FACTOR ANALYSIS PREPROCESSING FOR ICA

Shiro Ikeda

PRESTO, JST
BSI, RIKEN, 2–1 Hirosawa
Wako, Saitama, 351-0198, Japan
shiro@brain.riken.go.jp

ABSTRACT

One of the reasons ICA (Independent Component Analysis) became so popular is that ICA is a promising tools for a lot of applications. One of the attractive applications is the biological data analysis. There are a lot of works on neurobiological data analysis such as EEG (Electroencephalography), fMRI (functional Magnetic Resonance Imaging), and MEG (Magnetoencephalography), and they show interesting results. However, there still remain some problems to be solved. Many neurobiological data includes a large amount of noises, and also the number of independent components is unknown. These problems make it difficult to obtain good results by ICA algorithms. We discuss an approach to separate the data which contain additive noise without knowing the number of independent components. Our approach uses factor analysis as the preprocessing of the ICA algorithm, instead of PCA (Principal Component Analysis), which is the major preprocessing in many ICA algorithms. In the new preprocessing, the number of the sources and the amount of sensor noise are estimated. After the preprocessing, an ICA algorithm is used to estimate the separation matrix and mixing system. Through the experiments with MEG data and fMRI data, we show this approach is effective.

1. INTRODUCTION

The original ICA problem assumes the linear relation between the observed signal $\boldsymbol{x}$ and the source signal $\boldsymbol{s}$ defined as,

$$\boldsymbol{x} = A\boldsymbol{s} \tag{1}$$
$$\boldsymbol{x} \in R^n, \quad \boldsymbol{s} \in R^m, \quad A \in R^{n\times m}.$$

The assumptions in ICA are that each component of $\boldsymbol{s}$ is mean 0, mutually independent and drawn from different probability distribution which is not Gaussian expect for at most one. We also restrict m to be smaller or equal to n for the existence of linear solution.

The goal of ICA is to estimate a separation matrix W which satisfies the following equation,

$$WA = PD \tag{2}$$
$$W \in R^{m\times n}, \quad P \in R^{m\times m}, \quad D \in R^{m\times m},$$

where P is a permutation matrix which has a single entry of 1 in each row and column, and D is representing amplitude ambiguity which is given as a diagonal matrix. With the separation matrix W, we can reconstruct the independent source as $\boldsymbol{y} = W\boldsymbol{x}$. A lot of algorithms are proposed to solve this problem[1, 2, 3, 4, 5]. Most of the algorithms are based on a semi parametric approach[6], and giving a lot of interesting results.

Although each ICA algorithm is simple and its theory is elegant, we sometimes face problems when we apply ICA to real data, especially neurobiological data. The reason is that the relation in (1) does not describe the characteristics of the measurements well. In many cases, we cannot avoid noises, and the number of the sources m is sometimes difficult to be known in advance. For example, in the case of MEG[7, 8], n, which is the number of the sensors, is large (50$\sim$200), but we think the source number m is not so large within a short period. Also we know there are a large amount of noises which are not negligible.

In many neurobiological data, it is natural to categorize noises into two categories. One is the noise which affects every sensor. This noise is called artifacts. Electric power supply, earth magnetism, movements of the head, and the biological signals in which we are not interested (heartbeat, α-wave, and so on), are categorized into artifacts. The other is the noise on each sensor. Since the signals of the brain is usually very weak, we cannot avoid sensor noise through amplifying, or averaging process. This sensor noise is independent to each other.

Therefore, it is better to extend the problem in (1) to a noisy ICA problem which is defined as,

$$\boldsymbol{x} = A\boldsymbol{s} + \boldsymbol{\epsilon} \tag{3}$$

$$\boldsymbol{x}, \boldsymbol{\epsilon} \in R^n, \quad \boldsymbol{s} \in R^m, \quad A \in R^{n\times m}$$
$$\boldsymbol{s}\text{: sources and artifacts,}$$
$$\boldsymbol{\epsilon} \sim N(\mathbf{o}, \Sigma)\text{: noise on each sensor,}$$
$$\Sigma \in R^{n\times n}\text{: diagonal matrix.}$$

In this new problem, we make an assumption that the sensor noise is independent to each other and drawn from different normal distributions. This extension may not be sufficient in some applications, but this is the first step and we believe there are many cases in which it is effective.

In this article, we propose an approach to solve this problem. The idea is to use factor analysis for the preprocessing of data. We show the detail of the approach and some experimental results of neurobiological data analysis.

2. FACTOR ANALYSIS FOR PREPROCESSING OF ICA

2.1. Factor analysis

Many ICA algorithms separate the estimation process into two parts[1, 3, 4]. First part is to estimate the matrix which makes the data uncorrelated. This preprocessing is called sphering or whitening. After sphering, the remaining part of the estimation is a rotation matrix and many algorithms have been proposed[2, 3, 4, 5]. But for the sphering, most algorithms are using PCA[1]. Let $\{\boldsymbol{x}_t\}$ $(t = 1, \ldots, N)$ be the data set, and C be the covariance matrix of $\{\boldsymbol{x}_t\}$, $(C = \sum_t \boldsymbol{x}_t \boldsymbol{x}_t^T/N)$. By defining $P = C^{1/2}$, where $C = PP^T$, observation is transformed as, $\boldsymbol{x}' = P^{-1}\boldsymbol{x}$, and $\boldsymbol{x}'$ is uncorrelated because $\sum \boldsymbol{x}'\boldsymbol{x}'^T/N = I_n$ (I_n is n dimensional identity matrix). This works for noiseless case, but in the noisy case, P does not make the source signals uncorrelated.

Factor analysis is one of the well-known techniques for analyzing noisy multi-variant data. In factor analysis, real valued n-dimensional observation $\boldsymbol{x}$ is modeled as,

$$\boldsymbol{x} = A\boldsymbol{f} + \boldsymbol{\epsilon} \tag{4}$$

$$\boldsymbol{x}, \boldsymbol{\epsilon} \in R^n, \quad \boldsymbol{f} \in R^m, \quad A \in R^{n\times m}$$
$$\boldsymbol{f} \sim N(\mathbf{o}, I_m), \quad \boldsymbol{\epsilon} \sim N(\mathbf{o}, \Sigma),$$
$$\Sigma \in R^{n\times n}\text{: diagonal matrix}$$

The goal of factor analysis is to estimate m, A (factor loading matrix), and Σ (unique variance matrix) using the second order statistics C. The difference of (3) and (4) is the difference of $\boldsymbol{s}$ and $\boldsymbol{f}$. The variable $\boldsymbol{f}$ in (4) is assumed to be normally distributed, but $\boldsymbol{s}$ in (3) is not.

Let $Q \in R^{m\times n}$ be a pseudo-inverse matrix of A in (4), where $Q = (A^t\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}$. It can be easily checked that $AQA = A$ holds, and we transform the data as, $\boldsymbol{z} = Q\boldsymbol{x}$. Then $\boldsymbol{z}$ becomes the sphered data because,

$$\frac{1}{N}\sum \boldsymbol{z}\boldsymbol{z}^T = I_m + Q\Sigma Q^T = I_m + (A^t\Sigma^{-1}A)^{-1}.$$

This result shows that we can make the part of observation $\boldsymbol{x}$ due to the sources uncorrelated, and this is the aim of sphering. Therefore, we can have sphering of noisy observation by factor analysis.

For the estimation of m, A, and Σ, we can use some techniques of factor analysis. There are various estimation methods for A and Σ when m is given. We are going to use MLE (Maximum Likelihood Estimator). MLE is defined as,

$$(\hat{A}, \hat{\Sigma})_{\mathrm{MLE}} = \operatorname*{argmax}_{A,\Sigma}\Big(-\frac{1}{2}\Big\{ \operatorname{tr}\left(C(\Sigma + AA^T)^{-1}\right) + \log(\det(\Sigma + AA^T))\Big\}\Big). \tag{5}$$

For solving the equation, we can use the gradient decent algorithm or Gauss-Newton method. Also the EM (Expectation Maximization) algorithm can be applied.

In order to select the number of factors, m, there are also many approaches and we are going to use the model selection approach with an information criterion, MDL (Minimum Description Length). MDL is defined as follows,

$$\mathrm{MDL} = -L(\hat{A}, \hat{\Sigma}) + \frac{\log N}{N} \times \#\text{ of free parameters.}$$

The number of free parameters in factor analysis model is defined as follows. There are $n(m+1)$ parameters in A and Σ. But A has an ambiguity of rotation and $m(m-1)/2$ is the freedom of this ambiguity. Subtracting $m(m-1)/2$ from $n(m+1)$, the number of free parameters is $n(m+1) - m(m-1)/2$, and MDL for a factor analysis model is defined as,

$$\mathrm{MDL} = -L(\hat{A}, \hat{\Sigma}) + \frac{\log N}{N}\left(n(m+1) - \frac{m(m-1)}{2}\right).$$

For the existence of the estimates, a necessary condition for A to be estimable has been derived[9]. It comes from the fact that $n(n+1)/2 \geq n(m+1) - m(m-1)/2$ has to be satisfied, since C only has $n(n+1)/2$ degrees of freedom. By taking $m < n$ into account, the following bound is obtained

$$m \leq \frac{1}{2}\left\{2n + 1 - \sqrt{8n+1}\right\}. \tag{6}$$

2.2. Factor Analysis and ICA

After the sphering with factor analysis, $\boldsymbol{x}$ is linearly transformed by Q as $\boldsymbol{z} = Q\boldsymbol{x}$. What is left for estimation is the rotation matrix. This is also one big problem of factor analysis. In the last subsection, we assumed that $\boldsymbol{f}$ and $\boldsymbol{\epsilon}$ are normally distributed. We break a part of the assumption. We still assume that $\boldsymbol{\epsilon}$ is normally distributed, but $\boldsymbol{f}$ is not normally distributed and each component is independent. We can use some ICA algorithm now.

The ICA algorithm we use here, should not be affected by the second order statistics since even if data are preprocessed by factor analysis, $\boldsymbol{z}$ still has second order correlations. Therefore, an algorithm based on higher order statistics is preferable here. We use the JADE algorithm by J.-F. Cardoso, which is based on the 4th order cumulant[10]. Suppose a matrix for separation is estimated as $B \in R^{m \times m}$ by JADE. The separated signal $\boldsymbol{y}$ is obtained as,

$$\boldsymbol{y} = B\boldsymbol{z} = BQ\boldsymbol{x} = BQ(A\boldsymbol{s} + \boldsymbol{\epsilon}). \tag{7}$$

The goal of ICA is to estimate BQA to be $PD : P, D \in R^{m \times m}$ as in eq.(2). And finally, we obtain separation matrix W as $W = BQ$.

Also, we can estimate the mixing system by using W. Let us denote A estimated by factor analysis as A_{FA}, and the new mixing system as A_{ICA} as,

$$A_{\mathrm{ICA}} = A_{\mathrm{FA}} B^T. \tag{8}$$

This A_{ICA} does not have rotation ambiguity, and we can estimate the mixing system.

3. NEUROBIOLOGICAL DATA ANALYSIS

In this section, we show the results of the approach applied to neurobiological data.

3.1. MEG data analysis: phantom data

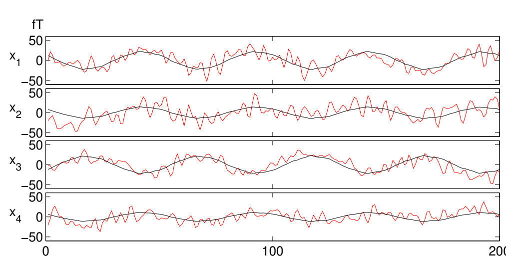

Figure 1: Sensor inputs of phantom data (dotted line), and after removing artifacts (solid line)

First, we show the result of our algorithm applied to MEG phantom data. In the phantom, there is a small platinum electrode and a current signal of 20Hz triangle wave is supplied to the electrode. The strength of the current is designed to be roughly the same as the brain response. The data are averaged over 100 trials. The averaged 4 signals out of 126 active sensors are shown in Fig.1.

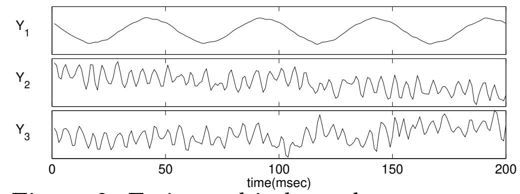

Figure 2: Estimated independent component

The number of the sources and the separation matrix are estimated by our algorithm. In this experiment, the number of the sources is estimated as 3. The separated independent components are shown in Fig.2.

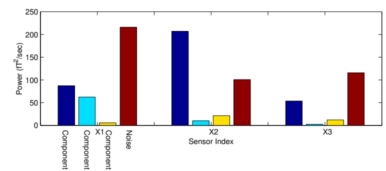

Figure 3: The Power of each component in the sensors

In this experiment, the input to the electrode is y_1 in Fig.2. After selecting the source, we reconstructed the signal using the mixing matrix A_{ICA} in (8). Each column of $\hat{A}_{\mathrm{ICA}}$ corresponds to the coefficients of each independent component to sensors. Let $\hat{A}_{\mathrm{ICA}} = (\hat{\boldsymbol{a}}_{\mathrm{ICA},1}, \ldots, \hat{\boldsymbol{a}}_{\mathrm{ICA},m})$, where $\hat{\boldsymbol{a}}_{\mathrm{ICA},i}$ is an n dimensional column vector. In this case, we can reconstruct the data with, $\boldsymbol{x}'(t) = \hat{\boldsymbol{a}}_{\mathrm{ICA},1} y_1(t)$. The recovered signals are shown in Fig.1. We can see that the noises are reduced.

Figure3 shows estimated power of source, artifacts, and noise on 3 out of 126 sensors. From the figure, we can see that some sensors contain large amount of noise and artifacts than the signal even after averaging over 100 trials.

One of the difficulties in neurobiology is that, we don't know the true signals in the brain. We could show that we can suppress noises in MEG measurement using ICA techniques, but we want to know it more clearly. For the visualization of the result, we implemented SF (spatial filter) technique[11]. SFs are a set of virtual sensors which are located on a hemisphere defined in the brain. We can estimate the current flows on those virtual sensors which describe the MEG observations well. The estimation is obtained by solving an inverse problem and it has a form of a linear mapping from the MEG sensors to SFs. In this experiment, 21×21 SFs are located on a hemisphere and one of the SFs corresponds to the position of the electrode in the phantom .

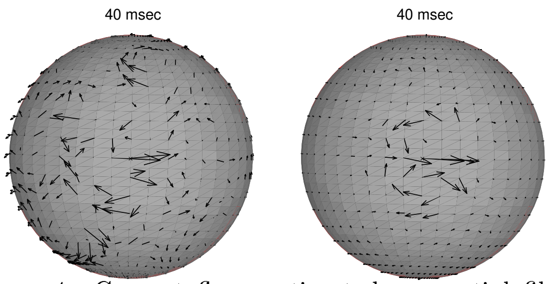

Figure 4: Current flows estimated on spatial filters: Original averaged data (left) and the result of ICA approach (right)

Left side of Fig.4 shows the original averaged data. There are a lot of noises but after the processing with ICA(right side), we clearly see the current flows generated by the electrode.

3.2. MEG data analysis: brain data

We applied our algorithm to the data of brain activity evoked by visual stimulation. The expected results of ICA for MEG data analysis would be summarized as 1) Separating brain signals from artifacts, and 2) Separate brain activities coming from different parts of the brain into different components.

We believe that 1) is possible because the artifacts and the brain signals would be independent. But 2) is difficult since it is more natural to think that even though the signals are coming from different parts of the brain, they might be dependent.

First, we show the averaged data in Fig.5(left). A kind of visual stimulations are given to a subject. The data are recorded by 120 sensors in this case (because only 120 of 129 sensors were working correctly). The duration of recording is from 100msec before and to 412msec after the stimulation with 1kHz sampling rate. The same procedure is applied to the subject for 100 times and we averaged the data. Three of the sensors are shown in Fig.5(left). It is observed that there are still a lot of noises after averaging.

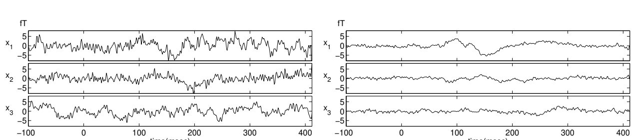

Figure 5: Averaged data(left), and after removing artifacts(right)

We applied our method to the data and 17 independent components were selected by MDL in this experiment. The independent components are shown in Fig.6. We also applied the method to the data from different subjects (4 more), and in all the cases, the selected numbers of sources are roughly the same (from 16 to 19).

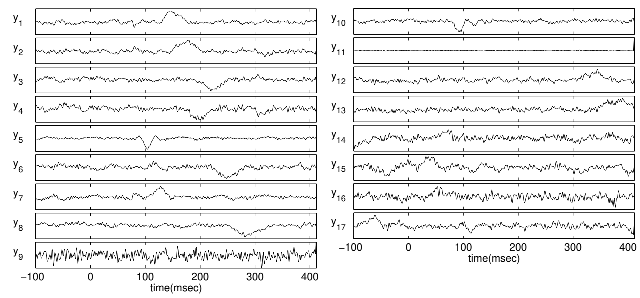

Figure 6: Estimated independent components of visual evoked response

Based on the results in Fig.6, we have to discriminate the brain signals and the artifacts. For example, we can see $y_9(t)$ is mainly a high frequency (180Hz) signal which seems to be an artifact. And $y_{11}(t)$ has a very large value at the very end of the record which seems to be some software noise. But for the other 15 components, it is difficult to know if they are brain sources or not. Fortunately, this experiment is a study of evoked response by visual stimulation, and we are not interested in the components which have some power before the stimulation. Therefore, we defined a threshold such that, if a signal has some power before the stimulation, we regarded the signal as an artifact. In this experiment, $y_1(t)$, $y_2(t)$, $y_3(t)$, $y_5(t)$, $y_6(t)$, $y_8(t)$, and $y_{10}(t)$, are selected as the brain signals.

After picking those sources up, we put them back to the orignal sensor signal space, and the result is shown in Fig.5 (right). It looks like that the noises are removed and the data are clear.

For the visualization, we also used the SF (spatial filter) technique[11]. In this experiment, the part of the brain we are interested in is the visual cortex, and we put 21×21 SFs as the center of the SFs is located at V1.

Fig.7 shows the output of the SFs before and after factor analysis and ICA processing. We recorded the response of a subject from 100msec before the visual stimulation to 412msec after the stimulation. The original data includes a lot of noises even before the stimulation and we can remove them very well. The response of the brain is known to be high around 100∼200msec after the stimulation[11]. From the figure we can see the characteristics of the response is preserved very well.

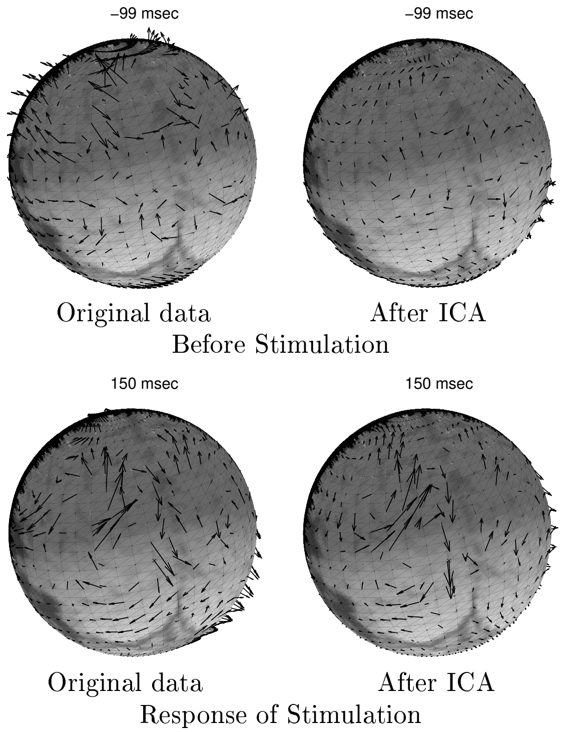

Figure 7: Result of the approach applied to MEG data: outputs of SFs are superimposed on a image of a brain obtained by MRI

3.3. fMRI Data Analysis

ICA is also applied to fMRI data[12], and interesting results have been reported. We also applied the algorithm to fMRI data.

The main difference between MEG data and fMRI data is that, MEG data has more samples in time than the dimension of the sensors, but in fMRI data, usually, the number of samples in time is from 10^1 to 10^3, but the dimension of the observation is huge ($\sim 10^4$, since we take each pixel as a sensor input. In the following fMRI recording, the number of pixels is 21085). Therefore, we usually set index i, t, of $x_i(t)$ in a different way, that is, index i of each component $x_i(t)$ in (1), describes the index of the sample in time, while index t runs through the index of pixels.

This fMRI experiment is designed to study the relationship between grip force and arm movement (they call it, grip force–load force coupling). Subject is forced to keep an object with one's thumb and index finger with/without hand movement and with/without arm movement. Therefore, there are 4 conditions and 6 scans are recorded for each condition after 2 scans for no task period. This set, which consists of $(2+6)\times 4 = 32$scans is repeated for 4 times and we have 128 scans.

Same set of experiment is given to 6 subjects, 6 times for each subject. We averaged the data over this 36sets of experiments using SPM (Statistical Parametric Mapping)[1].

Usual approach is to separate the signals into 128 independent components. But it seems that many of 128 components are rather distributed over wide part of the brain. If we can select smaller number of the independent components through our approach, it will be better for the analysis of brain activity.

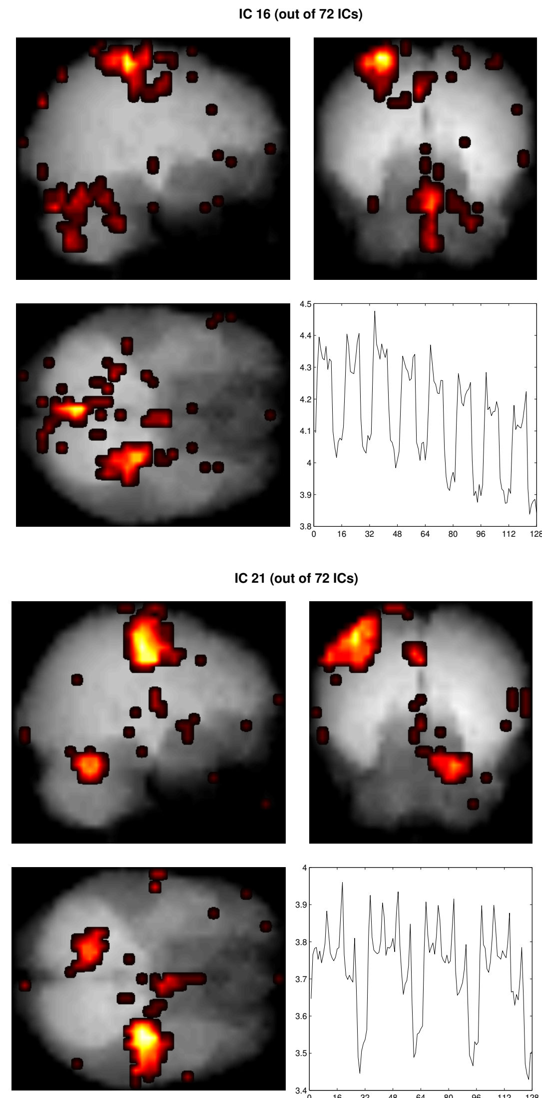

Figure 8: Independent components of fMRI

We applied our algorithm to this data, and finally, we selected 72 independent components. We show 2 of them in Fig.8. There are some meaningless components which are not shown here, and also some artifacts. This

[1] `http://www.fil.ion.ucl.ac.uk/spm/` for further information.

is a preliminary result and we are now studying further analysis.

4. DISCUSSION

In this article, we proposed a new approach of ICA to noisy data. We applied the algorithm to MEG data and fMRI, and have shown the approach is effective. We can estimate the number of the sources, and the power of the noise on each sensor which is independent to each other. This is one of the serious problem which has not been well treated in conventional ICA approaches, and this article gives one effective approach. We have proposed a method which solves this problem based on the semi parametric approach which is one attractive point of ICA.

Our approach also gives a new concept for factor analysis. How to determine the rotation is one of the big problems in factor analysis. And it is not common to use higher order statistics. Therefore our approach gives a new pathway to factor analysis, too.

Although, there still remain a lot of open problems. In the factor analysis, there are a lot of methods to estimate the parameters and the number of the sources, and each of them has each characteristics. We applied MLE for estimation and MDL for estimating the number of the sources. But there are different combinations, and there might be a method which suits better for some particular problems. Also the same thing may be true for the ICA algorithms. We used JADE but there might be a better algorithm. Another problem is the noise distribution. We assumed Gaussian distribution, but if we can have a better model, the algorithm will be improved further.

For neurobiological data, we also should add more biological knowledge to the data analysis. We can separate the signals into independent components, but the meaning of each component is not clear for us, and biological knowledge will help us to understand them.

Acknowledgment

We are grateful to Shun-ichi Amari, Keisuke Toyama, Noboru Murata, and Jianting Cao for their comments and suggestions for this research. We also thank Shigeki Kajihara, Shimadzu Inc. for MEG data and Tomoe Tamada for fMRI data.

5. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

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

[4] J.-F. Cardoso, "Higher-order contrasts for independent component analysis," *Neural Computation*, vol. 11, pp. 157–192, 1999.

[5] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.

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

[7] J. Cao, N. Murata, S. Amari, A. Cichocki, and T. Takeda, "ICA approach with pre & post-processing techniques," in *Proceedings of NOLTA'98*, vol. 1, pp. 287–290, 1998.

[8] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja, "Independent component analysis for identification of artifacts in Magnetoencephalographic recordings," in *Advances in Neural Information Processing Systems*, vol. 10, pp. 229–235, The MIT Press, 1998.

[9] W. Ledermann, "On the rank of the reduced correlational matrix in multiple-factor analysis," *Psychometrika*, vol. 2, pp. 85–93, 1937.

[10] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE-Proceedings-F*, vol. 140, pp. 362–370, 1993.

[11] K. Toyama, K. Yoshikawa, Y. Yoshida, Y. Kondo, S. Tomita, Y. Takanashi, Y. Ejima, and S. Yoshizawa, "A new method for magnetoencephalography: A three dimensional magnetometer-spatial filter system," *Neuroscience*, vol. 91, no. 2, pp. 405–415, 1999.

[12] M. J. Mckeown, T.-P. Jung, S. Makeig, G. Brown, S. S. Kindermann, T.-W. Lee, and T. J. Sejnowski, "Spatially independent activity patterns in functional magnetic resonance imaging data during the Stroop color-naming task," in *Proceedings of the National Academy of Sciences*, vol. 95, pp. 803–810, 1998.