1 Artificial evoked fields.

	peak time	dipole location(x,y,z)	dipole moment
EF1	0.25 (sec.)	10, 50, 50 (mm)	20 (nAm)
EF2	0.27 (sec.)	−40, 40, 40 (mm)	30 (nAm)

4.2 Source Decomposition

In this subsection, we demonstrate the decomposition

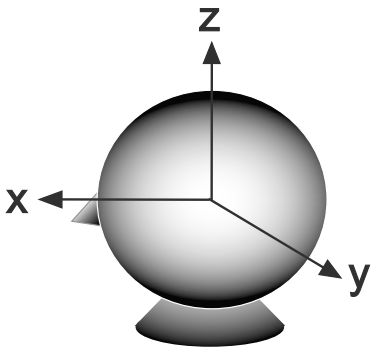

Fig. 3 3D frames of reference.

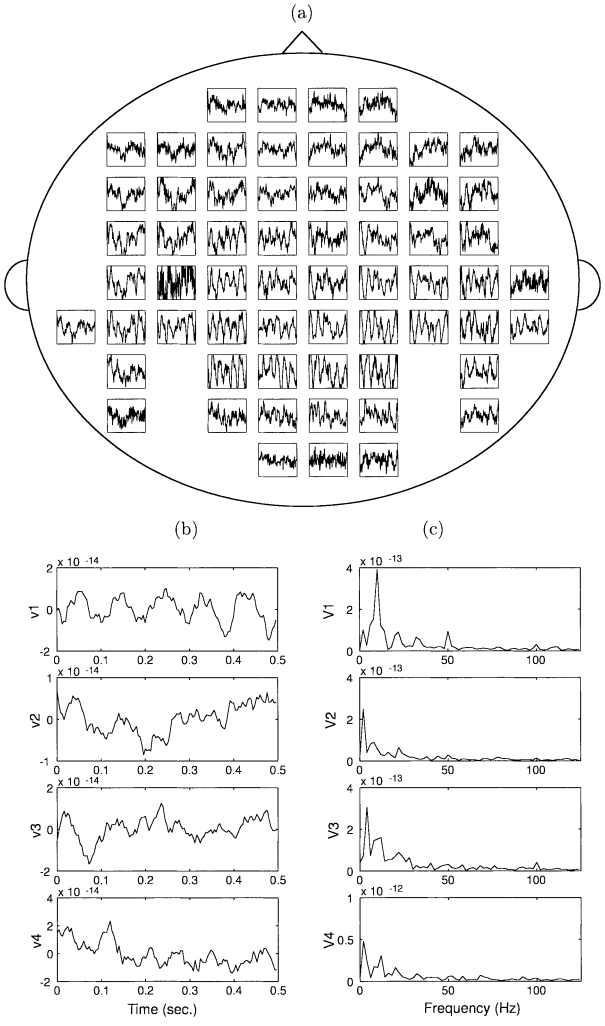

Fig. 4 (a) 1st. single-trial data, top view. The horizontal axis expresses time from 0 to 0.5 sec. and the vertical axis expresses amplitude from −0.5 to 0.5 pT. (b) Result of ICA with the robust pre-whitening technique. (c) Frequency contents of (b).

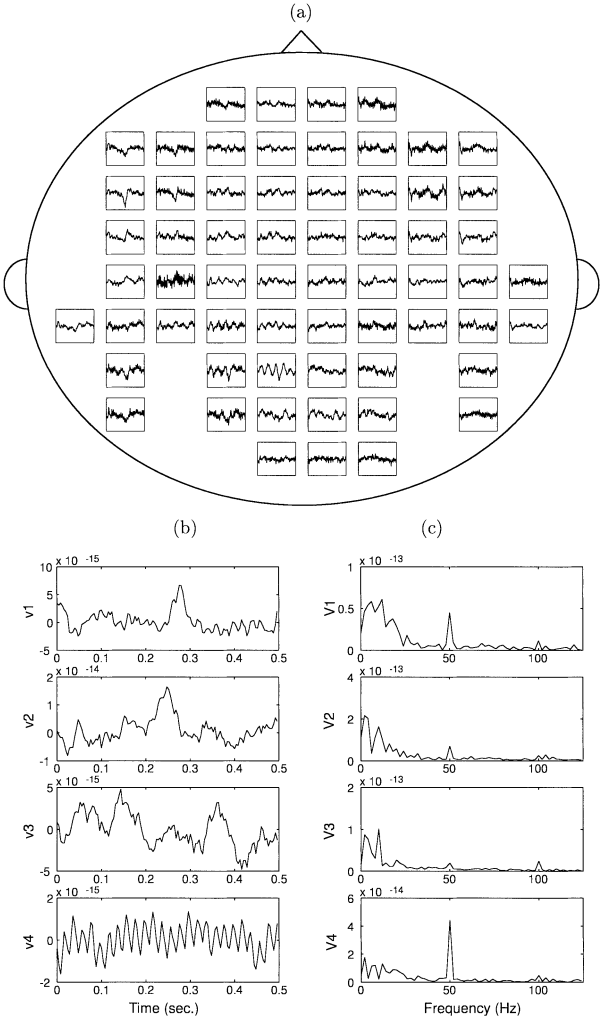

Fig. 5 (a) Averaged of 10 trials from 1 to 10 trial data, top view. (b) Result of ICA with the robust pre-whitening technique. (c) Frequency contents of (b).

of the source by means of the robust pre-whitening technique with the JADE algorithm for the single-trial data, the averaged 10-trials data and the averaged 100-trials data, as shown in Figs. 4–6(a). In each result, the number of sources is assumed to be $\widehat{n} = 4$. The results $\mathbf{v}_k(t)$ $(t = 0, \cdots, 0.5\,\text{sec.},\ k = 1, \cdots, 4)$ are shown in Figs. 4–6(b). Using a 125-point short time Fourier transform for $\mathbf{v}_k(t)$, we calculated the power spectrum of each decomposed component $\mathbf{V}_k(f)$ $(f = 0, \cdots, 125\,\text{Hz})$. See Figs. 4–6(c).

In our study, we focused on the decomposed signals EF1 and EF2, the 50 Hz electrical power interference and the α-wave component. We will now demonstrate the procedure for automatic classification of the decomposed components into signals EF1 and EF2, the 50 Hz electrical power interference and the α-wave component.

First, we calculated the power of each decomposed component in the time domain $P_{\mathbf{v}k} = \sum_{t=0}^{0.5} \mathbf{v}_k(t)\mathbf{v}_k^T(t)$ and frequency domain $P_{\mathbf{V}k} = \sum_{f=0}^{125} \mathbf{V}_k(f)\mathbf{V}_k^T(f)$, where $\mathbf{v}_k(t)$ denotes the k-th decomposed component and $\mathbf{V}_k(f)$ denotes the power spectrum of $\mathbf{v}_k(t)$. To define the criterion for classifying the decomposed components into EF1 and EF2, we calculated the power of

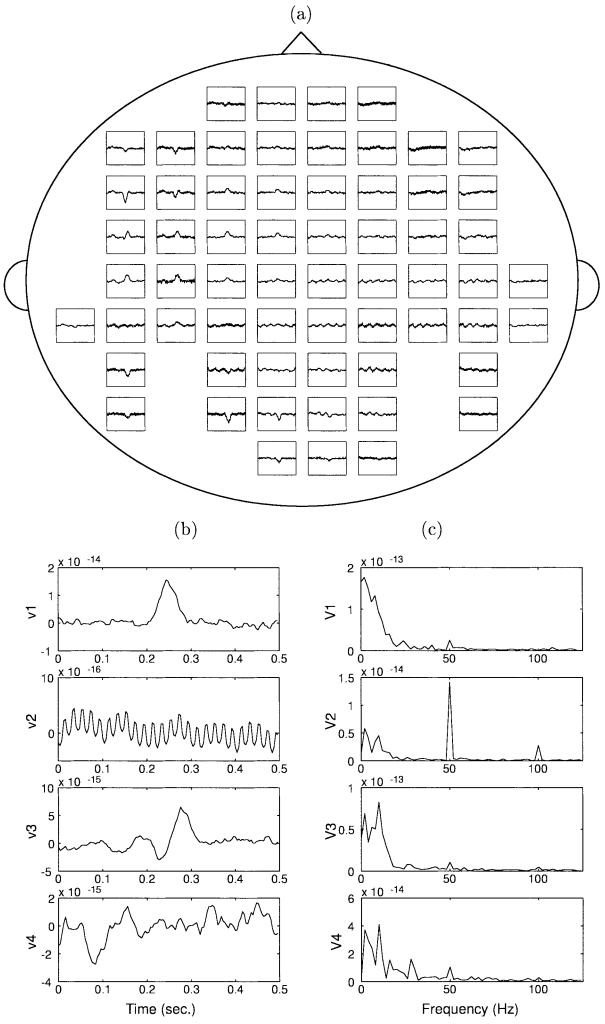

Fig. 6 (a) Averaged of 100 trials from 1 to 100 trial data, top view. (b) Result of ICA with the robust pre-whitening technique. (c) Frequency contents of (b).

each decomposed component in the duration from 0.2 to 0.3 sec. $P'_{\mathbf{v}k} = \sum_{t=0.2}^{0.3} \mathbf{v}_k(t)\mathbf{v}_k^T(t)$ and from 0.22 to 0.32 sec. $P''_{\mathbf{v}k} = \sum_{t=0.22}^{0.32} \mathbf{v}_k(t)\mathbf{v}_k^T(t)$. To define the criterion for classifying the decomposed components into the 50 Hz electrical power interference and α-wave component, we calculated the power in the frequency domain from 48 to 52 Hz $P'_{\mathbf{V}k} = \sum_{f=48}^{52} \mathbf{V}_k(f)\mathbf{V}_k^T(f)$ and from 8 to 12 Hz $P''_{\mathbf{V}k} = \sum_{f=8}^{12} \mathbf{V}_k(f)\mathbf{V}_k^T(f)$. Here we define the ratios $R'_{\mathbf{v}k} = \frac{P'_{\mathbf{v}k}}{P_{\mathbf{v}k}}$, $R''_{\mathbf{v}k} = \frac{P''_{\mathbf{v}k}}{P_{\mathbf{v}k}}$, $R'_{\mathbf{V}k} = \frac{P'_{\mathbf{V}k}}{P_{\mathbf{V}k}}$ and $R''_{\mathbf{V}k} = \frac{P''_{\mathbf{V}k}}{P_{\mathbf{V}k}}$. When $R'_{\mathbf{v}k} \geq k_{EF1}$, the decomposed component $\mathbf{v}_k$ is the signal EF1, since the signal EF1 was artificially evoked from 0.2 to 0.3 sec. where k_{EF1} is a positive constant. Similarly, when $R''_{\mathbf{v}k} \geq k_{EF2}$, $R'_{\mathbf{V}k} \geq k_e$ or $R''_{\mathbf{V}k} \geq k_\alpha$, the decomposed component $\mathbf{v}_k$ is the signal EF2, the 50 Hz electrical power interference or α-wave component, respectively, where k_{EF2}, k_e and k_α are positive constants. Based on prior experience, for this experiment we set these parameters as $k_{EF1} = 0.6$, $k_{EF2} = 0.6$, $k_e = 0.4$ and $k_\alpha = 0.35$, respectively. The parameters k were determined that only one component (EF1, EF2, electrical interference or α-wave) could be decided from one decomposed component and the number of definite components from all decomposed components will become the maximum.

We demonstrate the results for the single-trial data shown in Fig. 4. The result is that $\mathbf{v}_1$ has satisfied the equation $R''_{\mathbf{V}1} \geq k_\alpha$, so this component can be confirmed as α-wave component. As $\mathbf{v}_2$, $\mathbf{v}_3$ and $\mathbf{v}_4$ have not satisfied the other criteria, we know that these components cannot be decomposed.

In the results for the averaged 10-trials data shown in Fig. 5, $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ and $\mathbf{v}_4$ have satisfied the equations $R''_{\mathbf{v}1} \geq k_{EF2}$, $R'_{\mathbf{v}2} \geq k_{EF1}$, $R''_{\mathbf{V}3} \geq k_\alpha$ and $R'_{\mathbf{V}4} \geq k_e$, respectively, and can thus be regarded as signals EF2, EF1, α-wave component and electrical power interference, respectively.

In the results for the averaged 100-trials data shown in Fig. 6, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ have satisfied the equation $R'_{\mathbf{v}1} \geq k_{EF1}$, $R'_{\mathbf{V}2} \geq k_e$ and $R''_{\mathbf{v}3} \geq k_{EF2}$, respectively, and can thus be regarded as EF1, electrical power interference and EF2, respectively. $\mathbf{v}_4$ has not satisfied the other criterion, revealing that this component cannot be decomposed. It is actually environment interference, whose power is reduced by taking an average.

4.3 Dipole Localization

In this subsection, we discuss dipole location of the decomposed component treated in Sect. 4.2. We used the standard spatio-temporal dipole fitting routine, MEG v3.3a (CTF System Inc., Canada), to find dipole location. The estimated maps are shown in Fig. 7 and dipole locations of individual components are shown in Table 2. In Fig. 7, the top 'Measured' map was derived from a decomposed component. The middle 'Theoretical' map was computed by the moving dipoles. The bottom 'Difference' map represents the difference between the 'Measured' and 'Theoretical' maps. The amplitude information appears in the color scale bar.

In this study, since dipole locations of EF1 and EF2 were known in advance, we can compare them to the estimated one. Here we define the distance between the true dipole location $[x,\ y,\ z]$ mm and the estimated dipole location $[\widehat{x},\ \widehat{y},\ \widehat{z}]$ mm as

$$r = \sqrt{(x-\widehat{x})^2 + (y-\widehat{y})^2 + (z-\widehat{z})^2}, \tag{19}$$

where the true dipole location of EF1 was [10, 50, 50] and dipole location of EF2 was [−40, 40, 40].

Applying this routine, we demonstrate dipole localization derived by analyzing the single-trial data. The estimated map of α-wave component is shown in Fig. 7(a). In this result, the estimated map of α-wave component appears on the back area of the brain. The estimated dipole location is [−46.5, − 6.2, 57.8] mm

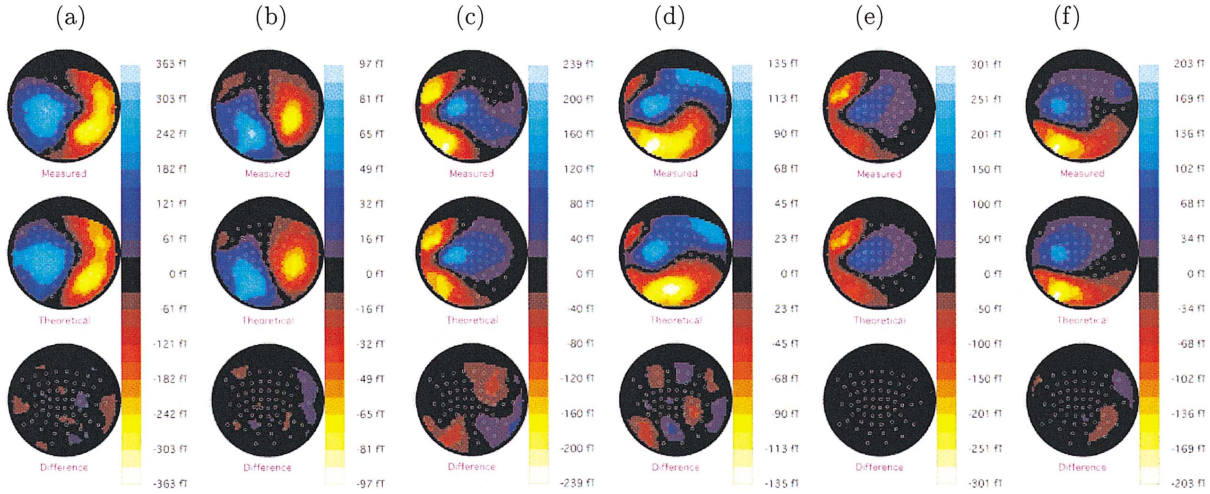

Fig. 7 Estimated maps: (a) Estimated map for single-trial data focus on α-wave. (b) Estimated map for averaged 10-trials data focus on α-wave. (c) Estimated map for averaged 10-trials data focus on EF1. (d) Estimated map for averaged 10-trials data focus on EF2. (e) Estimated map for averaged 100-trials data focus on EF1. (f) Estimated map for averaged 100-trials data focus on EF2. In each map, the measured map (top), the theoretical map (middle) and the difference between measured and theoretical map (bottom) are represented.

Table 2 Estimated dipole location (mm).

(a) Single-trial data.

	x	y	z	r
decomposed α-wave	−46.5	−6.2	57.8	-

(b) Averaged 10-trials data.

	x	y	z	r
decomposed α-wave	−50.8	−4.4	55.0	-
averaged EF1	7.8	16.5	46.4	33.8
decomposed EF1	−5.7	46.4	47.5	16.3
averaged EF2	−38.2	12.5	52.3	30.2
decomposed EF2	−37.5	24.1	52.9	20.6

(c) Averaged 100-trials data.

	x	y	z	r
averaged EF1	5.6	57.3	48.1	8.7
decomposed EF1	8.9	47.0	50.9	3.4
averaged EF2	−43.0	37.5	44.4	5.9
decomposed EF2	−37.1	37.0	36.3	5.6

and the maximum amplitude is 363 fT (see Table 2(a)).

The estimated maps of α-wave component, signals EF1 and EF2, derived by analyzing the average of 10 trials, are shown in Figs. 7(b), (c), (d), respectively. The result of α-wave component is that the estimated map and dipole location $[-50.8,\ -4.4,\ 55.0]$ mm are close to those in the previous result, but the amplitude is 97 fT, which is smaller than the result derived by the single-trial data. Comparing the two results of the single-trial and the averaged 10-trials data, we can conclude that the power of decomposed α-wave component is reduced by taking an average across the data trials. As for the result of EF1, note that the evoked response appears on the left-front area of the brain. The estimated dipole location becomes $[-5.7,\ 46.4,\ 47.5]$ mm. Using Eq. (19), the distance between the true and estimated dipole location of EF1 is 16.3 mm. For the result of EF2, the evoked response appears on the left-back area of the brain. The estimated dipole location is $[-37.5,\ 24.1,\ 52.9]$ mm. Therefore the distance between the true and estimated dipole locations of EF2 is 20.6 mm. Here, we describe the results of dipole estimation for averaged 10 trials data. The results of dipole estimation for averaged 10 trials data become $[7.8,\ 16.5,\ 46.4]$ mm for EF1 and $[-38.2,\ 12.5,\ 52.3]$ mm for EF2. Therefore the distances between the true and estimated dipole locations are 33.8 mm for EF1 and 30.2 mm for EF2. Comparing the results for applying ICA and taking averages, we see that dipole location derived by applying ICA is more accurate than that of taking averages. We conclude that dipole locations of evoked fields become more accurate by applying ICA approach (see Table 2(b)).

The estimated maps of the EF1 and EF2 derived by analyzing the averaged 100-trials data are shown in Figs. 7(e), (f), respectively. In the results of EF1 and EF2, note that the evoked responses appear like the maps derived by analyzing the averaged 10-trials data. The estimated dipole locations are $[8.9,\ 47.0,\ 50.9]$ mm at EF1 and $[-37.1,\ 37.0,\ 36.3]$ mm at EF2. Therefore the distances between the true and estimated dipole locations are 3.4 mm for EF1 and 5.6 mm for EF2. The results of dipole estimation for averaged 100 trials data become $[5.6,\ 57.3,\ 48.1]$ mm for EF1 and $[-43.0,\ 37.5,\ 44.4]$ mm for EF2. Therefore the distances between the true and estimated dipole locations are 8.7 mm for EF1 and 5.9 mm for EF2. Comparing the results for applying ICA and taking averages, we see

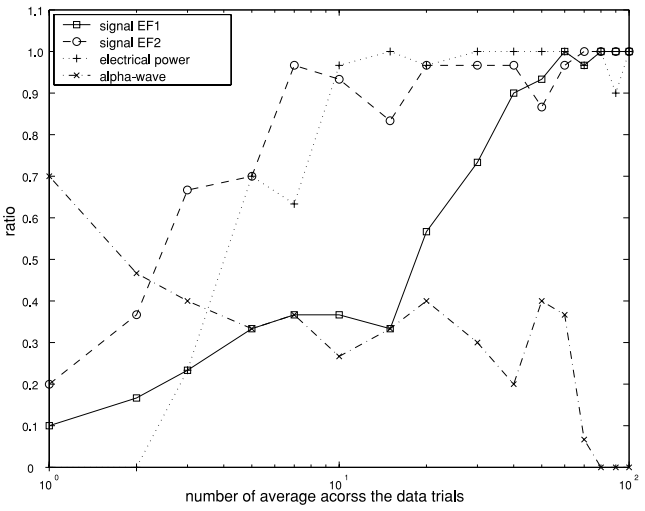

Fig. 8 Decomposition rates and the number of averages across data trials.

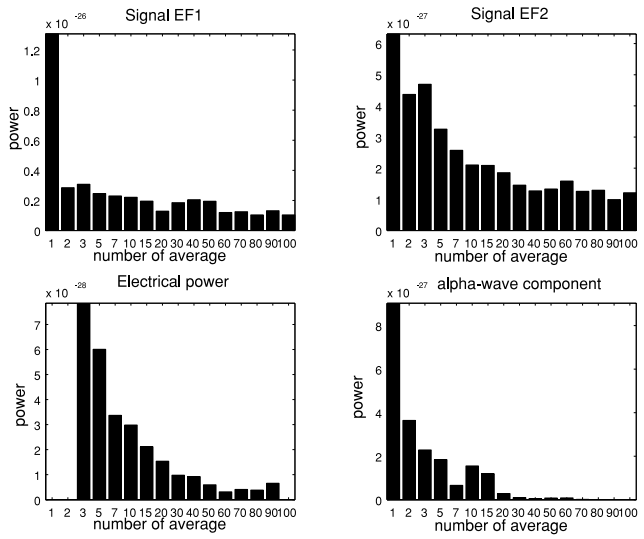

Fig. 9 Powers of decomposed components and the number of averages across data trials.

that dipole location derived by applying ICA is more accurate than that of taking averages. We conclude that the ICA approach is very efficient for dipole estimation (see Table 2(c)). Comparing the results of the averaged 10 and 100-trials data, we see that dipole location derived from the averaged 100-trials data is more accurate than that of 10-trials data. We conclude that dipole locations of evoked fields are more accurate when an average is taken across sets of data trials.

4.4 Comparison of the Decomposed Components

In this subsection, we demonstrate the relationship between the decomposed component and the number of averages across data trials. For the input signal we take averages across different number of trials from 1 to 100-trials. In each number of averages, from 1 to 70-trials, we simulate by using 30 types of moving data sets and count the number of decomposed signal EF1 (n_{EF1}), signal EF2 (n_{EF2}), electrical power interference (n_e) and α-wave component (n_α). To investigate the performance of source decomposition, we calculate the ratios $\frac{n_{EF1}}{30}$, $\frac{n_{EF2}}{30}$, $\frac{n_e}{30}$ and $\frac{n_\alpha}{30}$. In case the number of averages is 80, 90 or 100-trials, we simulate by using 20, 10, 1 types of moving data sets, respectively, and calculate the ratios.

The relationship between these ratios and the number of averages across the data trials is shown in Fig. 8. The horizontal axis expresses the number of averages from 1 to 100 trials and the vertical axis expresses these ratios. There are several things to note in Fig. 8. The ratios at EF1, EF2 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. Also, the ratio at EF2 is higher than at EF1 since the power of EF2 (30 nAm) is larger than that of EF1 (20 nAm). Finally, the ratio of α-wave component is high when the number of averages is small, but the larger the number of averages, the lower the ratio becomes.

Next, we demonstrate the relationship between the power of each decomposed component and the number of averages. The powers of the decomposed components P are calculated by using Eq. (18). The results are shown in Fig. 9. The powers of decomposed components are high when the number of averages is small. This means that these decomposed components are influenced by high power components such as α-wave component. In contrast, the power of α-wave component is almost zero, when the number of average is large. This means that the power of α-wave component is reduced by taking an average.

Given these results we can confirm that the analysis of averaged multiple-trials data is efficient for obtaining the evoked response. However, the analysis of non-averaged single-trial data is more efficient for obtaining the strength and dynamic components such as α-wave component.

5. Conclusions

In this paper, we performed source decomposition of single-trial and averaged multiple-trials MEG data using a robust pre-whitening technique and the JADE algorithm. Our results showed that the analysis of averaged data effectively determines dipole location of the evoked fields. For strength and dynamic components such as α-wave, the single-trial data analysis is more efficient. When one requires both the range of evoked strength and dipole locations, the analysis of the averaged limited-trials data is the most efficient option.

The authors hope these results will help neuroscientists to further their understanding of the temporal

cortex.

References

[1] J. Cao and A. Cichocki, "Blind source separation algorithm based on 4th-order self- and cross-cumulant criterion applied for EEG data analysis," Proc. 1997 International Symposium on Nonlinear Theory and its Applications, vol.2, pp.1005–1008, 1997.
[2] 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.
[3] 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. Neural Netw., vol.14, no.3, pp.631–645, June 2003.
[4] 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.
[5] S. Ikeda and K. Toyama, "Independent component analysis for noisy data—MEG data analysis," Neural Networks 13, pp.1063–1074, 2000.
[6] S. Ikeda, "ICA on noisy data: A factor analysis approach," in Advances in Independent Component Analysis, ed. M. Girolami, Springer, 2000.
[7] 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.
[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] 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.
[11] 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.
[12] 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

Yoshio Konno received the B.E. degree from Sophia University, Tokyo, Japan, in 2002. He is currently a master course student in the Department of Electrical and Electronics Engineering, Sophia University. His research interests include blind signal processing.

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.

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 and Univ. of California (Berkeley, CA, USA). He is currently Associate Professor of the Department of Electrical and Electronics Engineering, Sophia Univ. He was a short-term visiting scientist at several institutions, including M.I.T. and Max Planck Institute for Psycholinguistics. His research interests include signal processing, acoustics, speech and hearing sciences, and spoken language processing.

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).