A Robust Approach to Independent Component Analysis of Signals With High-Level Noise Measurements

Jianting Cao, *Member, IEEE*, Noboru Murata, Shun-ichi Amari, *Fellow, IEEE*, Andrzej Cichocki, and Tsunehiro Takeda

***Abstract*—In this paper, we propose a robust approach for independent component analysis (ICA) of signals that observations are contaminated with high-level additive noise and/or outliers. The source signals may contain mixtures of both sub-Gaussian and super-Gaussian components, and the number of sources is unknown. Our robust approach includes two procedures. In the first procedure, a robust prewhitening technique is used to reduce the power of additive noise, the dimensionality and the correlation among sources. A cross-validation technique is introduced to estimate the number of sources in this first procedure. In the second procedure, a nonlinear function is derived using the parameterized t-distribution density model. This nonlinear function is robust against the undue influence of outliers fundamentally. Moreover, the stability of the proposed algorithm and the robust property of misestimating the parameters (kurtosis) have been studied. By combining the t-distribution model with a family of light-tailed distributions (sub-Gaussian) model, we can separate the mixture of sub-Gaussian and super-Gaussian source components. Through the analysis of artificially synthesized data and real-world magnetoencephalographic (MEG) data, we illustrate the efficacy of this robust approach.**

***Index Terms*—Cross-validation method, independent component analysis (ICA), parametric estimation method, principal component analysis (PCA), robust prewhitening, t-distribution density model, unaveraged single-trial MEG data analysis.**

I. Introduction

BLIND separation of independent sources has received a great deal of attention due to various applications in science and technology. The problem of blind source separation (BSS) and/or ICA has been studied by many researchers in the fields of neural networks and statistical signal processing [1]–[5], [9], [15], [16], [18], [22], [26], [31], [36] during the past ten years, and many interesting theoretical and practical results have been achieved.

The particular ICA model considered in this paper is

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\xi}(t), \qquad t = 1, 2, \ldots \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ represents the transpose of m observations at time t. Each observation $x_i(t)$ contains n common components (sources) $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ and a unique component (additive noise), which is included in the vector $\boldsymbol{\xi}(t) = [\xi_1(t), \ldots, \xi_m(t)]^T$. $\mathbf{A}$ is a full-rank $m \times n$ mixing matrix.

In the model, the sources $\mathbf{s}$ and their number n, additive noise $\boldsymbol{\xi}$ and matrix $\mathbf{A}$ are unknown, but the sensor signals $\mathbf{x}$ are accessible. The sensor signals $\mathbf{x}$ contain mixtures of both sub- and super-Gaussian source components. It is assumed that the components of $\mathbf{s}$ are mutually statistically independent, as well as statistically independent of the noise components $\boldsymbol{\xi}$. Moreover, the noise components $\boldsymbol{\xi}$ themselves are assumed to be mutually decorrelated. Our goal is to estimate the independent sources under these challenging conditions or assumptions.

The topics considered in the present paper are related to various original contributions. They include the ICA noisy model in [7], [20], [24], [27]–[29], the estimation of the number of sources in [21], [32], the natural gradient and/or the relative gradient-based ICA algorithms with stability analysis in [1]–[5], [16], the optimal nonlinear functions or separation of mixtures of sub- and super-Gaussian source signals in [11], [17], [19], [23], [25], [34], [37], and the applications of ICA to averaged and unaveraged MEG/EEG data in [10], [12]–[14], [29], [30], [35], [38]. These references will be referred to again in the subsequent sections.

In the analysis of real-world MEG/EEG data one is faced with problems such as the different nature of source signals (e.g., both sub- and super-Gaussian sources exist), unknown number of sources, and contamination of the sensor signals with a high level (power) of additive noise and outliers. We propose a robust approach to the solution of these problems based on the subspace and the parametric methods to analyze the independent components. Our robust approach includes two procedures. In the first procedure, observations (noisy data) possessing high dimensionality are first decomposed into a source signal subspace and a noise subspace; the dimensionality is reduced optimally. In the second procedure, the transformed low-dimensional signals containing both sub-Gaussian and super-Gaussian components are further separated using the parameterized t-distribution density model and the light-tailed distribution density model with the natural gradient-based ICA algorithm.

This paper is organized as follows. A robust prewhitening technique with noise reduction and a cross-validation technique with optimal dimensionality reduction are presented in Sec-

Manuscript received April 15, 2001; revised May 2, 2002.

J. Cao is with the Department of Electronic Engineering, Saitama Institute of Technology, Saitama 369-0293, Japan, and is also with the Brain Science Institute, RIKEN, Saitama 351-0198, Japan (e-mail: cao@sit.ac.jp).

N. Murata is with the Department of Electrical and Electronics Engineering, Waseda University, Tokyo 169-8555, Japan.

S. Amari and A. Cichocki are with the Brain Science Institute, RIKEN, Saitama 351-0198, Japan.

T. Takeda is with the Department of Complexity of Science and Engineering, Graduate School of Tokyo University, Tokyo 113-8656, Japan.

Digital Object Identifier 10.1109/TNN.2002.806648

tion II. The parameterized $\mathbf{t}$-distribution model and its robust properties, as well as the stability of the developed algorithm are presented in Section III. Experimental results using this new approach on artificially synthesized data and real-world unaveraged single-trial MEG data are presented in Section IV. The MEG data are from an experiment studying the auditory evoked fields (AEF) task. Some conclusions are drawn and presented in Section V.

II. Robust Prewhitening Technique

In this section, we first describe the standard principal components analysis (PCA) approach, which has been adopted in some promising ICA algorithms [9], [15], [26] for the prewhitening step. Next, we show that this standard PCA approach can be extended to prewhitening of data with noise reduction. Finally, the practical implementation of the robust prewhitening algorithm is presented.

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

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

where N denotes data samples. When the sample size N is sufficiently large, the covariance matrix of the observed data can be written as

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

where $\mathbf{\Sigma} = \mathbf{X}\mathbf{X}^T/N$, and $\mathbf{\Psi} = \mathbf{\Xi}\mathbf{\Xi}^T/N$ is a diagonal matrix. Since the sources are mutually independent, as well as independent from the noise, the terms $\mathbf{S}\mathbf{S}^T/N \to \mathbf{I}$ and $\mathbf{S}\mathbf{\Xi}^T/N \to \mathbf{0}$, disappearing in (3). For convenience, we assume that $\mathbf{X}$ has been divided by $\sqrt{N}$ so that the covariance matrix can be given by $\mathbf{C} = \mathbf{X}\mathbf{X}^T$.

A. Standard PCA for Prewhitening

Let us first consider an ideal case in which the noise variance $\mathbf{\Psi}$ is close to zero, or in other words, the signal-to-noise ratio (SNR) is very high. In the context of biomedical signal processing, this condition can usually be achieved by taking the average of a large number of trials.

For this kind of data, a cost function for fitting the model to the data can be employed to make $\mathbf{C} - \mathbf{A}\mathbf{A}^T$ as small as possible. It is well known that the standard PCA approach is used to find the principal components using the eigenvalue decomposition method; that is, the solution of $\mathbf{A}\mathbf{A}^T$ for seeking n principal components can be obtained by

$$\hat{\mathbf{A}}\hat{\mathbf{A}}^T = \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^T \tag{4}$$

where $\mathbf{\Lambda}_n$ is a diagonal matrix whose elements are the n largest eigenvalues of $\mathbf{C}$. The columns of $\mathbf{U}_n$ are the corresponding eigenvectors. In (4), let one possible solution for $\hat{\mathbf{A}}$ be

$$\hat{\mathbf{A}} = \mathbf{U}_n\mathbf{\Lambda}_n^{1/2}. \tag{5}$$

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

$$\mathbf{z} = \mathbf{\Lambda}_n^{-1/2}\mathbf{U}_n^T\mathbf{x}. \tag{6}$$

Using this result, the covariance matrix is obtained as $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}_n$, which means that $\mathbf{z}$ are orthogonal, or the components are decorrelated in the new set of transformation data.

This standard approach or a similar decorrelation procedure has been adopted in some promising ICA algorithms [9], [15], [26] for the prewhitening procedure. Applying these algorithms to the analysis of averaged EEG/MEG data (SNR is high), some successful results have been reported [35], [38].

B. Prewhitening With Noise Reduction

Let us consider a more difficult or challenging case, however, in which the power of the noise is larger than that of the source signal (SNR < 0 dB). This situation usually happens with MEG raw data. In this case, the diagonal elements in the matrix $\mathbf{\Psi}$ are relatively large and, therefore, cannot be ignored in the model.

Similar to the standard PCA approach, here we can fit $\mathbf{A}\mathbf{A}^T$ to $\mathbf{C} - \mathbf{\Psi}$ using the eigenvalue decomposition method. We choose the columns of $\mathbf{A}$ as eigenvectors of $\mathbf{C} - \mathbf{\Psi}$ corresponding to the n largest eigenvalues so that the sum of the squares in each column is identical to the corresponding eigenvalue.

It should be noted that the noise covariance $\mathbf{\Psi}$ is assumed to be known in the above case. However, the noise covariance $\mathbf{\Psi}$ is usually unknown in the real-world problem and, therefore, it has to be estimated. Motivated by this, we employ the cost function

$$L(\mathbf{A}, \mathbf{\Psi}) = \operatorname{tr}\left[\mathbf{A}\mathbf{A}^T - (\mathbf{C} - \mathbf{\Psi})\right]\left[\mathbf{A}\mathbf{A}^T - (\mathbf{C} - \mathbf{\Psi})\right]^T \tag{7}$$

and minimize it by $(\partial L(\mathbf{A}, \mathbf{\Psi}))/\partial\mathbf{\Psi} = \mathbf{0}$, whereby the estimate of $\mathbf{\Psi}$ can be obtained as

$$\hat{\mathbf{\Psi}} = \operatorname{diag}\left(\mathbf{C} - \hat{\mathbf{A}}\hat{\mathbf{A}}^T\right). \tag{8}$$

In (8), the estimate $\hat{\mathbf{A}}$ can be obtained in the same way as in (5).

Note that both the matrix $\mathbf{A}$ and the diagonal elements of $\mathbf{\Psi}$ need to be estimated together from the data. The estimate $\hat{\mathbf{A}}$ is obtained by the standard PCA approach using the eigenvalue decomposition method. The estimate $\hat{\mathbf{\Psi}}$ is obtained by the unweighted least squares method, which is one of the estimation methods used in factor analysis [33].

Once the estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{\Psi}}$ converge to stable values, we need to finally compute the score matrix, or the pseudoinverse matrix. Since the solution for a pseudoinverse matrix is not unique, in this paper, we employ the Bartlett method [8], which is an unbiased model. The noise variance $\mathbf{\Psi}$ is taken into the calculation, that is

$$\mathbf{Q} = \left[\hat{\mathbf{A}}^T\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{A}}\right]^{-1}\hat{\mathbf{A}}^T\hat{\mathbf{\Psi}}^{-1}. \tag{9}$$

Using this result, a new set of transformation data can be obtained by $\mathbf{z} = \mathbf{Q}\mathbf{x}$.

Note that the covariance matrix is $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}_n + \mathbf{C}\mathbf{\Psi}\mathbf{C}^T$, which implies that the source signals are decorrelated in this subspace. Therefore, the robust prewhitening technique plays the same role in decorrelation as in standard PCA, but the noise variance $\mathbf{\Psi}$ is taken into account with the former. The difference between the two methods is that standard PCA fits both

diagonal and off-diagonal elements of $\mathbf{C}$, whereas the robust prewhitening technique fits only the off-diagonal elements of $\mathbf{C}$. Based on this property, one can take advantage in reducing the high-power noise using the robust prewhitening technique, particularly in the case in which the power of noise is larger than that of the signal. Another advantage is that the robust prewhitening technique is robust to non-Gaussian unique outlier, since it does not assume close adherence to the normal distribution in the model assumption.

Some similar noise reduction approaches using statistical methods such as the maximum-likelihood method of factor analysis has been proposed in [7], [29]. The advantage of these methods is that there is a statistical measure available for assessing how well the model fits the data. However, when sensor signals are contaminated with high-power Gaussian noise or non-Gaussian outliers, such as with single-trial, MEG data, the algorithms derived using these methods work inefficiently.

C. Estimation of the Number of Sources

The cross-validatory techniques have been injudiciously applied in multivariate statistics [32]. The basic idea of a cross-validatory method is that it divides the data into several subgroups. One group is used to determine some features of the data, and the other groups are used to verify the features. Using this technique, we propose a criterion for estimating the number of sources $\hat{n}$ based on the error of estimated noise variance.

Let us first divide the data matrix $\mathbf{X}$ into several disjoint groups, such as $\mathbf{X}_i \in \mathbf{R}^{m\times N/K}$, where N is the number of data samples and the group number $i = 1, \ldots, K$. Next, we delete each group in turn from the data matrix and compute one estimate of the noise variance as $\mathrm{diag}(\hat{\mathbf{\Psi}}_i)$; we use remaining data to compute another estimate of the noise variance as $\mathrm{diag}(\hat{\mathbf{\Psi}}_j)$, where $j \neq i$. It is obvious when the estimate of source number $\hat{n}$ has not been matched to its true value, there will be a larger error between the noise variance and its estimate. Based on this property, we define the criterion for each conjectured source number $\hat{n}$ as

$$\mathrm{Error}(\hat{n}) = \frac{1}{K}\sum_{i=1}^{K} \mathrm{tr}\left[\mathrm{diag}\left(\hat{\mathbf{\Psi}}_i^{(\hat{n})}\right) - \mathrm{diag}\left(\hat{\mathbf{\Psi}}_j^{(\hat{n})}\right)\right]^2. \tag{10}$$

When the "disjoint" condition is relaxed, we can replace one of the estimates, $\mathrm{diag}(\hat{\mathbf{\Psi}}_i^{(\hat{n})})$ or $\mathrm{diag}(\hat{\mathbf{\Psi}}_j^{(\hat{n})})$, using the estimate $\mathrm{diag}(\hat{\mathbf{\Psi}}^{(\hat{n})})$ that will be computed for the total data samples.

It should be noted that it is unnecessary to compute all the estimates of the source number, such as from $\hat{n} = 1$ to $\hat{n} = m$, where m denotes the number of sensors, since the estimated number of sources should be within the bound [6] as $\hat{n} \leq (1/2)(2m + 1 - \sqrt{8m+1})$.

D. Summary

The computation for the robust prewhitening technique is summarized as follows.

1) Calculate the covariance matrix $\mathbf{C}$ from data and set an initial guess of $\mathbf{\Psi}$, e.g., $\mathbf{\Psi}(0) = \mathrm{diag}(\mathbf{C}^{-1})^{-1}$.
2) Given an initial number, such as $\hat{n} = 1$, divide the data matrix as $\mathbf{X}_i \in \mathbf{R}^{m\times N/K}$, then calculate the error using (10).
3) Calculate the $\hat{n}$ largest eigenvalues $\mathbf{\Lambda}_{\hat{n}}$ and the associated eigenvectors $\mathbf{U}_{\hat{n}}$ of $\mathbf{C}$.
4) Calculate the estimate $\hat{\mathbf{A}}$ using (5) and the estimate $\hat{\mathbf{\Psi}}$ using (8).
5) Repeat calculations 3) and 4) until the estimates converge to stable values.
6) Set the conjectured number as $\hat{n} \Leftarrow \hat{n} + 1$, repeat 2) to 5) until the condition $\hat{n} \leq (1/2)(2m + 1 - \sqrt{8m+1})$ is satisfied.
7) Seek the minimum error of $\hat{n}$ using (10) and calculate the pseudoinverse of estimate $\hat{\mathbf{A}}$ using (9); finally, obtain a new set of transformation data by $\mathbf{z} = \mathbf{Q}\mathbf{x}$.

It should be noted that it is necessary to use the robust prewhitening technique (i.e., decorrelation procedure) to reduce the power of the noise as well as the number of parameters to be estimated. It is insufficient to obtain the independent components, since an orthogonal matrix in general contains additional degrees of freedom. Therefore, the remaining parameters need to be estimated by ICA.

III. Robust Nonlinear Function in the ICA Algorithm

After prewhitening of the noisy observations, the transformed signals $\mathbf{z} = \mathbf{Q}\mathbf{x}$ are obtained through a procedure in which the power of noise, mutual correlation and dimensionality have been reduced. The decomposed independent sources $\mathbf{y} \in \mathbf{R}^n$ can then be obtained from a linear transformation as

$$\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t) \tag{11}$$

where $\mathbf{W} \in \mathbf{R}^{n\times n}$ is the demixing matrix, which can be computed using several kinds of ICA algorithms [1], [15], [26]. In this paper, we employ the nature gradient-based ICA algorithm to compute matrix $\mathbf{W}$.

The Kullback–Leibler divergence is one ICA contrast function that measures the mutual stochastic independence of the output signals y_i between the joint probability density function (pdf) $p_y(\mathbf{y})$ and the marginal pdfs $p_i(y_i)$ as

$$D(\mathbf{y}|\mathbf{W}) = \int p_y(\mathbf{y}) \log \frac{p_y(\mathbf{y})}{\prod_{i=1}^{n} p_i(y_i)}\, d\mathbf{y}. \tag{12}$$

In (12), the KL 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.

The natural gradient-based algorithm was developed by Amari [1], [4], [5] and independently by Cardoso, which was termed the relative gradient [16]. It has been proven that the natural gradient greatly improves learning efficiency in blind source separation. Applying the natural gradient to minimize KL divergence (12), the general learning rule for updating $\mathbf{W}$ can be developed as [1]

$$\Delta\mathbf{W}(t) = \eta\left[\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(t))\mathbf{y}^T(t)\right]\mathbf{W}(t). \tag{13}$$

When prewhitening has been performed and the demixing matrix $\mathbf{W}$ is restricted to be orthogonal in form, the algorithm can be extended to the following form as [5]:

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

When the prewhitening is included in the separation process, the algorithm becomes the EASI algorithm [16]

$$\Delta\mathbf{W}(t) = \eta\left[\mathbf{I} - \mathbf{y}(t)\mathbf{y}^T(t) - \boldsymbol{\varphi}(\mathbf{y}(t))\mathbf{y}^T(t) + \mathbf{y}(t)\boldsymbol{\varphi}^T(\mathbf{y}(t))\right]\mathbf{W}(t). \quad (15)$$

In (13)–(15), $\eta > 0$ is a learning rate, and $\boldsymbol{\varphi}(\cdot)$ is the vector of activation functions whose optimal components are

$$\varphi_i(y_i) = -\frac{d}{dy_i}\log p_i(y_i) = -\frac{\dot{p}_i(y_i)}{p_i(y_i)} \quad (16)$$

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

Typical ICA algorithms, including algorithms (13)–(15), rely on the appropriate choice of nonlinear score functions. The optimal function (16) depends on the probability distribution of the source, which is usually not available in the ICA task. Therefore, one practical solution to this problem is to employ a hypothetical probability density model. Several algorithms have been developed for separating the mixtures of sub- and super-Gaussian sources [11], [17], [19], [23], [25], [34], [37].

For bimodal distributed sources, employing a mixed Gaussian density model has been proposed in [25], [34]. The developed algorithm based on this model is elegant and enables us to separate the mixtures of sub- and super-Gaussian sources. However, the sign of the hyperbolic-tangent function in the algorithm is determined by the sign of the kurtosis, which does not always correspond to the source distribution.

The generalized Gaussian distributions are comprised of unimodal heavy-tailed (super-Gaussian) and light-tailed distributions (sub-Gaussian). The parametric method using the generalized Gaussian density model has been developed in [37], [17]. In the method, a shape parameter is introduced to represent not only the distribution of source but also the value of the source kurtosis. However, several discontinuities exist in the heavy-tailed distribution density model, and these discontinuities lead to the instability of algorithm.

In this paper, we introduce the $\mathbf{t}$-distribution, which is a family of heavy-tailed distributions with degrees of freedom (a shape parameter). Two advantages exist in applying the $\mathbf{t}$-distribution density model: 1) no discontinuity exists in the $\mathbf{t}$-distribution density model, hence, the derived algorithm is always locally stable; and 2) the nonlinear function derived from the parameterized $\mathbf{t}$-distribution density model is robust to the undue influence of outlier.

By combining the $\mathbf{t}$-distribution density model with a family of light-tailed distributions (sub-Gaussian) density model, we can separate the mixtures of sub-Gaussian and super-Gaussian source components. In this case, the nonlinear function is determined by the estimated value of the kurtosis, which corresponds to the source distribution. Preliminary results from applying the $\mathbf{t}$-distribution density model was reported in [11]. Some extensions, such as the robust property in misestimating the shape parameters (or kurtosis) and its application to a real-world problem, are presented in this paper.

A. Proposed Unimodal Distribution Density Model

In this paper, we present a unimodal distribution density model that combines the $\mathbf{t}$-distribution density model and a family of light-tailed distributions, which are a part of the generalized Gaussian distributions density model (see Fig. 1).

The pdf of the $\mathbf{t}$-distribution with a shape parameter β and a scaling factor λ_β is defined by

$$p_\beta(y) = \frac{\lambda_\beta\Gamma\left(\frac{\beta+1}{2}\right)}{\sqrt{\pi\beta}\Gamma\left(\frac{\beta}{2}\right)}\left(1 + \frac{(\lambda_\beta y)^2}{\beta}\right)^{-(1/2)(\beta+1)} \quad (17)$$

where $\Gamma(\cdot)$ is a Gamma function defined by

$$\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)\,dt. \quad (18)$$

Changing the shape parameter β within $0 < \beta < \infty$, we can obtain a family of heavy-tailed distributions (super-Gaussian) that have much higher peaks than the Gaussian distribution [see Fig. 1(a)]. In particular, when $\beta = 1$, it is identical to the Cauchy distribution, and when β approaches infinity, it reduces to the Gaussian distribution.

It should be noted that only the heavy-tailed distributions (super-Gaussian) exist in the $\mathbf{t}$-distribution. In order to establish a unimodal distribution density model that includes not only the heavy-tailed distributions but also the light-tailed distributions (sub-Gaussian), we subtract a family of light-tailed distributions from the generalized Gaussian distributions.

The pdf of the generalized Gaussian distribution with a shape parameter α and a scaling factor λ_α is represented by

$$p_\alpha(y) = \frac{\alpha\lambda_\alpha}{2\Gamma\left(\frac{1}{\alpha}\right)}\exp(-|\lambda_\alpha y|^\alpha) \quad (19)$$

where $\alpha = 2$ is the pdf of Gaussian distribution. When $\alpha < 2$, a family of the heavy-tailed distributions (super-Gaussian) are similar to that of the $\mathbf{t}$-distribution. When $\alpha > 2$, they are a family of light-tailed distributions (sub-Gaussian) and are much wider than the Gaussian distribution [see Fig. 1(b)].

The generalized Gaussian distribution density model is simple. Both heavy-tailed and light-tailed distributions can be controlled using only one parameter. However, since the parameterized heavy-tailed distributions model is not always stable and not as robust, it is worthwhile to replace it with the parameterized $\mathbf{t}$-distribution model, represented by an additional parameter. In this paper, we propose a unimodal distribution density model that is a combination of the $\mathbf{t}$-distribution density model and the light-tailed distribution density model [see Fig. 1(c)].

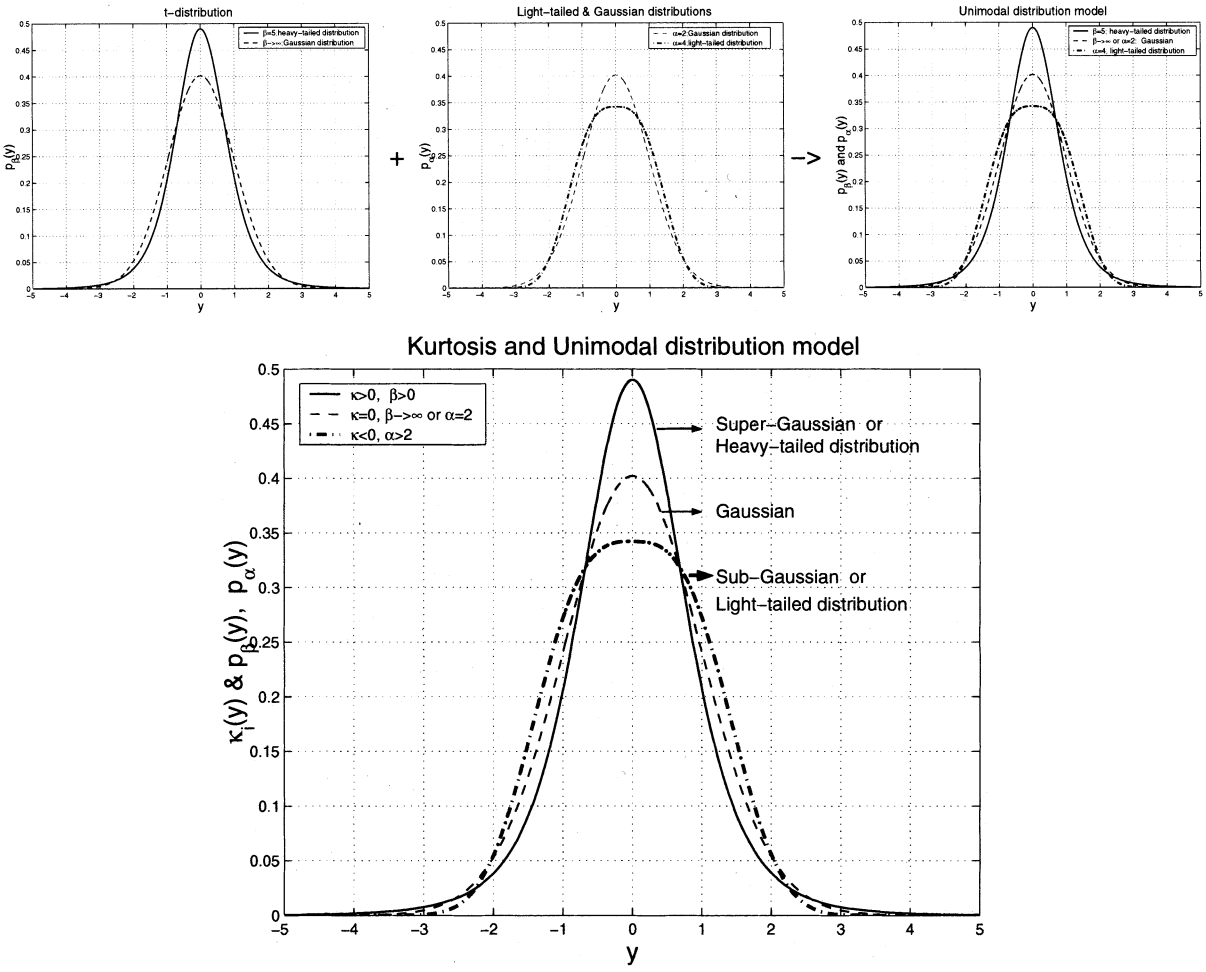

Fig. 1. Relationship between the kurtosis and the proposed unimodal distribution density model. (a) $\mathbf{t}$-distribution. (b) Light-tailed distributions subtracted from the generalized Gaussian distributions. (c) Proposed unimodal distribution density model. (d) Relationship between the kurtosis and the unimodal distribution model.

B. Kurtosis and Proposed Unimodal Distribution Density Model

Kurtosis is a quantity used to measure the peakedness or flatness of the frequency distribution of a random variable. The normalized kurtosis is defined by

$$\kappa(y) = \frac{m_4(y)}{m_2^2(y)} - 3 \tag{20}$$

where $m_2(y) = E[y^2]$ and $m_4(y) = E[y^4]$ are the second- and fourth-order moments, respectively. A positive kurtosis corresponds to a super-Gaussian distribution and a negative kurtosis corresponds to a sub-Gaussian distribution.

It should be emphasized that the value of the kurtosis corresponds to the shape parameter only in the unimodal distribution model. Based on this property, the relationship between the value of the kurtosis and the shape parameters in the $\mathbf{t}$-distribution model and the generalized Gaussian distribution density model can be established.

The second- and fourth-order moments of the $\mathbf{t}$-distribution can be derived by using formula (17). They are

$$m_2 = \int_{-\infty}^{\infty} y^2 p_\beta(y)\, dy = \frac{\beta\Gamma\left(\frac{\beta-2}{2}\right)}{2\lambda_\beta^2\Gamma\left(\frac{\beta}{2}\right)}, \qquad \beta > 2 \tag{21}$$

$$m_4 = \int_{-\infty}^{\infty} y^4 p_\beta(y)\, dy = \frac{3\beta^2\Gamma\left(\frac{\beta-4}{2}\right)}{4\lambda_\beta^4\Gamma\left(\frac{\beta}{2}\right)}, \qquad \beta > 4 \tag{22}$$

where the scaling factor λ_β can be derived using one of the equations in (21) and (22). For example, using (21), we obtain

$$\lambda_\beta = \left[\frac{\beta\Gamma\left(\frac{\beta-2}{2}\right)}{2m_2\Gamma\left(\frac{\beta}{2}\right)}\right]^{1/2}. \tag{23}$$

Substituting (21) and (22) into (20), we obtain the relative formula between the kurtosis κ_β and the shape parameter β as

$$\kappa_\beta = \frac{m_4}{m_2^2} - 3 = \frac{3\Gamma\left(\frac{\beta-4}{2}\right)\Gamma\left(\frac{\beta}{2}\right)}{\Gamma^2\left(\frac{\beta-2}{2}\right)} - 3. \tag{24}$$

It should be noted that when parameter β is given, it is easy to calculate the value of kurtosis κ_β using (24). However, in most cases, we need to obtain parameter β when given kurtosis κ_β. This can be done by establishing a lookup table in advance, and then seeking the value of β in the table that is close to κ_β.

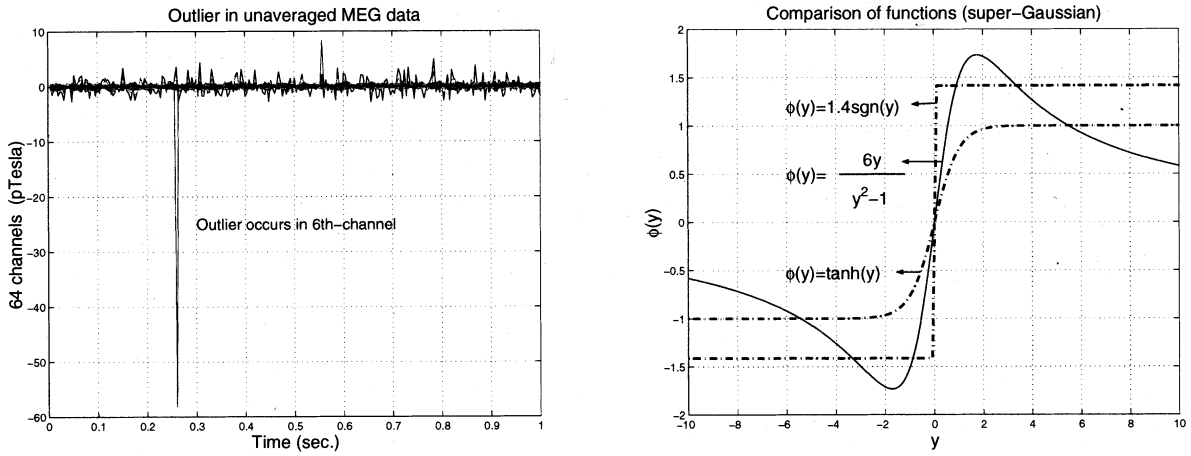

Fig. 2. (a) Example of an outlier occurring in unaveraged MEG data. (b) Comparison of the functions derived by from the heavy-tailed distribution models: 1) $\mathbf{t}$-distribution model (solid line; $\beta = 5$); 2) heavy-tailed distributions from the generalized Gaussian distribution model (dash-dotted line; $\alpha = 1$); and 3) mixture of Gaussian density model (dash-dotted line). It is clear that the function derived by from the $\mathbf{t}$-distribution model is robust to the undue influence of an outlier, which often occurs in MEG records.

Similarly, the second- and fourth-order moments of the generalized Gaussian distribution and the scaling factor can be obtained by

$$m_2 = \int_{-\infty}^{\infty} y^2 p_\alpha(y)\, dy = \frac{\Gamma\left(\frac{3}{\alpha}\right)}{\lambda_\alpha^2 \Gamma\left(\frac{1}{\alpha}\right)} \tag{25}$$

$$m_4 = \int_{-\infty}^{\infty} y^4 p_\alpha(y)\, dy = \frac{\Gamma\left(\frac{5}{\alpha}\right)}{\lambda_\alpha^4 \Gamma\left(\frac{1}{\alpha}\right)} \tag{26}$$

$$\lambda_\alpha = \left[\frac{\Gamma\left(\frac{3}{\alpha}\right)}{m_2 \Gamma\left(\frac{1}{\alpha}\right)}\right]^{1/2} \tag{27}$$

and the relative formula between the kurtosis κ_α and shape parameter α can be obtained by

$$\kappa_\alpha = \frac{\Gamma\left(\frac{5}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma^2\left(\frac{3}{\alpha}\right)} - 3 \tag{28}$$

where the heavy-tailed distributions $\alpha < 2$ are restricted.

The relationship between the kurtosis and the proposed unimodal distribution density model is shown in Fig. 1(d). Note that the curves in Fig. 1(d) are similar to those in Fig. 1(c). This means that the unimodal distribution curves can be represented not only by the parameterized unimodal distribution density model but also by the kurtosis.

C. Proposed Algorithm and Its Implementation

The optimal functions [see (16)] derived using the $\mathbf{t}$-distribution density model (17) and the light-tailed distribution density model (19) are

$$\varphi_i(y_i) = \frac{(1+\beta)y_i}{y_i^2 + \frac{\beta}{\lambda_\beta^2}}, \qquad \kappa_\beta = \hat{\kappa}_i > 0 \tag{29}$$

$$\varphi_i(y_i) = \alpha\lambda_\alpha \operatorname{sgn}(y_i)|\lambda_\alpha y_i|^{\alpha-1}, \qquad \kappa_\alpha = \hat{\kappa}_i \le 0 \tag{30}$$

where $\hat{\kappa}_i$ is the estimate of kurtosis defined by $\hat{\kappa}_i = \hat{m}_4/\hat{m}_2^2 - 3$. It can be obtained by estimating the second- and fourth-order moments using the moving-average algorithm as

$$\hat{m}_j(t) = [1-\eta]\hat{m}_j(t-1) + \eta y_i^j(t), \qquad j = 2, 4 \tag{31}$$

where $0 < \eta < 1$ is a learning rate. According to the estimate $\hat{\kappa}_i$, the function in (29) or (30) is selected automatically. Some coefficients in the functions, such as λ_β or λ_α, can be calculated using (23) or (27), respectively.

Let us consider an example in which we compare the activation functions derived by the $\mathbf{t}$-distribution density model, the heavy-tailed distributions in the generalized Gaussian distribution density model and the mixed Gaussian density model. The result is shown in Fig. 2(b). From this result, it is clear that the function $\varphi_i(y_i)$ derived by the $\mathbf{t}$-distribution density model approaches zero when the value of y_i abruptly increases. This means that the proposed function is robust to the undue influence of an outlier that often occurs in MEG raw data [see Fig. 2(a)].

The implementation of the proposed ICA algorithm is summarized as follows.

- Calculate the output $\mathbf{y}$ using (11) when given observations $\mathbf{z}$ and an initial value of $\mathbf{W}$.
- Calculate the kurtosis using (20) where the second- and fourth-order moments are estimated using (31).
- Establish two look-up tables for (24) and (28) in advance, and seek α or β from the table according to the value of $\hat{\kappa}_i$.
- Calculate the scaling factor using (23) or (27) according to the value of $\hat{\kappa}_i$.
- Calculate the nonlinear function using (29) or (30) and update $\mathbf{W}$ using (13), (14) or (15).

D. Stability Analysis of the Proposed Algorithm

The stability conditions for algorithm (13) were developed by Amari *et al.* in [2]. The goal of the stability analysis is to

seek some conditions that guarantee the behavior of the learning system approaching the equilibrium. It has been proven that when the Jacobian $\partial(E[\mathbf{I}-\boldsymbol{\varphi}(\mathbf{y})\mathbf{y}^T]\mathbf{W})/\partial\mathbf{W}$ has negative real parts, the equilibrium is asymptotically stable. For each of the pairwise sources i, j ($i \neq j$), the necessary and sufficient conditions are developed as [2]

$$E\left[y_i^2\dot{\varphi}_i(y_i)\right]+1>0 \tag{32}$$

$$E[\dot{\varphi}_i(y_i)]>0 \tag{33}$$

$$E\left[y_i^2\right]E\left[y_j^2\right]E[\dot{\varphi}_i(y_i)]E[\dot{\varphi}_j(y_j)]-1>0 \tag{34}$$

where $E[\cdot]$ denotes expectation and $\dot{\varphi}=d\varphi/dy$.

Applying these conditions, we investigate the stability effect of the nonlinear functions based on the $\mathbf{t}$-distribution density model and the generalized Gaussian distribution density model.

First, we investigate function (29) derived from the $\mathbf{t}$-distribution density model for super-Gaussian signals. In this case, the conditions in (32)–(34) are always satisfied since the terms on the left-hand side in (35)–(37) are positive (see Appendix I)

$$E\left[y_i^2\dot{\varphi}_i(y_i)\right]+1=\frac{2\beta}{\beta+3}>0 \tag{35}$$

$$E[\dot{\varphi}_i(y_i)]=\frac{\lambda_\beta^2(\beta+1)}{\beta+3}>0 \tag{36}$$

$$E\left[y_i^2\right]E[\dot{\varphi}_i(y_i)]-1=\frac{\beta(\beta+1)\Gamma\left(\frac{\beta-2}{2}\right)}{2(\beta+3)\Gamma\left(\frac{\beta}{2}\right)}-1>0. \tag{37}$$

In (37), the same assumption $\beta > 2$, as in (21), is necessary. Moreover, $E[y_i^2]E[\dot{\varphi}_i(y_i)]$ approaches one if and only if β approaches infinity. This is the case for the Gaussian distributed signal.

Next, we investigate function (30) derived from the generalized Gaussian distribution density model. Similar to the $\mathbf{t}$-distribution model, the stability conditions can be derived as (see Appendix II):

$$E\left[y_i^2\dot{\varphi}_i(y_i)\right]+1=\alpha>0 \tag{38}$$

$$E[\dot{\varphi}_i(y_i)]=\frac{\lambda_\alpha^2\alpha(\alpha-1)\Gamma\left(\frac{\alpha-1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}>0 \tag{39}$$

$$E\left[y_i^2\right]E[\dot{\varphi}_i(y_i)]-1=\frac{\alpha(\alpha-1)\Gamma\left(\frac{3}{\alpha}\right)\Gamma\left(\frac{\alpha-1}{\alpha}\right)}{\Gamma^2\left(\frac{1}{\alpha}\right)}-1>0. \tag{40}$$

Let us consider the following three cases

Case 1: $\alpha > 2$ (For Sub-Gaussian Signals): When $\alpha > 2$, the stability conditions (38)–(40) are always satisfied since the left terms are positive.

Case 2: $\alpha = 2$ (For Gaussian Signals): When $\alpha = 2$, (38) and (39) are positive. In (40), for the Gaussian distributed signal y_i, the term $E[y_i^2]E[\dot{\varphi}_i(y_i)] = 1$. If another signal y_j is also Gaussian distributed, then stability condition (34) is not satisfied since $E[y_i^2]E[\dot{\varphi}_i(y_i)]E[y_j^2]E[\dot{\varphi}_j(y_j)]-1=0$. It is well known that two Gaussian signals cannot be separated. We prove this from the stability analysis point of view. However, if another signal y_j is super-Gaussian or sub-Gaussian, then it is easy to prove that condition (34) is satisfied since $E[y_j^2]E[\dot{\varphi}_j(y_j)] > 1$.

Case 3: $\alpha < 2$ (For Super-Gaussian Signals): When $\alpha < 2$, (38) is positive. However, (39) and (40) do not always have a solution since many discontinuities (for example, $\alpha = 0.2, 0.5$, etc.) exist in the Gamma function $\Gamma((\alpha-1)/\alpha)$, and these discontinuities lead to stability conditions (33) and (34) being unsatisfied in some situations.

As mentioned in the previous subsections, one advantage of employing the $\mathbf{t}$-distribution density model is that the developed nonlinear function is robust to outliers [see Fig. 2(b)]. As we can see here, another advantage of employing the $\mathbf{t}$-distribution model is that it always maintains stability. The cost for employing the $\mathbf{t}$-distribution model is that it introduces an additional shape parameter β.

E. Robust to Misestimation of the Parameter

In Section 3–D, the stability of developed functions was investigated under an ideal case in which the parameter α or β (obtained from the estimate $\hat{\kappa}_i$) is a true value, or is estimated correctly. In this subsection, we investigate the case in which the estimate $\hat{\alpha}$ or $\hat{\beta}$ deviates from the true value α or β with an error ε as

$$\hat{\beta}=\beta+\varepsilon \tag{41}$$

$$\hat{\alpha}=\alpha+\varepsilon. \tag{42}$$

For the case of misestimation, the following questions may arise: 1) Are the stability conditions (32)–(34) still satisfied? 2) If they are conditionally satisfied, what are the new conditions? or How large an interval ε is allowed for the estimate deviating from the true value? In this case, the functions (29) and (30) become

$$\varphi_{i,\hat{\beta}}(y_i)=\frac{(1+\beta+\varepsilon)y_i}{y_i^2+\frac{\beta+\varepsilon}{\lambda_\beta^2}} \tag{43}$$

$$\varphi_{i,\hat{\alpha}}(y_i)=\alpha+\varepsilon\lambda_\alpha\,\mathrm{sgn}(y_i)|\lambda_\alpha y_i|^{\alpha+\varepsilon-1}. \tag{44}$$

Applying stability conditions (32)–(34) to investigate functions (43) and (44), we can derive new conditions as

$$E\left[y_i^2\dot{\varphi}_{i,\hat{\beta}}(y_i)\right]+1=\frac{(\beta+\varepsilon+1)(\beta-3)}{\beta^2+4\beta+3}+1>0 \tag{45}$$

$$E[\dot{\varphi}_{i,\hat{\beta}}(y_i)]=\frac{\lambda_\beta^2(\beta+\varepsilon+1)}{\beta+3}>0 \tag{46}$$

$$E\left[y_i^2\right]E[\dot{\varphi}_{i,\hat{\beta}}(y_i)]-1=\frac{\beta(\beta+\varepsilon+1)\Gamma(\beta/2-1)}{2(\beta+3)\Gamma(\beta/2)}-1>0 \tag{47}$$

and

$$E\left[y_i^2\dot{\varphi}_{i,\hat{\alpha}}(y_i)\right]+1=\frac{((\alpha+\varepsilon)^2-\alpha-\varepsilon)\Gamma[(\alpha+\varepsilon+1)/\alpha]}{\Gamma(1/\alpha)}+1>0 \tag{48}$$

$$E[\dot{\varphi}_{i,\hat{\alpha}}(y_i)]=\frac{\lambda_\alpha^2(\alpha^2+\alpha\varepsilon)\Gamma(2\alpha+\varepsilon-1/\alpha)}{\Gamma(1/\alpha)}>0 \tag{49}$$

$$E\left[y_i^2\right]E[\dot{\varphi}_{i,\hat{\alpha}}(y_i)]-1=\frac{(\alpha^2+\alpha\varepsilon)\Gamma(3/\alpha)\Gamma(2\alpha+\varepsilon-1/\alpha)}{\Gamma^2(1/\alpha)}-1>0 \tag{50}$$

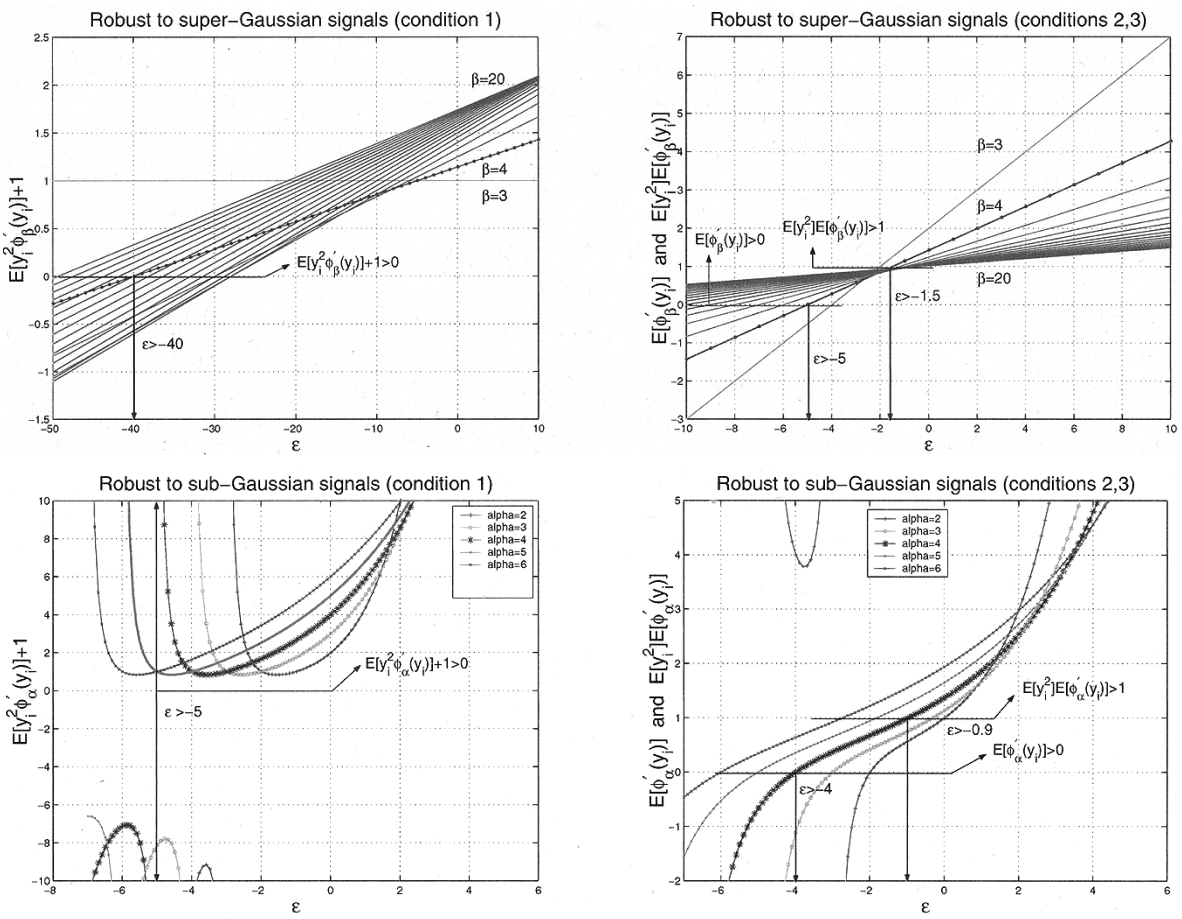

Fig. 3. Robustness to parameter misestimation. (a) and (b) Robustness to misestimating the super-Gaussian signals under the conditions in (45)–(47). (c) and (d) robustness to misestimating the sub-Gaussian signals under the conditions in (48)–(50).

respectively. In (45)–(47) and (48)–(50), we seek the value of ε that will satisfy the new stability conditions.

Let us consider two examples to illustrate these results (see Fig. 3). The first example shows that function (43) derived using the $\mathbf{t}$-distribution density model is robust to misestimation of super-Gaussian signals [see Fig. 3(a) and (b)]. As shown in Fig. 3(a), when the true value $\beta = 4$, the condition (45) is satisfied if $\varepsilon > -40$. Similarly, in Fig. 3(b), when $\beta = 4$, conditions (46) and (47) are satisfied if $\varepsilon > -5$ and $\varepsilon > -1.5$, respectively. In summary, conditions (45)–(47) are satisfied when $\varepsilon > -1.5$, or when the estimate $\hat{\beta}$ is greater than 2.5 (noted that it is not necessary to be exactly equal to four). The second example shows that function (44) derived from the light-tailed distribution density model is robust to misestimation of sub-Gaussian signals [see Fig. 3(c) and (d)]. In this case, the true value $\alpha = 4$, when the estimate $\hat{\alpha}$ is greater than 3.1 (or $\varepsilon > -0.9$); conditions (48)–(50) are always satisfied.

IV. Experimental Results

A. Experiment With Artificially Synthesized Data

We have performed a simulation experiment with one super-Gaussian source, $\kappa_\beta = 4$ and one sub-Gaussian source, $\kappa_\alpha = -1.5$. The total number of data samples was 10 000. We plot 2000 samples in Fig. 4(a). Two sources were artificially mixed using a 7×2 random numeric matrix $\mathbf{A}$. Seven uncorrelated Gaussian additive noise signals with their variances $E[\boldsymbol{\xi}_i^2]$, or

$$\text{diag}(\boldsymbol{\Psi}) = [1.2499 \quad 101.0902 \quad 0.2577 \quad 0.0001 \quad 0.0001 \quad 0.9823 \quad 1.0078]$$

and an outlier with amplitude 100 were added to an associated element of $\mathbf{x}(t)$ [see Fig. 4(b)].

To compare the power of the source to that of the noise and outlier, we define the signal-to-noise ratio (SNR) and the signal-to-outlier ratio (SOR) as

$$\text{SNR} = 10 \log \frac{E[s_i^2]}{E[\xi_i^2]}, \qquad \text{SOR} = \frac{\max(|s_i|)}{\max(|\xi_i|)}.$$

Using these formulas, we know that the maximum SNR $= -20$ dB (high-level noise) at sensor x_2. This means that the power of the noise is 100 times that of the source signal. The amplitude of the outlier (at sensor x_1) is 21 times that of the maximum amplitude of a source, as SOR $= 21$ (an overwhelming transient).

The proposed robust algorithm (see Section II-D) was used for prewhitening with noise reduction. The estimated variances of the additive noise [see (8)] are

$$\text{diag}\left(\hat{\boldsymbol{\Psi}}\right) = [1.2594 \quad 101.0757 \quad 0.2877 \quad 0.1125 \quad 0.0073 \quad 0.9931 \quad 1.0120]$$

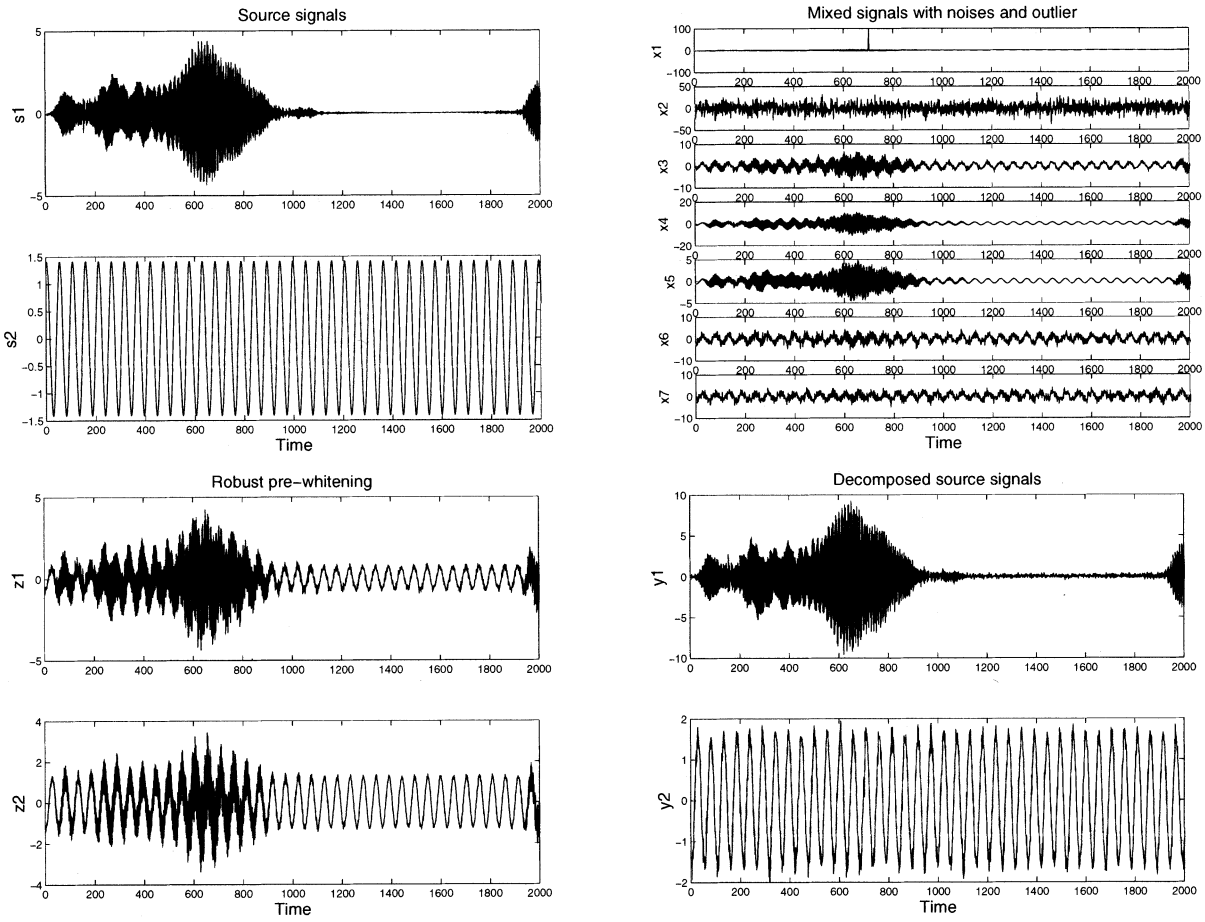

Fig. 4. Results from application of proposed ICA procedures for artificially synthesized signals. (a) Two super- and sub-Gaussian source signals. (b) Seven mixed signals with noise and outlier. (c) Result for robust prewhitening using the proposed algorithm in Section II. (d) Result for decomposing sub- and super-Gaussian source signals using proposed algorithm in Section III.

which are very close to the true values given above. As seen from the result shown in Fig. 4(c), the high-level noise was almost completely removed, but the sources still overlapped. Following this result, the proposed algorithm [see Section III-C; $\eta = 0.005$ for (15) and $\eta = 0.01$ for (31)] was used to further separate the overlapping sub- and super-Gaussian components. The result shown in Fig. 4(d) indicates that the source signals are accurately estimated.

In the previous experiment, the number of sources and its estimate are assumed to be known as $n = \hat{n} = 2$. In the next experiment, we assume the number of sources is unknown, and we apply the proposed criterion (10) to estimate the number of sources. We first divided the data matrix $\mathbf{X}$ into five disjoint groups, such as $\mathbf{X}_i \in \mathbf{R}^{7\times 2000}$ $(i = 1, \ldots, 5)$ Next, we use one of the groups' data to compute one estimate of the noise variance $\mathrm{diag}(\hat{\mathbf{\Psi}}_i)$, and use the remaining data to compute another estimate of the noise variance $\mathrm{diag}(\hat{\mathbf{\Psi}}_j)$ where $j \neq i$. Repeating this calculation, we obtain the error of the estimated number of sources $\mathrm{Error}(\hat{n})$ for $\hat{n} = 1, \ldots, 7$ (see Fig. 5). As seen in Fig. 5, when the estimated number of sources is $\hat{n} = 2$, the error is at a minimum. This result indicates that the dimensionality was reduced optimally.

It should be noted that although we have computed all of the estimates from $\hat{n} = 1$ to $\hat{n} = 7$, it is not necessary when applying the condition $\hat{n} \leq (1/2)(2m + 1 - \sqrt{8m+1})$. Under the condition $\hat{n} \leq 3$, we know the result is the same.

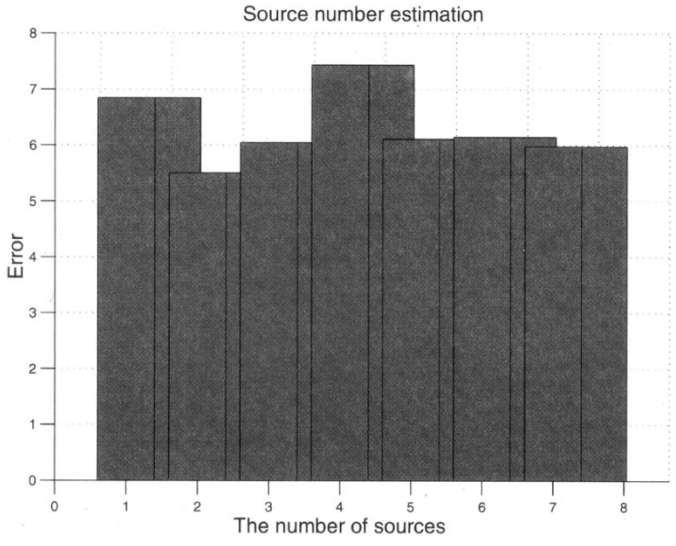

Fig. 5. Result for estimation of the source number. The optimal number is $\hat{n} = 2$.

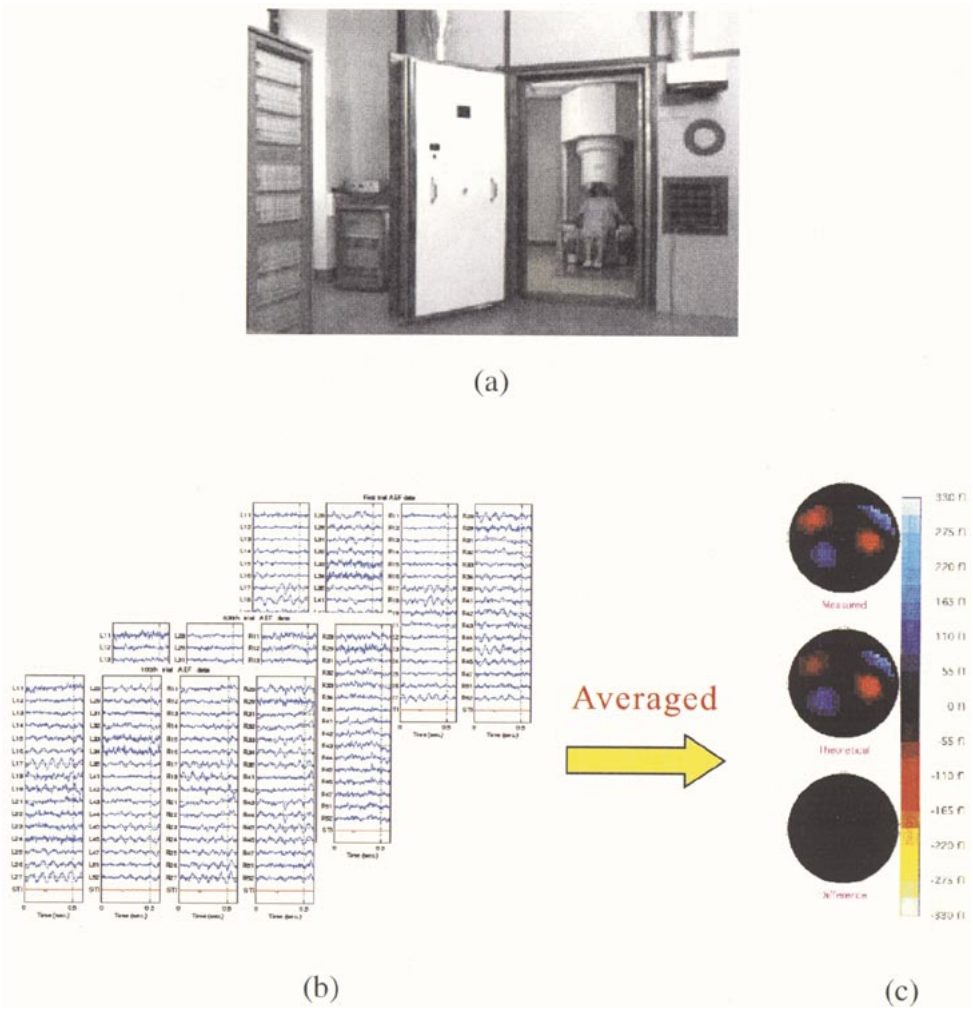

Fig. 6. Experiment, records, and mapping of averaged AEF. (a) Experimental setup for the AEF task. (b) MEG data with 64-channels recorded over 630 trails. (c) Ttypical result for the localization of averaged AEF. Two dipoles appear on the left and right sides of the brain.

B. Experiment With Unaveraged Single-Trial AEF Data

In the application of ICA to real-world MEG/EEG data, most researchers have treated the subjected averaged data to the analysis [35], [38], [29]. We are more interested in analyzing unaveraged single-trial data [12]–[14], [30], since many kinds of important information, such as the strength (amplitude) of an evoked response and its dynamics, can be visualized, which otherwise might be "smoothed" out in averaged trials.

The MEG data for an AEF task were recorded using an Omega-64 (CTF Systems Inc., Canada), whole-cortex MEG system [see Fig. 6(a)]; the experiment was conducted at the National Institute of Bioscience and Human-Technology, Japan. The sensor arrays consist of 64 MEG channels. A healthy, male adult participated in the AEF experiment. Auditory stimulation consisted of a 1 kHz tone, presented binaurally through headphones. There were 630 sets of single-trial data recorded over 379.008 s. The duration of each trial was 0.6016 s, and the stimulus was presented 0.2 s after recording. The sampling rate for the MEG was 312.5 Hz [see Fig. 6(b)].

Taking the average of 630 trials and localizing the evoked fields using the dipole fitting method, we obtain a typical result for AEF analysis, as shown in Fig. 6(c). This result illustrates that the estimated locations are reliable, but the amplitude of each evoked response corresponding to the stimulus is not available. In order to visualize the amplitude of the evoked response, we apply the proposed ICA with the robust prewhitening approach (Sections II and III) to unaveraged single-trial data. For the first single-trial data, two active components (IC1 and IC3), which correspond to N100 evoked responses, are successfully extracted [see Fig. 7(a)]. The IC2 component is a typical 10 Hz alpha-wave, and the high frequency IC4 component may be an environmental interference component.

Projecting the decomposed components IC1, IC2 and IC3 onto the sensor space individually, we can virtually visualize a contribution by a single source at the sensors. This can be done by using $\hat{\mathbf{x}}(k) = \hat{\mathbf{a}}_i \hat{s}_i(k)$, where k is an index of data samples, $\hat{s}_i$ is decomposed source, and $\hat{\mathbf{a}}_i$ is ith column of the estimated matrix $\hat{\mathbf{A}}\mathbf{W}^{-1}$ obtained from Sections II and III, respectively.

Localizing the components IC1, IC2 and IC3 independently using the virtual contribution $\hat{\mathbf{x}}$, we obtain the head maps, as shown in Fig. 7(b)–(d), respectively. The map in Fig. 7(b) indicates that IC1 is located on the right temporal cortex, and the maximum amplitude of IC1 is 184 fT, which corresponds

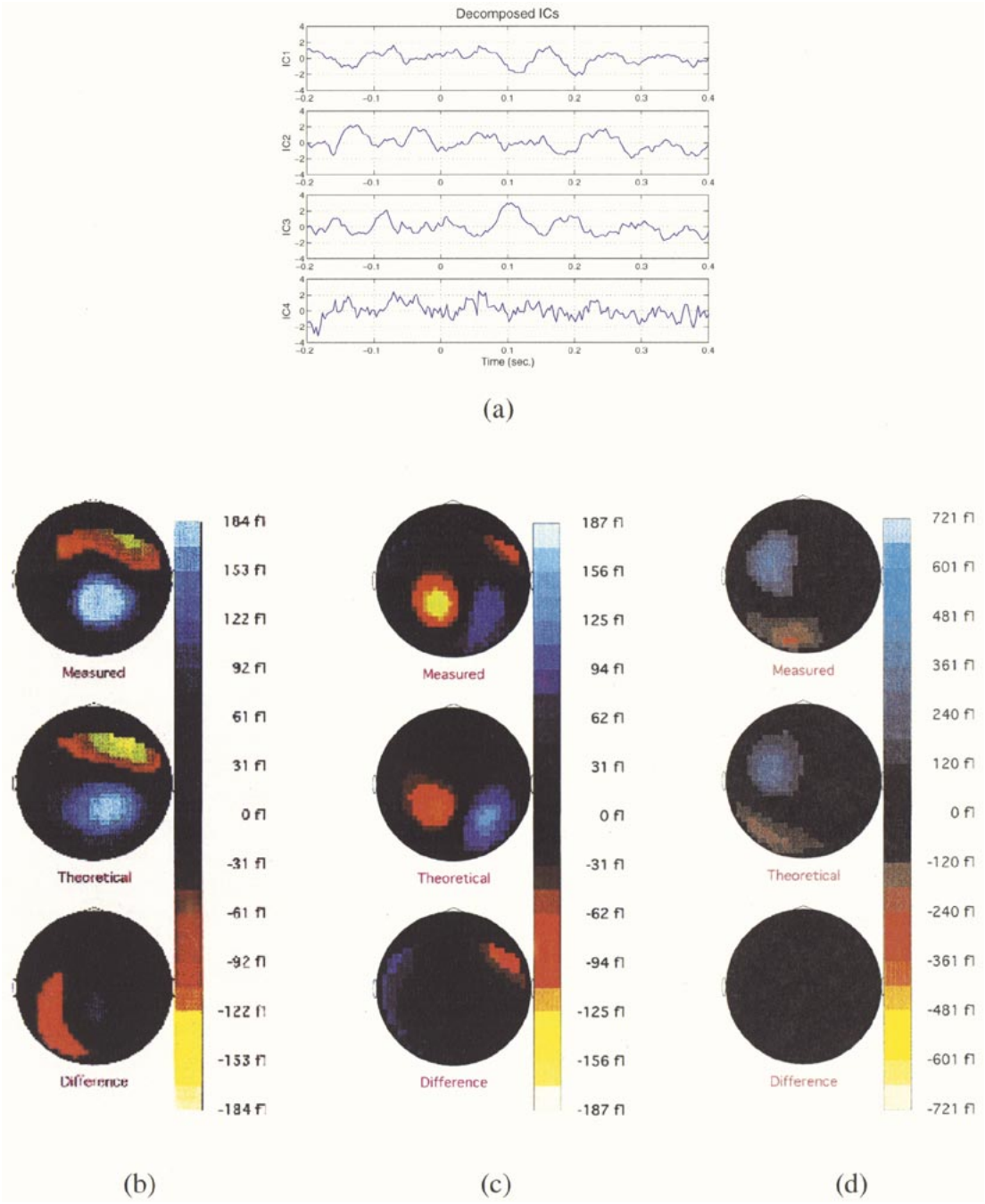

Fig. 7. Results for unaveraged AEF data using the proposed approach. (a) Result for decomposed independent sources. (b) Source localization for IC1 (evoked response N100 on the right side). (c) Source localization for IC2 (a typical alpha-wave component). (d) Source localization for IC3 (evoked response N100 on the left side).

to the first stimulus. The map in Fig. 7(c) indicates that IC2 is a 10 Hz alpha wave and is located near the back of the head. The map in Fig. 7(d) indicates that IC3 is on the left temporal cortex, and the maximum amplitude of IC3 is 721 fT (i.e., very strong). Comparing two maps derived using ICA in Fig. 7(b) and (d) with the averaged map in Fig. 6(c), we find that the two evoked responses, IC1 and IC3, correspond to the averaged one in terms of the source location. It is impossible to obtain amplitude information corresponding to a particular stimulus from the averaged data. However, by applying the proposed approach to unaveraged, single-trial data, the amplitude information (activity strength) of each individual evoked response has been visualized. Moreover, we find that the evoked response on the left temporal cortex, IC3, [see Fig. 7(d)] is much stronger than that on the right side, IC1, [see Fig. 7(b)] when the first stimulus is presented.

The same robust prewhitening technique with JADE algorithm [15] was applied to the same unaveraged single-trial data as above. The result for decomposed individual components is shown in Fig. 8(a). The maps for IC1, IC2, and IC3 are shown in Fig. 8(b)–(d), respectively. Comparing the results derived using JADE with those derived using the natural gradient-based algo-

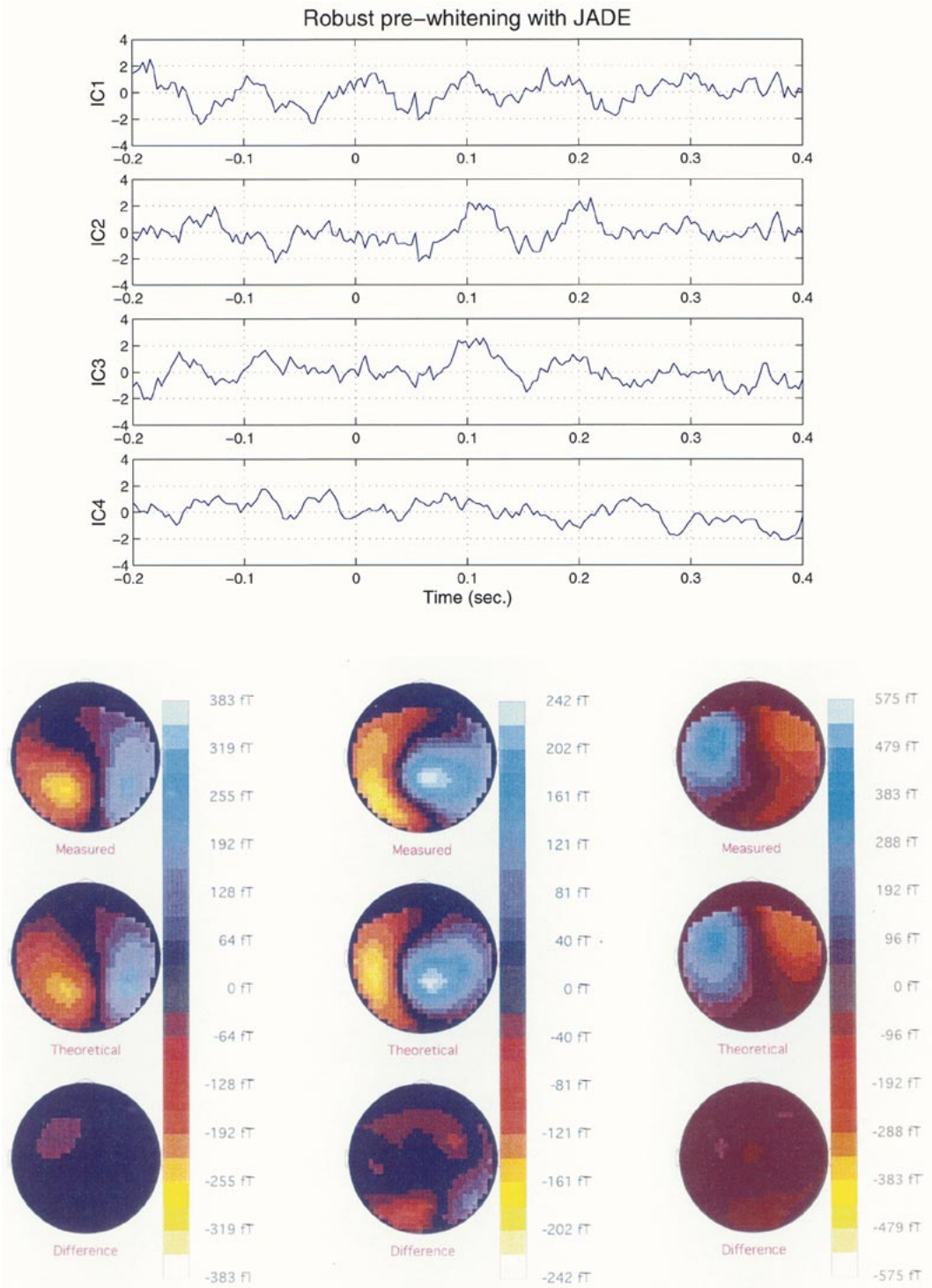

Fig. 8. Results for unaveraged AEF data using JADE with robust prewhitening. (a) Result for decomposed sources. (b) Source localization for IC1 (a typical alpha-wave component). (c) Source localization for IC2. (d) Source localization for IC3. Note that the location of IC1, IC2 and IC3 are different from ones located on the left and right side of the brain shown in Fig. 6(c) or Fig. 7(b) and (d).

rithm, we find that the time courses are very similar in Figs. 7(a) and 8(a). However, the maps for the evoked N100 responses [see Fig. 8(c) and (d); IC2 and IC3] derived using JADE are not identical to the averaged map [Fig. 6(c)] in terms of location. This result proves that the natural gradient-based ICA algorithm works efficiently for real-world measured AEF data.

V. Conclusion

In this paper, we have proposed a novel approach for independent component analysis under the conditions of the sensor signals are contaminated with a high-level additive noise and outliers, overlapped sub-Gaussian and super-Gaussian source components, and the number of sources is unknown.

The main advantages of our approach are as follows: 1) A robust prewhitening technique based on the subspace method is presented for reducing high-level additive noise and for reducing the dimensionality in an optimal manner. It plays the same role in decorrelation as standard PCA, but the noise variance is taken into account. 2) The parameterized unimodal distribution density model, which combines the $\mathbf{t}$-distribution density model with the light-tailed distribution model, is proposed for separating the mixtures of sub-Gaussian and super-Gaussian sources. It has been proven that functions (29) and (30) derived from the proposed model are robust to misestimation of kurtosis. Moreover, function (29) derived by the $\mathbf{t}$-distribution density model is always locally stable (without any discontinuity) and is robust to outliers.

Applying the proposed approach to the analysis of unaveraged single-trial AEF data, we obtained the following novel results: 1) overlapping N100 responses were successfully decomposed into individual responses; 2) the locations of the decomposed N100 responses were identical to the dipoles in the most reliable averaged map; and 3) the strength (amplitude) of each evoked response corresponding to a single stimulus in one single-trial can be visualized. We believe this amplitude information will be very useful for neuroscientists in their studies of the information processing mechanisms of the temporal cortex.

Appendix I
Derivation of the Stability Conditions of $\mathbf{t}$-Distribution Based Function

The differential of function (29) based on the $\mathbf{t}$-distribution density model is derived as

$$\dot{\varphi}_i(y_i) = \frac{(1+\beta)(c-y_i^2)}{(y_i^2+c)^2} \tag{51}$$

where we define $c = \beta/\lambda_\beta^2$.

Let us first calculate the expectation of y_i^2 as

$$\begin{aligned} E[y_i^2] &= \int_{-\infty}^{\infty} y_i^2 p_\beta(y_i)\,dy \\ &= \int_{-\infty}^{\infty} y_i^2 c_\beta \left(1+\frac{y_i^2}{\beta}\right)^{-(1/2)(\beta+1)} dy_i \\ &= 2c_\beta c^{(1/2)(\beta+1)} \left[\frac{1}{c^{q-1}} \sum_{r=1}^{q-1} \frac{(-1)^{r+1}}{2r+1}\, {}_{q-2}C_{r-1} \right. \\ &\qquad \left. \times \left(\frac{y_i}{\sqrt{y_i^2+c}}\right)^{2r+1}\right]_0^{\infty} \end{aligned} \tag{52}$$

where

$$c_\beta = \frac{\lambda_\beta \Gamma\left(\frac{\beta+1}{2}\right)}{\sqrt{\pi\beta}\Gamma\left(\frac{\beta}{2}\right)} \quad \text{and} \quad q = \beta/2$$

we use the formula of the finite series including the binomial coefficients

$$\sum_{r=0}^{n} \frac{(-1)^r}{r+a}\, {}_nC_r \binom{n}{r} = \frac{\Gamma(n+1)\Gamma(a)}{\Gamma(n+a+1)} \tag{53}$$

where ${}_nC_r$ denotes the combination of n things taken r at a time.

Using (53), the result can be obtained as

$$E[y_i^2] = \frac{\beta\Gamma\left(\frac{\beta-2}{2}\right)}{2{\lambda_\beta}^2\Gamma\left(\frac{\beta}{2}\right)}. \tag{54}$$

Note that this result is the same as (21).

Next, using (17) and (51), we can calculate the expectation of $\dot{\varphi}_i(y_i)$ as

$$\begin{aligned} &E[\dot{\varphi}_i(y_i)] \\ &= \int_{-\infty}^{\infty} \dot{\varphi}_i(y_i) p_\beta(y_i)\,dy_i \\ &= 2c_\beta c^{(1/2)(\beta+1)}(\beta+1) \\ &\quad \times \left[\int_0^{\infty} \frac{c}{(y_i^2+c)^{0.5(\beta+5)}}\,dy_i - \int_0^{\infty} \frac{y_i^2}{(y_i^2+c)^{0.5(\beta+5)}}\,dy_i\right] \\ &= 2c_\beta c^{(1/2)(\beta+1)}(\beta+1) \\ &\quad \times \left[\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\beta+4}{2}\right)}{2c^{(\beta/2)+1}\Gamma\left(\frac{\beta+5}{2}\right)} - \frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{\beta+2}{2}\right)}{2c^{(\beta/2)+1}\Gamma\left(\frac{\beta+5}{2}\right)}\right] \\ &= \frac{\lambda_\beta^2(\beta+1)}{\beta+3}. \end{aligned} \tag{55}$$

Similar to (55), the expectation of $\dot{\varphi}_i(y_i)$ with y_i^2 can be calculated as

$$\begin{aligned} &E[y_i^2\dot{\varphi}_i(y_i)] \\ &= \int_{-\infty}^{\infty} y_i^2 \dot{\varphi}_i(y_i) p_\beta(y_i)\,dy_i \\ &= 2c_\beta c^{(1/2)(\beta+1)}(\beta+1) \\ &\quad \times \left[\int_0^{\infty} \frac{cy_i^2}{(y_i^2+c)^{0.5(\beta+5)}}\,dy_i - \int_0^{\infty} \frac{y_i^4}{(y_i^2+c)^{0.5(\beta+5)}}\,dy_i\right] \\ &= 2c_\beta c^{(1/2)(\beta+1)}(\beta+1) \\ &\quad \times \left[\frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{\beta+2}{2}\right)}{2c^{\beta/2}\Gamma\left(\frac{\beta+5}{2}\right)} - \frac{\Gamma\left(\frac{5}{2}\right)\Gamma\left(\frac{\beta}{2}\right)}{2c^{\beta/2}\Gamma\left(\frac{\beta+5}{2}\right)}\right] = \frac{\beta-3}{\beta+3}. \end{aligned} \tag{56}$$

Using the results of (54) and (55), we obtain

$$E[y_i^2]E[\dot{\varphi}_i(y_i)] = \frac{\beta(\beta+1)\Gamma\left(\frac{\beta-2}{2}\right)}{2(\beta+3)\Gamma\left(\frac{\beta}{2}\right)}. \tag{57}$$

Using the results of (55)–(57), we can easily obtain the stability conditions (35)–(37) for the function derived from the $\mathbf{t}$-distribution density model.

Appendix II
Derivation of the Stability Conditions of Generalized Gaussian Distribution–Based Function

The differential of function (30) based on the generalized Gaussian distribution density model is derived as

$$\dot{\varphi}_i(y_i) = \alpha(\alpha-1)\lambda_\alpha^2 |\lambda_\alpha y_i|^{\alpha-2}. \tag{58}$$

Using the definition in (19), we can calculate the expectation of y_i^2 as

$$E[y_i^2] = \int_{-\infty}^{\infty} y_i^2 p_\alpha(y_i)\,dy = \frac{\alpha\lambda_\alpha}{\Gamma\left(\frac{1}{\alpha}\right)} \int_0^{\infty} y_i^2 \exp(-(\lambda_\alpha y_i)^\alpha)\,dy_i = \frac{\alpha\lambda_\alpha}{\Gamma\left(\frac{1}{\alpha}\right)} \times \frac{\Gamma\left(\frac{3}{\alpha}\right)}{\alpha\lambda_\alpha^3} = \frac{\Gamma\left(\frac{3}{\alpha}\right)}{\lambda_\alpha^2\Gamma\left(\frac{1}{\alpha}\right)}. \quad (59)$$

Using (19) and (58), we can calculate the expectation of $\dot{\varphi}_i(y_i)$ as

$$E[\dot{\varphi}_i(y_i)] = \int_{-\infty}^{\infty} \dot{\varphi}_i(y_i) p_\alpha(y_i)\,dy_i = \frac{\lambda_\alpha^3\alpha^2(\alpha-1)}{\Gamma\left(\frac{1}{\alpha}\right)} \int_0^{\infty} (\lambda_\alpha y_i)^{\alpha-2} \exp(-(\lambda_\alpha y_i)^\alpha)\,dy_i = \frac{\lambda_\alpha^{\alpha+1}\alpha^2(\alpha-1)}{\Gamma\left(\frac{1}{\alpha}\right)} \times \frac{\Gamma\left(\frac{\alpha-1}{\alpha}\right)}{\lambda_\alpha^{\alpha-1}\alpha} = \frac{\lambda_\alpha^2\alpha(\alpha-1)\Gamma\left(\frac{\alpha-1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}. \quad (60)$$

Similar to (60), the expectation of $\dot{\varphi}_i(y_i)$ with y_i^2 can be calculated as

$$E[y_i^2\dot{\varphi}_i(y_i)] = \int_{-\infty}^{\infty} y_i^2\dot{\varphi}_i(y_i) p_\alpha(y_i)\,dy_i = \frac{\lambda_\alpha^{\alpha+1}\alpha^2(\alpha-1)}{\Gamma\left(\frac{1}{\alpha}\right)} \int_0^{\infty} y_i^\alpha \exp(-(\lambda_\alpha y_i)^\alpha)\,dy_i = \frac{\lambda_\alpha^{\alpha+1}\alpha^2(\alpha-1)}{\Gamma\left(\frac{1}{\alpha}\right)} \times \frac{\Gamma\left(\frac{\alpha+1}{\alpha}\right)}{\lambda_\alpha^{\alpha+1}\alpha} = \alpha - 1. \quad (61)$$

Using the results of (59) and (60), we obtain

$$E[y_i^2]E[\dot{\varphi}_i(y_i)] = \frac{\alpha(\alpha-1)\Gamma\left(\frac{3}{\alpha}\right)\Gamma\left(\frac{\alpha-1}{\alpha}\right)}{\Gamma^2\left(\frac{1}{\alpha}\right)}. \quad (62)$$

Using the results of (60)–(62), we can easily obtain the stability conditions (38)–(40) for the function derived from the generalized Gaussian distribution density model.

In the same manner, the stability conditions (45)–(47) and (48)–(50) in the case of misestimation of the parameter can be derived.

ACKNOWLEDGMENT

The authors would like to thank H. Endo, the National Institute of Bioscience and Human-Technology, Japan, for the AEF experiment and useful comments.

REFERENCES

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

[2] S. Amari, T. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.

[3] S. Amari and J. F. Cardoso, "Blind source separation: Semiparametric statistical approach," *IEEE Trans. Signal Processing*, vol. 45, pp. 2692–2700, Nov. 1997.

[4] S. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, Feb. 1998.

[5] ——, "Natural gradient for over- and under-complete bases in ICA," *Neural Comput.*, vol. 11, no. 8, pp. 1875–1883, Nov. 1999.

[6] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1984.

[7] H. Attias, "Independent factor analysis," *Neural Comput.*, vol. 11, no. 4, pp. 803–851, May 1999.

[8] M. S. Bartlett, "Factor analysis in psychology as a statistician sees it," *Proc. Uppsala Symp. Psychological Factor Analysis*, no. 3, pp. 4–23, 1953.

[9] A. T. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1004–1034, 1995.

[10] J. Cao, N. Murata, S. Amari, A. Cichocki, and T. Takeda, "MEG data analysis based on ICA approach with pre- and post-processing techniques," in *Proc. 1998 Int. Symp. Nonlinear Theory Applications*, vol. 1, 1998, pp. 287–290.

[11] J. Cao and N. Murata, "A stable and robust ICA algorithm based on $\mathbf{t}$-distribution and generalized Gaussian distribution models," in *Neural Networks Signal Processing*. New York: IEEE Press, 1999, vol. IX, NNSP-99, pp. 283–292.

[12] J. Cao, N. Murata, S. Amari, A. Cichocki, T. Takeda, H. Endo, and N. Harada, "Independent source separation and localization for single-trial Magnetoencephalographic data," in *Proc. 1998 Int. Symp. Nonlinear Theory Applications*, vol. 2, 1999, pp. 719–722.

[13] ——, "Single-trial magnetoencephalographic data decomposition and localization based on independent component analysis approach," *IEICE Trans. Fundamentals of Electronics, Communications, Computer Sciences*, vol. E83-A, no. 9, pp. 1757–1766, Sept. 2000.

[14] J. Cao, N. Murata, S. Amari, A. Cichocki, and T. Takeda, "Independent component analysis for unaveraged single-trial MEG data decomposition and single-dipole source localization," *Neurocomputing*, vol. 49, pp. 255–277, 2002.

[15] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Math. Anal. Appl.*, vol. 17, no. 1, pp. 145–151, 1996.

[16] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.

[17] A. Cichocki and S. Amari, *Adaptive Blind SIgnal and Image Processing*. New York: Wiley, 2002.

[18] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electron. Lett.*, vol. 30, no. 17, pp. 1386–1387, Aug. 1994.

[19] A. Cichocki, I. Sabala, S. Choi, B. Orsier, and R. Szupiluk, "Self adaptive independent component analysis for sub-Gaussian and super-Gaussian mixtures with unknown number of sources and additive noise," in *Proc. 1997 Int. Symp. Nonlinear Theory Applications*, vol. 2, Honolulu, HI, Dec. 1997, pp. 731–734.

[20] A. Cichocki, S. Douglas, and S. Amari, "Robust techniques for independent component analysis (ICA) with noisy data," *Neurocomputing*, vol. 22, pp. 113–129, 1998.

[21] A. Cichocki, J. Karhunen, W. Kasprzak, and R. Vigario, "Neural networks for blind separation with unknown number of sources," *Neurocomputing*, vol. 24, pp. 55–93, 1999.

[22] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[23] S. C. Douglas, A. Cichocki, and S. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions," in *Proc. IEEE Workshop Neural Networks Signal Processing*, Almelia Island Plantation, FL, Sept. 1997, pp. 436–445.

[24] ——, "A bias removal technique for blind source separation with noisy measurements," *Electron. Lett.*, vol. 34, no. 14, pp. 1379–1380, July 1998.

[25] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalized ICA algorithms," in *NIPS'96 Workshop: Blind Signal Processing*, Snowmaas, CO, Dec. 1996.

[26] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, pp. 1483–1492, 1997.

[27] A. Hyvärinen, "Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood," *Neurocomputing*, vol. 22, pp. 49–67, 1998.

[28] ——, "Gaussian moments for noisy independent component analysis," *IEEE Signal Processing Lett.*, vol. 6, pp. 145–147, June 1999.

[29] S. Ikeda, "ICA on noisy data: A factor analysis approach," in *Advances in Independent Component Analysis*, M. Girolami, Ed. New York: Springer-Verlag, 2000.

[30] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Independent component analysis of single-trial event related potentials," in *Proc. ICA'99*, 1999, pp. 173–179.
[31] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.
[32] W. J. Krzanowski, "Cross-validation in principal component analysis," *Biometrics*, vol. 43, pp. 575–584, 1997.
[33] K. G. Jöreskog, "Factor analysis by least squares and maximum likelihood," in *Statistical Methods for Digital Computer*. New York: Wiley, 1976.
[34] 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 Comput.*, vol. 11, no. 2, pp. 417–441, 1998.
[35] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing System 8*. Cambridge, MA: MIT Press, 1996, pp. 145–151.
[36] J. C. Prinecipe, D. Xu, and J. W. Fisher, "Information-theoretic learning," in *Unsupervised Adaptive Filtering: Blind Source Separation*, S. Haykin, Ed., 2000, vol. 1, pp. 265–319.
[37] K. Sharifi and A. Leon Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 52–56, Feb. 1995.
[38] 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, pp. 589–593, May 2000.

Jianting Cao (S'95–M'96) 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 was a Researcher at the Institute of Technology and Equipment with the Ministry of Geological and Mineral in China. From 1996 to 1998, he was a Researcher with the Brain Science Institute, RIKEN (The Institute of Physical and Chemical Research) in Japan. From 1998 to 2002, he was an Associate, and an Assistant Professor with Sophia University, Japan. He is currently an Associate Professor at the Department of Electronic Engineering, Saitama Institute of Technology, and a Visiting Research Scientist at the Brain Science Institute, RIKEN, Saitama, Japan. His research interests include blind signal processing, biomedical signal processing, neural networks, and learning algorithms.

Dr. Cao received the Best Paper Award from the Telecommunications Advancement Foundation, Japan, in 1996. He is a Member of IEICE, Japan.

Noboru Murata received the B.Eng., M.Eng. and Dr.Eng. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1987, 1989, and 1992, respectively.

From 1995 to 1996, he was a Research Associate at the University of Tokyo, and he was a Visiting Researcher with the Research Institute for Computer Architecture and Software Technology of German National Research Center for Information Technology (GMD FIRST), supported by Alexander von Humboldt Foundation. From January 1997 to September 1997, he was a Frontier Researcher with the Brain Information Processing Group in the Frontier Research Program at the Institute of Physical and Chemical Research (RIKEN), Saitama, Japan. In October 1997, he became a Researcher at Laboratory for Information Synthesis in RIKEN Brain Science Institute, and from April 1999 to March 2000, he was a Senior Researcher there. In April 2000, he joined the Department of Electrical, Electronics, and Computer Engineering, Waseda University, Japan, where he is presently an Associate Professor. His research interests include the theoretical aspects of learning machines such as neural networks, focusing on the dynamics and statistical properties of learning.

Shun-ichi Amari (F'94) received the Bachelor degree majoring in mathematical engineering and the Dr.Eng. degree from the University of Tokyo, Tokyo, Japan, in 1958 and 1963, respectively.

He has worked at Kyushu University and then at University of Tokyo, where he is now a Professor Emeritus. He is currently the Director of RIKEN Brain Science Institute and Group Director in Brain-Style Information Processing Group, Brain Science Institute, RIKEN (The Institute of Physical and Chemical Research), Saitama, Japan. He has been engaged in research in wide areas of mathematical engineering or applied mathematics, such as topological network theory, differential geometry of continuum mechanics, pattern recognition, and information sciences. In particular, he has devoted himself to mathematical foundations of neural-network theory, including statistical neurodynamical, dynamical theory of neural fields, associative memory, self-organization, and general learning theory. Another main subject of his research is information geometry proposed by himself, which develops and applies modern differential geometry to statistical inference, information theory, and modern control theory, proposing a new powerful method of information sciences and probability theory.

Dr. Amari is the Past President of the International Neural Networks Society. He was the Conference Chair of the first International Joint Conference on Neural Networks, and he gave the Mahalanobis Lecture on Differential Geometry of Statistical Inference at the 47th session of the International Statistical Institute. He is the recipient of the 1995 Japan Academy Award, the 1992 IEEE Neural Networks Pioneer Award, and the 1997 IEEE Piore Award.

Andrzej Cichocki received the M.Sc. (with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering, from Warsaw University of Technology, Warsaw, Poland, in 1972, 1975, and 1982, respectively.

Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a Full Professor in 1991. He spent a few years as Alexander Humboldt Research Fellow and Guest Professor at the Lehrstuhl fuer Allgemeine und Theortische Elektrotechnikat, University Erlangen-Nuernberg, Germany. He is currently Head of the Laboratory for Advanced Brain Signal Processing in the Brain Science Institute RIKEN, Saitama, Japan, in the Brain-Style Information Processing Group. He is the co-author of three books: *Adaptive Blind Signal and Image Processing* (New York: Wiley, 2002), *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (New York: Springer-Verlag, 1989) and *Neural Networks for Optimization and Signal Processing* (New York: Teubner-Wiley, 1993/1994). His current research interests include biomedical signal and image processing, neural networks, unsupervised learning algorithms and nonlinear dynamic systems theory.

Tsunehiro Takeda was born in 1948. He received the Bachelor, Master, and Doctor degrees from Department of Mathematical Engineering and Information Physics, Tokyo University, Tokyo, Japan, in 1972, 1974, and 1977, respectively.

He was with the Industrial Products Research Institute, MITI, from 1977 to 1997. Since 1997, he has been a Professor of Tokyo University. Presently, 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).

Dr. Takeda received Awards from Agency of Science and Technology in 1991, the Agency of Industrial Science and technology in 1992, Tsukuba City in 1993 for outstanding research, and Agency of Science and Technology for prominent patents in 1988, 1991, and 1994.