A ROBUST ICA APPROACH FOR UNAVERAGED SINGLE-TRIAL AUDITORY EVOKED FIELDS DATA DECOMPOSITION

Jianting Cao [1,2]*, Noboru Murata*[3]*, Shun-ichi Amari*[2]*, Andrzej Cichocki*[2] *and Tsunehiro Takeda*[4]

[1]Dept. of Electrical and Electronics Engineering, Sophia University,
7-1 Kioicho, Chiyoda-ku, Tokyo 102-8554, Japan.
[2]Brain Science Institute, RIKEN,
2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan.
[3]Dept. of Electrical and Electronics Engineering, Waseda University,
3-4-1 Ookubo, Shinjuku-ku, Tokyo 169-8555, Japan.
[4]Dept. of Complexity of Science and Engineering,
Graduate School of Tokyo University,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
E-mail: cao@brain.riken.go.jp

ABSTRACT

Treating an averaged EFs (evoked-fields) or ERPs (event-related potentials) data is a main approach in recent topics on applying ICA to neurobiological signal processing. By taking the average, the signal-noise ratio (SNR) is increased but some important information such as the strength of an evoked response and its dynamics will be lost. The single-trial data analysis, on the other hand, can avoid this problem but the poor SNR arises.

In this paper, we present a robust approach for decomposing and localizing unaveraged single-trial MEG data. Our approach has two procedures. In the first step, a PCA-like pre-whitening with the high-level noise reduction and an optimal dimensionality reduction techniques are presented. In the second step, a robust nonlinear function derived by the parameterized **t**-distribution model is applied to separate the mixtures of sub-Gaussian and super-Gaussian source components. The results on unaveraged AEFs single-trial data analysis illustrate that not only the behavior and location but also the activity strength (amplitude) and dynamics of the individual evoked response can be visualized by using the proposed method.

1. INTRODUCTION

Applying the ICA to physiological data has received a lot of attentions due to many practical results have been achieved [14], [15], [11], [16]. Recently, the analysis of single-trial data is interested in order for catching more information from human brain records [12]. This paper presents a practical method for decomposing and localizing unaveraged single-trial MEG data [5], [6].

The problem of the single-trial data analysis is formulated by

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

where $\mathbf{x}(t) = [x_1(t), \cdots, x_m(t)]^T$ represent the transpose of m observations at time t. Each observation $x_i(t)$ contains n common components (sources) $\mathbf{s}(t) = [s_1(t), \cdots, s_n(t)]^T$ and a unique component (additive noise) which is included in the vector $\boldsymbol{\xi}(t) = [\xi_1(t), \cdots, \xi_m(t)]^T$. Since the tissue and skull do not attenuate magnetic fields in an MEG measurement, $\mathbf{A} \in \mathbf{R}^{m \times n} = (a_{ij})$ can be represented by a numerical matrix whose element a_{ij} is a quantity related to the physical distance between i-th sensor and j-th source.

In this 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. It is assumed that the components of $\mathbf{s}$ are mutually statistically independent, as well as statistically independent to the noise components $\boldsymbol{\xi}$. Moreover, the noise components $\boldsymbol{\xi}$ themselves are assumed to be mutually independent.

Some remarks about the model :

- Regardless of the nature of components, according to the model, a source component s_i contributes to at least two sensors and a noise component ξ_i contributes at most to only one sensor. Based on this definition, we can easily distinguish a noise component from the source components by checking the distribution of amplitude of the observed signals on the nearby sensors. If the distribution of amplitude of these signals is not smooth, this means nothing but

a noise is added in one of these sensors. This technique sometimes can help us in selecting of a suitable model.

- Depending on the condition of the experiment, the sources $\mathbf{s}$ may include : (1) 'brain sources' such as evoked responses, spontaneous and artifacts; (2) 'interference sources' such as power interference; environmental interferences. These sources contain either a positive kurtosis or a negative kurtosis. Therefore, the problem of separating the mixture of sub-Gaussian and super-Gaussian signals arises.

- The number of sensors m is fixed depending on the MEG machine. The number of sources n is unknown in the model, and it has to be estimated or to be conjectured by using a prior.

There are two kinds of undesirable components have to be removed in our task. The first one is additive noises $\boldsymbol{\xi}$, their power will be reduced by the robust pre-whitening approach in the pre-processing step. The second one, they are usually to be called 'noise' (interference, brain noises etc.) but they are attributed to the source in our model. They will be discarded after ICA source decomposition.

2. METHOD OF DATA ANALYSIS

2.1. Pre-whitening with noise reduction

In this subsection, we first describe the standard PCA approach which has been adopted in some promising ICA algorithms [9], [3] for the pre-whitening. Next, we show that this standard PCA approach can be extended to pre-whitening with a high-level noise reduction.

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

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

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

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T + \boldsymbol{\Psi}, \qquad (3)$$

where $\boldsymbol{\Sigma} = \mathbf{X}\mathbf{X}^T/N$, and $\boldsymbol{\Psi} = \boldsymbol{\Xi}\boldsymbol{\Xi}^T/N$ is a diagonal matrix. For convenience, we assume that $\mathbf{X}$ has been divided by $\sqrt{N}$ so that the covariance matrix can be given by $\mathbf{C} = \mathbf{X}\mathbf{X}^T$.

For the averaged data, the noise variance $\boldsymbol{\Psi}$ is small or zero. A cost function for fitting the model to the data is to make $\mathbf{C} - \mathbf{A}\mathbf{A}^T$ as small as possible. It is well known that the standard PCA can find the principal components by employing the eigen-value decomposition. That is, the solution of $\mathbf{A}\mathbf{A}^T$ for seeking n principal components can be obtained by

$$\hat{\mathbf{A}}\hat{\mathbf{A}}^T = \mathbf{U}_n\boldsymbol{\Lambda}_n\mathbf{U}_n^T, \qquad (4)$$

where $\boldsymbol{\Lambda}_n$ 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 Eq. (4), let one possible solution for $\hat{\mathbf{A}}$ is

$$\hat{\mathbf{A}} = \mathbf{U}_n\boldsymbol{\Lambda}_n^{1/2}, \qquad (5)$$

and then the scores can be obtained from as $\mathbf{z} = \boldsymbol{\Lambda}_n^{-1/2}\mathbf{U}_n^T\mathbf{x}$. Note that the covariance matrix is $E\{\mathbf{z}\mathbf{z}^T\} = \boldsymbol{\Lambda}_n$, it means that $\mathbf{z}$ are orthogonal. Applying this algorithm for the averaged data analysis, some successful results have been reported [14], [15].

For unaveraged single-trial data, the SNR is very low. This means that the diagonal elements of $\boldsymbol{\Psi}$ cannot be ignored in the model. In this case, we can fit $\mathbf{A}\mathbf{A}^T$ to $\mathbf{C} - \boldsymbol{\Psi}$ by the eigen-value decomposition. That is, choosing the columns of $\mathbf{A}$ as eigenvectors of $\mathbf{C} - \boldsymbol{\Psi}$ corresponding to the n largest eigenvalues so that the sum of the squares in each column is identical to the corresponding eigenvalue.

Noted that the noise variance $\boldsymbol{\Psi}$ is assumed to be known. However, in practice $\boldsymbol{\Psi}$ is unknown, and it has to be estimated. Motivated by this, we employ the cost function

$$L(\mathbf{A}, \boldsymbol{\Psi}) = tr[\mathbf{A}\mathbf{A}^T - (\mathbf{C} - \boldsymbol{\Psi})]^2 \qquad (6)$$

and minimize it by $\frac{\partial L(\mathbf{A},\boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} = 0$, whereby the estimate

$$\hat{\boldsymbol{\Psi}} = diag(\mathbf{C} - \hat{\mathbf{A}}\hat{\mathbf{A}}^T) \qquad (7)$$

is obtained. The estimate $\hat{\mathbf{A}}$ can be obtained in the same way as in Eq. (5).

Both the matrix $\mathbf{A}$ and the diagonal elements of $\boldsymbol{\Psi}$ have to be estimated together from data. The estimate $\hat{\mathbf{A}}$ is obtained by the standard PCA. The estimate $\hat{\boldsymbol{\Psi}}$ is obtained by the so called unweighted least squares method that is one of the estimation methods in factor analysis. Once the estimates $\hat{\mathbf{A}}$ and $\hat{\boldsymbol{\Psi}}$ converge to stable values, we can finally compute the score matrix by using the Bartlett method [2] as

$$\mathbf{Q} = [\hat{\mathbf{A}}^T\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{A}}]^{-1}\hat{\mathbf{A}}^T\hat{\boldsymbol{\Psi}}^{-1}. \qquad (8)$$

Using the above result, the new transformation data can be obtained by $\mathbf{z} = \mathbf{Q}\mathbf{x}$. Note that the covariance matrix is $E\{\mathbf{z}\mathbf{z}^T\} = \boldsymbol{\Lambda}_n + \mathbf{C}\boldsymbol{\Psi}\mathbf{C}^T$, which implies that the subspace of the source signals are decorrelated.

The robust approach plays the same role in decorrelation as the standard PCA, but the noise variance $\boldsymbol{\Psi}$ is taken into account. The difference is that the standard PCA is to fit both diagonal and off-diagonal elements of $\mathbf{C}$, whereas

the robust pre-whitening technique is to fit off-diagonals elements of $\mathbf{C}$ only. Based on this property, the robust approach is enable us to reduce a high-level noise which is very important in the single-trial data analysis.

2.2. Optimal dimensionality reduction

The cross-validatory techniques have been wildly applied in multivariate statistics. It usually divides the data into two groups, and uses one group to determine some characteristics of the data, and then uses the other groups to verify the characteristics. Extending this concept, we propose a criterion for determining the estimation of the source number $\hat{n}$ by using the error of estimating the noise variance.

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

$$\mathrm{Error}(\hat{n}) = \frac{1}{K}\sum_{i=1}^{K} \mathrm{tr}[\mathrm{diag}(\hat{\mathbf{\Psi}}_i^{(\hat{n})}) - \mathrm{diag}(\hat{\mathbf{\Psi}}_j^{(\hat{n})})]^2. \quad (9)$$

It should be noted that we are not necessary to compute all of the estimates of the source number such as from $\hat{n} = 1$ to $\hat{n} = m$ when applying a sufficient condition as $\hat{n} \leq \frac{1}{2}(2m + 1 - \sqrt{8m+1})$. Within this bound, we know that the estimate of the source number is reliable.

2.3. Robust nonlinear function in ICA algorithm

After pre-processing the data, a new data vector $\mathbf{z} = \mathbf{Q}\mathbf{x}$ is obtained in which the power of noises, mutual correlation and dimensionality have been reduced. The decomposed independent sources $\mathbf{y} \in \mathbf{R}^n$ then can be obtained from a linear transformation as

$$\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t), \quad (10)$$

where $\mathbf{W} \in \mathbf{R}^{n \times n}$ is the demixing matrix that can be computed by using several BSS/ICA methods [1], [8]. For example, applying the natural gradient based approach, an updating rule is

$$\Delta\mathbf{W}(t) = \eta \left[\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(t))\mathbf{y}^T(t)\right]\mathbf{W}(t), \quad (11)$$

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

$$\varphi_i(y_i) = -\frac{d}{dy_i}\log p_i(y_i) = -\frac{\dot{p}_i(y_i)}{p_i(y_i)}, \quad (12)$$

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

Typical ICA algorithms rely on the choice of nonlinear functions. The form of the optimal function depend on the probability distribution of the source which is usually not available in the ICA task. Several algorithms have been developed for separating the mixtures of sub- and super-Gaussian sources [10], [13]. In this paper, we will use the recently developed the parameterized **t**-distribution model (unimodal) [4]. We will not go into detail in the theoretical analysis but show some advantages for the derived nonlinear function [7] : (1) the function is robust to outliers; (2) the nonlinear function is determined by the value of the kurtosis which corresponds to the source distribution; (3) the algorithm holds stability as well as it robust to the misestimation of kurtosis. The proposed nonlinear functions have the forms :

$$\varphi_i(y_i) = \alpha\lambda_\alpha \,\mathrm{sgn}(y_i)|\lambda_\alpha y_i|^{\alpha-1}, \qquad \hat{\kappa}_i \leq 0, \quad (13)$$

$$\varphi_i(y_i) = \frac{(1+\beta)y_i}{y_i^2 + \frac{\beta}{\lambda_\beta^2}} \qquad \hat{\kappa}_i > 0. \quad (14)$$

Implementation of the proposed ICA algorithm is summarized as [4] :

- Calculate the output $\mathbf{y}$ from given observations $\mathbf{z}$ and an initial value $\mathbf{W}$.
- Calculate the kurtosis by $\hat{\kappa}_i = \hat{m}_4/\hat{m}_2^2 - 3$ where the 2nd- and 4th-order moments are estimated by $\hat{m}_j(t) = [1-\eta(t)]\hat{m}_j(t-1) + \eta(t)y_i^j(t), (j = 2, 4)$.
- Establish the look-up tables by $\kappa_\alpha = \frac{\Gamma(\frac{5}{\alpha})\Gamma(\frac{1}{\alpha})}{\Gamma^2(\frac{3}{\alpha})} - 3$ and $\kappa_\beta = \frac{3\Gamma(\frac{\beta-4}{2})\Gamma(\frac{\beta}{2})}{\Gamma^2(\frac{\beta-2}{2})} - 3$, and determine α or β corresponding to the value of $\hat{\kappa}_i$.
- Calculate the scaling constants $\lambda_\alpha = \left[\frac{\Gamma(\frac{3}{\alpha})}{m_2\Gamma(\frac{1}{\alpha})}\right]^{\frac{1}{2}}$ or $\lambda_\beta = \left[\frac{\beta\Gamma(\frac{\beta-2}{2})}{2m_2\Gamma(\frac{\beta}{2})}\right]^{\frac{1}{2}}$ according to the value of $\hat{\kappa}_i$.
- Calculate the nonlinear function by Eq. (13) or Eq. (14) and update $\mathbf{W}$ by Eq. (11).

2.4. Single-dipole source localization

After pre-processing and applying ICA to the single-trial data, the individual sources have been extracted. We will post-processing the decomposed sources for seeking their location, amplitude and dynamics by using the spatio-temporal dipole fit method. Noted that since we localize the sources independently, it is as reliable as the single equivalent current dipole fitting.

Using the matrices $\mathbf{W}$ and $\mathbf{Q}$ or $\hat{\mathbf{A}}$, we can obtain the estimated behavior of the brain activities as

$$\hat{\mathbf{s}}(t) = \mathbf{y}(t) = \mathbf{G}\mathbf{x}(t), \quad (15)$$

where $\mathbf{G} = \mathbf{WQ}$. To seek the location of the decomposed independent sources, let $\hat{\mathbf{A}}_g = \hat{\mathbf{A}}\mathbf{W}^{-1}$, where $\hat{\mathbf{a}}_i$ is i-th column of matrix $\hat{\mathbf{A}}_g$ and $\hat{s}_i$ is ith component in the estimated sources vector $\hat{\mathbf{s}}$ ($i = 1, \cdots, n$). The virtual contribution of multi-sources or a single source to the sensors can be represented by

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{A}}_g \hat{\mathbf{s}}(k) \qquad \text{or} \qquad \hat{\mathbf{x}}(k) = \hat{\mathbf{a}}_i \hat{s}_i(k), \qquad (16)$$

where k is the sample index of data samples of a single-trial, $\hat{s}_i$ is an independent source which is usually selected from our interest.

3. STUDY ON AUDITORY EVOKED FIELDS

3.1. AEF experiment

The AEF data was recorded using an Omega-64 (CTF systems Inc., Canada) whole-cortex MEG system at the National Institute of Bioscience and Human Technology, Japan. The sensor arrays consists of 64 MEG channels.

The AEF experiment was performed on a normal male adult whose both ears were stimulated by 1 kHz tone. There were 630 sets of trial data recorded in 379.008 sec. The duration of each single-trial was 0.6016 sec and the stimulus was given at 0.2 sec. The sampling rate was 312.5 Hz and the samples were 188 in each trial. In the experiment, the model sphere is set at $x = -0.38$ cm, $y = 0$ cm, $z = 5.35$ cm and $r = 7.3$ cm. Each single-trial with 64-channel records in the time course is shown in Fig. 1.

Taking the average of total recorded trials and localizing the evoked fields by using the dipole fitting method, we obtain averaged map as shown in Fig. 1. This is a typical result of averaged AEF analysis in which the two dipoles appear correctly in the left and right regions of the head map. The latency was set at 0.096 sec in the fitting process. As seen from this map, the maximum evoked response is 330 fT. It only represents the averaged strength of two dipoles.

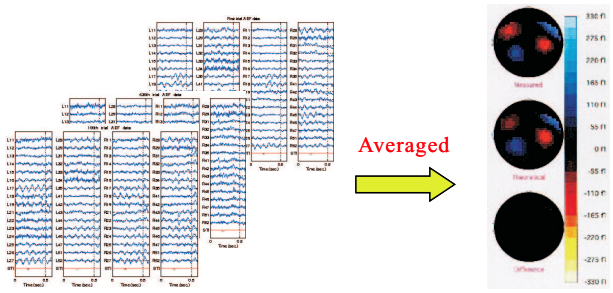


Fig. 1. AEF single-trial records and its averaged map.

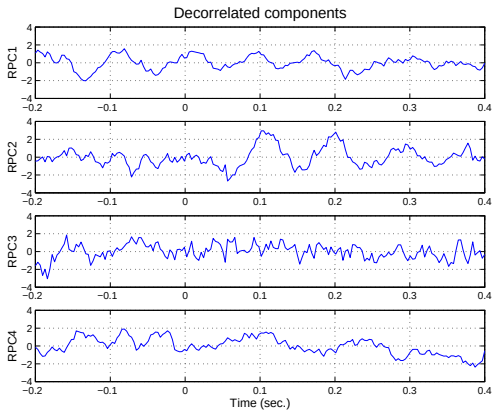
Further beyond the behavior and location of the evoked response as in averaged data, we search the activity strength information of each single evoked response and its dynamics (the strength of the evoked response corresponding to related stimulus). This paper presents results obtained by applying ICA to the unaveraged single-trial data.

3.2. Results for single-trial AEF by ICA

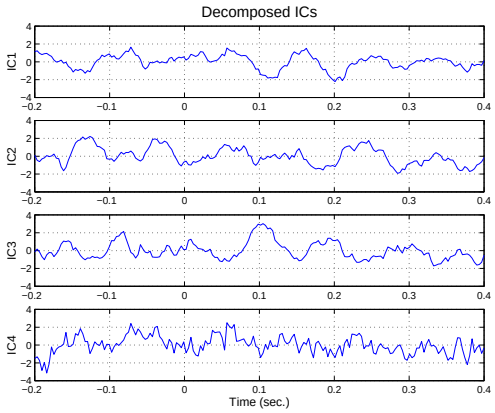
Various unaveraged single-trial data sets have been analyzed by using the robust pre-whitening technique with proposed ICA algorithms. As a typical example, we show the results for the first single-trial data in Figs. 2a and b. As shown in Fig. 2a, the result for the pre-whitening, only one component (RPC2) corresponds to the evoked responses (N100). Hence this result is not identical to the typical AEF analyzed result (Fig. 1) in which there two responses were evoked.

Applying proposed ICA approach, two independent components (IC1 and IC3) are successfully extracted corresponding to the N100 evoked responses (see Fig. 2b). IC2 is a typical alpha-wave component of 11 Hz and IC4 has a high frequency which may has been effected by environment interference. Projecting these components $\hat{s}_i (i = 1, 2, 3)$ onto the sensor space by using Eq. (16) and localizing them independently, the head map for the individual components is obtained as shown in Fig. 2c left, center and right, respectively.

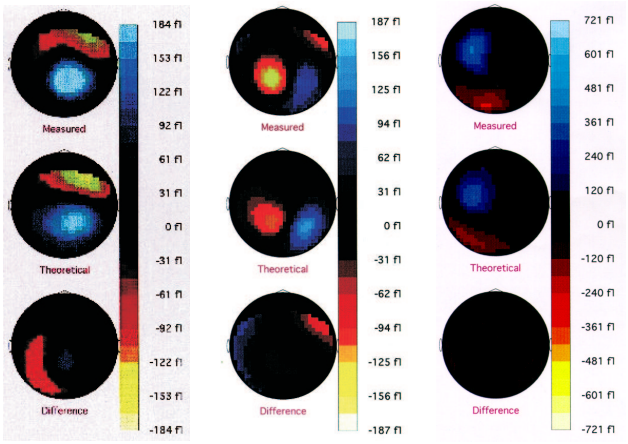
The map in Fig. 2c left shows that the magnetic field distribution for the decomposed IC1 is in the right side of the temporal cortex. The maximum response for IC1 is 184 fT evoked at 110 ms in the first trial. The map of decomposed IC2 in Fig. 2c center shows that an alpha-wave with 11 Hz is located near the back of the head (a typical result). The map in Fig. 2c right shows that the magnetic field distribution for the decomposed IC3 is in the left side of the temporal cortex. The maximum response for IC3 is 721 fT at 101 ms in the first trial.



(a) Result for pre-whitening.



(b) Decomposed ICs for the first trial records.



(c)Source localization of IC1 to IC3.

Fig. 2. Results for 1st. trial AEF data.

Let us focus the discussion on N100 evoked responses by the ICA decomposition and the average of stimulus trials. Comparing the two maps derived by ICA in Figs. 2c with the averaged map in Fig. 1, we can easily find that the two evoked individual responses IC1 and IC3 correspond to the averaged map in their locations. It is impossible to obtain the amplitude information of a single evoked response from the averaged map. However, by using ICA approach for the unaveraged single-trial data, the amplitude information (activity strength) for each individual evoked response has been obtained. Moreover, based on the results in Fig. 2c, we note that the evoked response in the left side of the temporal cortex IC3 is stronger than that in the right side IC1 when given the first stimulus.

In general, it is possible for ICA to separate the different nature of overlapped sources such as the evoked responses, spontaneous, interference and artifacts, since they would be mutually independent. For the same nature sources such as the two evoked responses, it is difficult to presume statistical independence between them. However, based on the results in Figs. 2c, we can find that the maximum amplitude (IC1: 184 fT; IC3: 721 fT) and the corresponding latency (IC1: 110 ms; IC3: 101 ms) of each evoked response are different. This means that the amplitude distributions of these signals are different. This may be a reason for applying ICA to decompose the overlapped same nature responses into two individual responses successfully.

In order to observe the dynamics of the evoked response, we investigated sample data from 1st to 100th trial (100 trials). As an example, we show the results of the decomposed IC1 and IC3 maps corresponding to the 4th stimulus in Fig. 3. Comparing these results with the results shown in Fig. 2c, we have found that the locations of the IC1 and IC3 are almost similar for both the 1st and 4th stimulus. However, the maximum amplitude of IC1 changes as 184 $\rightarrow$ 516 fT and IC3 as 721 $\rightarrow$ 201 fT. This implies that the evoked response IC3 in the left side maybe 'familiar' since the first evoked response is the strongest one in the 100 trials, and the IC1 in the right side maybe 'novel' since the 4th evoked response is the strongest one in 100 trials.

Other phenomena for IC1 and IC3 have been observed in other stimulus trials. The locations are similar for the 1st and 4th trial but their amplitude are relatively weaker. For example, the alpha-wave component IC2 is the strongest in a trial but IC1 and IC3 are relatively weaker, this may be because of the subject lost his attentions.

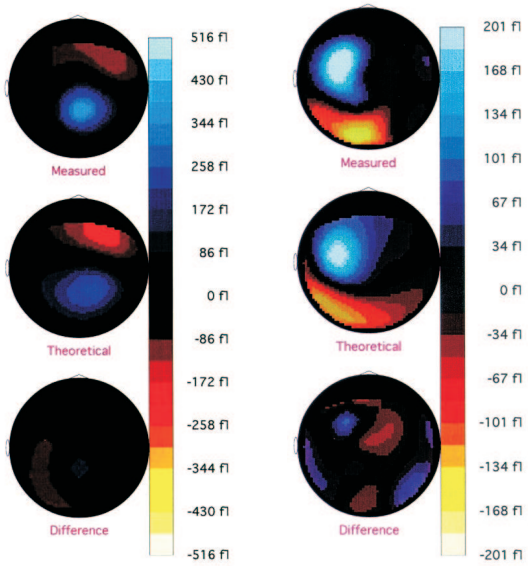


Fig.3. Evoked responses on the 4th. trial.

4. CONCLUSIONS

We have proposed a novel unaveraged single-trial data analysis method and applied it to the MEG AEF tasks. The proposed method is based on the ICA approach with the robust pre-processing and post-processing techniques.

Through the analysis of unaveraged single-trial AEF data by our method, The relationship between the stimulus and the evoked response becomes causal for each singal trial.

Moreover, much important and useful information such as the amplitude, location and dynamics of evoked individual response corresponding to the related stimulus can be obtained.

Based on our data analysis results (without knowledge from neuroscience), some novel features of AEF we believe have been discovered : (1) The activity strength of each evoked response (inside of the brain) is possible to be visualized; (2) The evoked response in the left side of the brain is stronger than the right one for given the first stimulus (maybe depend on the subject), but their locations correspond to those in the averaged map; (3) Related to the evoked responses, some phenomena such as 'familiar', 'novel' and 'lost attentions' have been observed; (4) The evoked responses may not move to other regions in the cortex for the AEF task, they stay almost in the same locations but the strength of the left response alters with the right one. The authors hope that this kind of data analysis procedure can help neuroscientists advance their studies of the mechanism of the temporal cortex.

Acknowledgments

The authors would like to acknowledge Mr. Hiroshi Endo and Dr. Nobuyoshi Harada at the National Institute of Bioscience and Human-Technology, Japan for the AEF experiment and useful comments.

5. REFERENCES

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

[2] M. S. Bartlett, "Factor analysis in psychology as a statistician sees it," *Uppsala Symposium on Psychological Factor Analysis*, no. 3, pp.4-23, 1953.

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

[4] J. Cao and N. Murata, "A stable and robust ICA algorithm based on **t**-distribution and generalized Gaussian distribution models," in *Neural Networks for Signal Processing*, vol. IX, IEEE Press, N.Y., NNSP-99, pp. 283-292, 1999.

[5] 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," *Proc. of 1999 International Symposium on Nonlinear Theory and its Applications*, vol. 2, pp.719-722, 1999.

[6] 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*, 2001 (in print).

[7] J. Cao, N. Murata, S. Amari, A. Cichocki, "A robust approach to independent component analysis with high-level noise measurements," *Submitted to IEEE Trans. Neural Networks*, 2001.

[8] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Signal Processing*, vol. 44, no. 12, pp.3017-3030, 1996.

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

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

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

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

[13] T-W. Lee, M. Girolami and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources," *Neural Computation*, vol. 11, no. 2, pp.417-441, 1998.

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

[15] R. Vig*á*rio, J. S*ä*rel*ä*, V. Jousm*ä*ki, M. H*ä*m*ä*l*ä*inen and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE trans. Biomed. Eng.*, vol. 47, no. 5, pp.589-593, 2000.

[16] A. Ziehe, K. -R. M*ü*ller, G. Nolte, B. -M. Mackert and G. Curio, "Artifact reduction in magnetoneurography based on time-delayed second-order correlations," *IEEE trans. Biomed. Eng.*, vol. 47, no. 1, pp.75-87, 2000.