EEG SIGNAL ENHANCEMENT USING MULTI-CHANNEL WIENER FILTER WITH A SPATIAL CORRELATION PRIOR

Hayato Maki[†]*, Tomoki Toda, Sakriani Sakti, Graham Neubig, Satoshi Nakamura*

Graduate School of Information Science
Nara Institute of Science and Technology (NAIST), Japan
[†]maki.hayato.lt3@is.naist.jp

ABSTRACT

Event-related potentials (ERPs) of electroencephalogram (EEG) are often used as features for brain machine interfaces or for analysis of brain activities. However, as EEG signals easily suffer from various artifacts, ERPs are often collapsed and hard to observe. There are several attempts at using multi-channel EEG signals to enhance EEG signals of interest and make ERPs more clearly observed. For example, a previous work has proposed a blind EEG signal separation method using a multi-channel Wiener filter designed with a probabilistic generative model of observed EEG signals. This method copes with the under-determination of EEG signal separation by assuming sparseness of each EEG component in the time-frequency domain. Although this method blindly separates EEG signals into individual EEG components using time-varying scaled spatial correlation matrices, target EEG components, such as P300 of ERP, are often known in advance in some applications. In this paper, inspired by this previous work, we propose a probabilistic EEG signal enhancement method using a multi-channel Wiener filter, newly incorporating prior information of the spatial correlation matrices related to the target EEG component in the probabilistic generative model to improve performance of EEG signal enhancement. An experimental evaluation for P300 enhancement shows that the proposed method significantly reduces artifacts.

Index Terms— EEG signal enhancement, ERP, Wiener filter, spatial correlation prior

1. INTRODUCTION

Electroencephalography (EEG) is the recording of electrical activity along the scalp. It has been used as a tool for medicine [1], cognitive science [2], and development of new brain machine interfaces (BMI) [3,4,5]. However, it is often hard to analyze and interpret EEG signals because of their poor signal-to-noise ratio (SNR). EEG recordings capture a mixture of endogenous brain activities and extraneous and physiological artifacts such as power grid noise, eye blinks, or muscle activities, while signals of interest evoked by some brain activities generally have lower energy than the artifacts [6]. For the above reasons, synchronous signal averaging is often employed, especially in the study of event-related potentials (ERPs). The whole recording is cut into smaller intervals containing a single stimulus. Each of these intervals is called a trial. Trial signals are averaged synchronously, which attenuates signals from the artifacts and background brain activities while preserving the amplitude of signals of interest. However, it is known that the ERP waveforms have variability between both trials and subjects, so information is lost by averaging [7].

Thus, techniques of artifact removal or feature extraction from a single-trial EEG data have become a highly active research topic in neuroscience, engineering and signal processing [8,9]. One framework for artifact removal using multi-channel EEG signals is Independent Component Analysis (ICA) [10]. ICA defines a generative model for the observed multivariate data, which are assumed to be linear mixtures of some unknown latent variables. The mixing system is also unknown. There are two major assumptions in applying ICA to EEG signals. First, the ICA component projections are summed linearly at scalp electrodes. Second, the time course of EEG activity and artifacts are statically independent. ICA has a limitation of the number of separable signal sources, up to N sources from N electrodes. However, EEG signals are generated by numerous synapses, so the number of electrodes is actually much fewer than sources. This problem is called *under-determined* and it is known to be essentially difficult to solve using linear filtering including ICA.

Another approach uses a multi-channel Wiener filter that has been proposed in the context of under-determined blind source separation [12,13] and applies it to EEG signals [14]. In [14], they define an event as a phenomenon that contributes to an observation signal, e.g. a cognitive process, an eye blink and so on, and assume sparseness of each event in the time-frequency domain. They don't separate EEG signals into individual signals evoked by sources, but separate by each event. This approach has the merit that noise reduction is done by a time-variant filter, while ICA is a time-invariant filter. To design such a filter, they use a probabilistic generative model that models the contributions of each source to all mixture channels in the time-frequency domain as zero-mean complex Gaussian random variables whose covariance matrix encodes the spatial characteristics of the source.

In this paper, inspired by this previous work, we propose a probabilistic EEG signal enhancement method using a multi-channel Wiener filter, newly incorporating prior information of the spatial correlation matrices related to the target EEG component in the probabilistic generative model. The previous work blindly separates EEG signals into individual EEG components using time-varying scaled spatial correlation matrices without any prior information. However, target EEG components, such as P300 of ERP, are often known in advance in some applications. This allows us to obtain prior information for the estimation of parameters and improve performance of EEG signal enhancement. In addition, we don't have to select the target event signal from the separated event signals as in the conventional blind separation methods including ICA, because the target signal corresponds to the mixture component with the target prior distribution.

2. PREVIOUS WORK

Previous works [12,13] have proposed, in the context of under-determined blind sound source separation, a blind signal separation method using a multi-channel Wiener filter designed based on a probabilistic generative model. This method has been successfully applied to unsupervised EEG event signal separation using multi-channel EEG signals [14], which we describe in this section.

2.1. Observation Model

In this paper, we refer to the multi-channel EEG signal related to the k-th event simply as the "k-th event signal." Given I channels, the k-th event signal $\boldsymbol{c}_k(n,f) = [c_{k,1}(n,f), \cdots, c_{k,I}(n,f)]^\top$ at time n and frequency f in the time-frequency domain is expressed as

$$\boldsymbol{c}_k(n,f) = \sum_{l \in E_k} \boldsymbol{h}_l s_l(n,f), \tag{1}$$

$$\boldsymbol{h}_l = [h_{1l}, \cdots, h_{Il}]^\top, \tag{2}$$

where $\{\cdot\}^\top$ is the transpose, $s_l(n,f)$ is the source signal from various sources, such as synapse, muscle, and so on, given by a complex value, E_k is a set of the sources activated in the k-th event, and h_{il} is the transfer function from the l-th source to the i-th channel assuming that $0 \le h_{il} \le 1$. The observed multi-channel EEG signal is $\boldsymbol{x}(n,f) = [x_1(n,f), \cdots, x_I(n,f)]^\top$ expressed as

$$\boldsymbol{x}(n,f) = \sum_{k=1}^{K} \boldsymbol{c}_k(n,f), \tag{3}$$

where K is the number of events.

2.2. Probabilistic Generative Model

We assume that the probability density function of the source signal $s_l(n,f)$ is modeled by the following zero-mean complex Gaussian distribution,

$$p\left(s_l(n,f)\right) = \mathcal{N}_c\left(s_l(n,f); 0, v_k(n,f)\right), \quad l \in E_k, \tag{4}$$

where the variance $v_k(n,f)$ varies in the time-frequency domain depending on the k-th event. Further assuming that the source signals are non-correlated to each other, the probability density function of the k-th event signal $\boldsymbol{c}_k(n,f)$ is modeled by a multivariate complex Gaussian distribution as follows:

$$p\left(\boldsymbol{c}_k(n,f)\right) = \mathcal{N}_c\left(\boldsymbol{c}_k(n,f); \boldsymbol{0}, \boldsymbol{R}_{c_k}(n,f)\right), \tag{5}$$

$$\boldsymbol{R}_{c_k}(n,f) = E\left[\boldsymbol{c}_k(n,f)\boldsymbol{c}_k(n,f)^H\right] \tag{6}$$

$$= v_k(n,f)\boldsymbol{R}_k, \tag{7}$$

$$\boldsymbol{R}_k = \sum_{l \in E_k} \boldsymbol{h}_l \boldsymbol{h}_l^\top, \tag{8}$$

where $\{\cdot\}^H$ is the complex conjugate transpose. The spatial covariance matrix $\boldsymbol{R}_{c_k}(n,f)$ is factorized into the time-frequency invariant spatial covariance matrix $\boldsymbol{R}_k$ and the time-frequency variant variance component $v_k(n,f)$.

We also assume that only one event signal is active in each time-frequency slot as follows:

$$\boldsymbol{x}(n,f) = \boldsymbol{c}_{z(n,f)}(n,f), \tag{9}$$

where $z(n,f)$ is the index of the active event signal. Consequently, the probability density function of the observation signals $\boldsymbol{x}$ in the time-frequency domain is modeled by a Gaussian mixture model as follows:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{n,f} p(\boldsymbol{x}(n,f)|\boldsymbol{\theta})$$

$$= \prod_{n,f}\sum_{k=1}^{K} \alpha_k \mathcal{N}_c\left(\boldsymbol{x}(n,f); \boldsymbol{0}, v_k(n,f)\boldsymbol{R}_k\right), \tag{10}$$

where α_k is a prior probability of the k-th event signal to be active. The model parameter set $\boldsymbol{\theta}$ consists of α_k, $v_k(n,f)$, and $\boldsymbol{R}_k$ of each mixture component.

2.3. Event Separation

2.3.1. Model Parameter Estimation

Given the observation signals $\boldsymbol{x}$, the model parameter set is estimated by maximizing the likelihood function of the generative model given by Eq. (10) as follows:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{x}|\boldsymbol{\theta}). \tag{11}$$

The EM algorithm can be effectively used in this maximization process. In the E-step, the following posterior probability is calculated at each time-frequency slot,

$$m_k(n,f) = \frac{\alpha_k \mathcal{N}_c\left(\boldsymbol{x}(n,f); \boldsymbol{0}, v_k(n,f)\boldsymbol{R}_k\right)}{\sum_{k'=1}^{K} \alpha_{k'} \mathcal{N}_c\left(\boldsymbol{x}(n,f); \boldsymbol{0}, v_{k'}(n,f)\boldsymbol{R}_{k'}\right)} \tag{12}$$

In the M-step, the model parameter set is updated as follows:

$$\hat{\alpha}_k = \frac{\sum_{n,f} m_k(n,f)}{\sum_{n,f,k'} m_{k'}(n,f)}, \tag{13}$$

$$\hat{v}_k(n,f) = \frac{1}{I}\boldsymbol{x}(n,f)^H \boldsymbol{R}_k^{-1}\boldsymbol{x}(n,f), \tag{14}$$

$$\hat{\boldsymbol{R}}_k = \frac{1}{\sum_{n,f} m_k(n,f)} \sum_{n,f} \frac{m_k(n,f)}{\hat{v}_k(n,f)} \boldsymbol{x}(n,f)\boldsymbol{x}(n,f)^H. \tag{15}$$

Note that $\hat{v}_k(n,f)$ and $\hat{\boldsymbol{R}}_k$ are iteratively updated as they depend on each other.

2.3.2. Multi-channel Wiener Filter

Using the estimated spatial covariance matrices related to individual event signals, a multi-channel Wiener filter is designed as follows:

$$\hat{\boldsymbol{c}}_k(n,f) = \hat{\boldsymbol{R}}_{c_k}(n,f)\hat{\boldsymbol{R}}_x^{-1}(n,f)\boldsymbol{x}(n,f), \tag{16}$$

$$\hat{\boldsymbol{R}}_{c_k}(n,f) = m_k(n,f)\hat{v}_k(n,f)\hat{\boldsymbol{R}}_k, \tag{17}$$

$$\hat{\boldsymbol{R}}_x(n,f) = \sum_{k=1}^{K} \hat{\boldsymbol{R}}_{c_k}(n,f), \tag{18}$$

where $\hat{\boldsymbol{c}}_k(n,f)$ is the k-th event signal separated from $\boldsymbol{x}(n,f)$.

2.4. Problem

This method does not use any prior information for estimation of spatial correlation matrices and blindly separates EEG signals into individual EEG components using time-varying scaled spatial correlation matrices. However, target EEG components, such as P300 of ERP, are often known in advance in some applications, which allows us to obtain prior information for the estimation of spatial correlation matrices. In addition, separated signals have permutation ambiguity, so we have to select the target event signal from the separated event signals as in the conventional blind separation methods including ICA.

3. PROPOSED EEG SIGNAL ENHANCEMENT METHOD

In BMI or analysis of brain activities, we often know which event we would like to enhance, such as P300 of ERP, motor imagery, or steady state visual evoked potentials (SSVEP). In such a case, we need to enhance the target EEG event signal from the observed EEG signal, rather than blindly separating it into multiple EEG event signals. Moreover, we can also record EEG signals related to the target event beforehand and use them as prior knowledge for enhancement. In this section, we propose an EEG signal enhancement method that incorporates prior distributions of the spatial covariance matrices to the conventional blind separation framework.

3.1. Spatial Correlation Prior

The Wishart distribution is known as the conjugate prior distribution of the precision matrix of a multivariate Gaussian distribution with known mean vectors. The prior distributions of time-frequency invariant spatial covariance matrices are designed as follows:

$$p\left(\boldsymbol{R}_k^{-1}|\boldsymbol{\Psi}_k^{-1}, q\right) = \frac{1}{Z}|\boldsymbol{R}_k^{-1}|^{\frac{q-I-1}{2}} \exp\left(-\frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\Psi}_k \boldsymbol{R}_k^{-1}\right]\right), \quad (19)$$

where Z is the normalizing constant, $\boldsymbol{\Psi}_k$ is a I-by-I symmetric positive definite matrix, and q is the degrees of freedom.

Using previously recorded multi-channel EEG signals related to a specific event, we can determine the hyper parameters, $\boldsymbol{\Psi}_k$ and q as follows:

$$\boldsymbol{\Psi}_k = \sum_{n,f} \boldsymbol{x}'_k(n,f)\boldsymbol{x}'_k(n,f)^H, \quad (20)$$

$$q = NF, \quad (21)$$

where $\boldsymbol{x}'_k(n,f)$ is the pre-recorded EEG signal at time n and frequency f, N is the total number of time frames, and F is the total number of frequency slots.

It is ideal to use the event signal $\boldsymbol{c}_k$ but it is essentially difficult to record such a signal because the recorded EEG signal easily suffers from multiple events. In this paper, we use carefully recorded EEG signals to establish a contrast between signals in which the target event exists or doesn't exist. For instance, if we apply our proposed method to P300 ERP enhancement, we record several EEG signals in which P300 ERP is supposed to be observed and also record EEG signals in which P300 ERP is supposed to not be observed. Then, we calculate hyper parameters using each of these to develop the prior probability distribution for P300 ERP and that for the background EEG signals, respectively. We may also be able to create prior probability distributions for other specific events, such as an eye blink.

3.2. EEG Signal Enhancement with Spatial Correlation Prior

Given the observation signals $\boldsymbol{x}$, the model parameter set $\boldsymbol{\theta}$ is estimated by maximizing the posterior probability density function of the time-frequency invariant spatial covariance matrices, $\boldsymbol{R} = \{\boldsymbol{R}_1, \cdots, \boldsymbol{R}_K\}$, as follows:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{R}|\boldsymbol{x}, \boldsymbol{\theta}_{\backslash \boldsymbol{R}}, \boldsymbol{\Psi}_k, q) \quad (22)$$

$$= \arg\max_{\boldsymbol{\theta}} \prod_{n,f} p\left(\boldsymbol{x}(n,f)|\boldsymbol{\theta}\right) \prod_{k=1}^{K} p\left(\boldsymbol{R}_k^{-1}|\boldsymbol{\Psi}_k, q\right), \quad (23)$$

where $\boldsymbol{\theta}_{\backslash \boldsymbol{R}}$ is the model parameter set except for $\boldsymbol{R}$. This maximization process can also be effectively solved with EM algorithm. The following auxiliary function is maximized with respect to $\boldsymbol{\theta}$,

$$\begin{aligned}\mathcal{Q} = \sum_{n,f,k} m_k(n,f)\Big(&\log(\alpha_k) - I\log(v_k(n,f)) \\ &+ \log(|\boldsymbol{R}_k^{-1}|) - \frac{1}{v_k(n,f)}\boldsymbol{x}(n,f)^H \boldsymbol{R}_k^{-1}\boldsymbol{x}(n,f)\Big) \\ &+ \sum_{k=1}^{K}\left(\frac{q-I-1}{2}\log|\boldsymbol{R}_k^{-1}| - \frac{1}{2}\mathrm{Tr}\left[\boldsymbol{\Psi}_k \boldsymbol{R}_k^{-1}\right]\right) \\ &+ \mathrm{Const.}\end{aligned} \quad (24)$$

In the E-step, the posterior probability $m_k(n,f)$ is calculated at each time-frequency slot as shown in Eq. (12). In the M-step, $\hat{\alpha}_k$ and $\hat{v}_k(n,f)$ are updated as shown in Eqs. (13) and (14), and $\boldsymbol{R}_k$ is updated as follows:

$$\hat{\boldsymbol{R}}_k = \frac{1}{\frac{q-I-1}{2} + \sum_{n,f} m_k(n,f)}\left(\frac{1}{2}\boldsymbol{\Psi}_k + \sum_{n,f}\frac{m_k(n,f)}{\hat{v}_k(n,f)}\boldsymbol{x}(n,f)\boldsymbol{x}(n,f)^H\right), \quad (25)$$

where $\hat{v}_k(n,f)$ and $\hat{\boldsymbol{R}}_k$ are iteratively updated as they depend on each other. Finally, the target event signal is extracted from the observed EEG signals using a multi-channel Wiener filter as described in **Section 2.3.2**.

In the proposed enhancement method, we don't have to select the target event signal from the separated event signals as in the conventional blind separation method, because the target signal corresponds to the mixture component with the target prior distribution. We may also deal with event signals not modeled with the prior distributions by just using additional mixture components without the prior distributions.

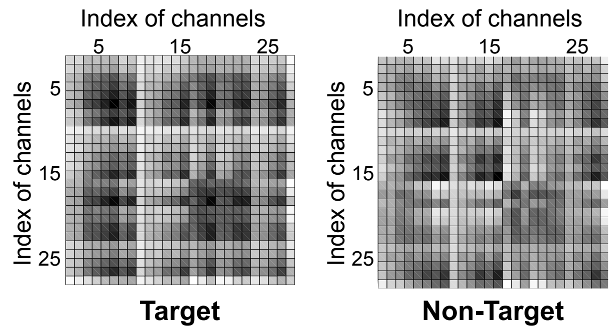

Fig. 1. Visualization of the spatial correlation matrices, the left is calculated using target trials and the right is calculated using non-target trials. Both axes represent channels. Darker slots have stronger correlations.

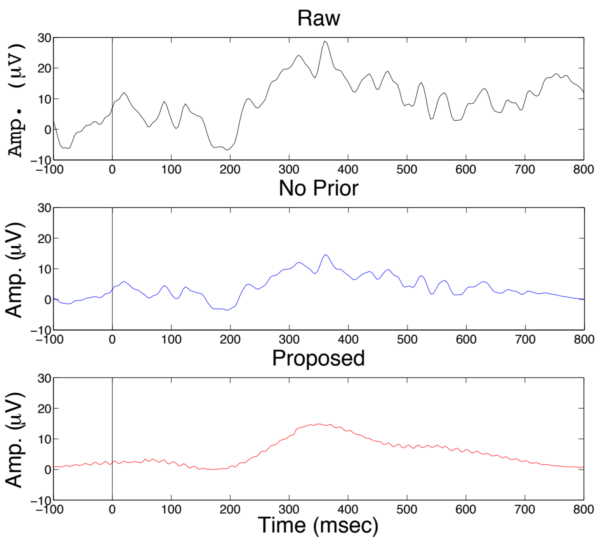

Fig. 2. Comparison of single-trial EEG waveforms. These are, in the order of appearance, the raw signal, the signal denoised by the previous method using no prior, and the signal denoised by the proposed method.

4. EXPERIMENTAL EVALUATION

4.1. Data Acquisition

The EEG data were recorded from a single subject with 27 channels at positions of the extended 10/20-System. The measured signal was digitized at 1000 Hz and downsampled to 200 Hz. We conducted the oddball paradigm experiment, which is a classical experimental design evoking an ERP P300. The subject was presented a sequence of two types of audio stimuli in random order, one is a 2000 Hz sine wave and the other is a 1000 Hz one. 1000 Hz sounds were frequently (200 times) presented and 2000 Hz sounds were rare (50 times). The subject was told that the rare 2000 Hz sound was the target and to count the number of presentations of target stimuli during the experiment. It has been found that P300 usually appears around 300ms after the target stimuli [15]. After the oddball paradigm, the subject was told to be relaxed and we recorded an EEG signal during the resting state for 2 minutes.

Twenty-five EEG signals during target stimuli are chosen at random to use learn hyper parameters and the remaining twenty-five signals are used as test data. The test data are separated using the previous method without prior and the proposed method. Both methods assume the number of events $K = 2$. The learned hyper parameters are visualized in Fig. 1.

4.2. Evaluation

A comparison of waveforms of a single-trial EEG signal in the time domain is shown in Fig. 2. It can be seen that the raw signal is contaminated by artificial spikes. The Wiener filter with no prior does little to remove these spikes, but the proposed method largely succeeds in removing them. Consequently, the appearance of ERP of P300 evoked by target stimuli is more clearly shown.

To perform further quantitative analysis, we assume that the synchronous averaging provides a reasonable approximation for the true target signal, and use the distance from this signal as an automatic evaluation measure. To do so, we averaged 25 target trials of EEG data as a reference signal, and evaluate the performance of the proposed method and previous method with no prior by calculating correlation coefficients between the reference signal and noise reduced signals with only a single or a few trials.

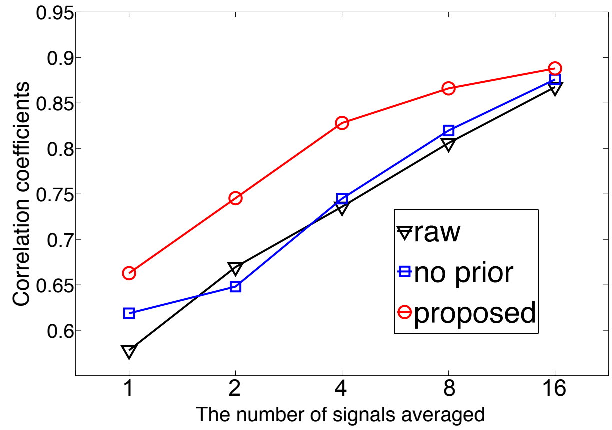

Fig. 3. Mean of correlation coefficients between the reference signal and averaged trial signals. The x-axis that is log-scaled with base 2 represents the number of signals averaged. The y-axis is correlation coefficient

Fig. 3 illustrates the performance of noise reduction combined with synchronous averaging. To calculate the correlation coefficients for each number of signals averaged, we first select 25 random sets with the appropriate number of signals, average the signals in each set, measure the correlation coefficient for each averaged signal, then take the mean of the correlation coefficients for the 25 sets. The performance of noise reduction of proposed method was superior to the previous method and generally needs only half as many trials to achieve a signal of the same quality as previous methods.

5. CONCLUSION

In this study we proposed a method to improve the performance of parameter estimation of a multi-channel Wiener filter for EEG signal noise reduction. While previous works doesn't use a prior information for the estimation of spatial correlation matrices, we used a Wishart distribution as a prior. The experimental evaluation showed the effectiveness of using the prior to estimate the parameters and performance of noise reduction was improved compared to the method without a prior. Future work will consider setting a prior distribution for other parameters and adaptation of hyper parameters learned from other people.

6. ACKNOWLEDGEMENT

Part of this work was supported by JSPS KAKENHI Grant Number 24240032.

7. REFERENCES

[1] Keith H Chiappa. *Evoked potentials in clinical medicine*. Lippincott Williams & Wilkins, 1997.

[2] Wolfgang Klimesch. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, 1999.

[3] Guido Dornhege. *Toward brain-computer interfacing*. MIT press, 2007.

[4] Fabien Lotte, Marco Congedo, Anatole Lécuyer, and Fabrice Lamarche. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of neural engineering*, 2007.

[5] Brent J Lance, Scott E Kerick, Anthony J Ries, Kelvin S Oie, and Kaleb McDowell. Brain-computer interface technologies in the coming decades. *Proceedings of the IEEE*, 2012.

[6] Otavio G Lins, Terence W Picton, Patrick Berg, and Michael Scherg. Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain topography*, 1993.

[7] André Mouraux and Gian Domenico Iannetti. Across-trial averaging of event-related EEG responses and beyond. *Magnetic resonance imaging*, 2008.

[8] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Müller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 2008.

[9] Benjamin Blankertz, Steven Lemm, Matthias Treder, Stefan Haufe, and Klaus-Robert Müller. Single-trial analysis and classification of ERP components - A tutorial. *NeuroImage*, 2011.

[10] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*. John Wiley & Sons, 2004.

[11] Arnaud Delorme, Terrence Sejnowski, and Scott Makeig. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 2007.

[12] Ngoc QK Duong, Emmanuel Vincent, and Rémi Gribonval. Under-determined reverberant audio source separation using a full-rank spatial covariance model. *IEEE Transactions on Audio, Speech, and Language Processing*, 2010.

[13] Ryutaro Sakanashi, Shigeki Miyabe, Takeshi Yamada, and Shoji Makino. Comparison of superimposition and sparse models in blind source separation by multichannel Wiener filter. *Proceedings of the Annual Summit and Conference of the Asia-Pacific Signal and Information Processing Association (APSIPA ASC)*, 2012.

[14] Yusuke Kurihana, Shigeki Miyabe, Tomasz M Rutkowski, Yoshihiro Matsumoto, Takeshi Yamada, and Shoji Makino. (in Japanese) Signal separation of EEG using multivariate probabilistic model. *IEICE Technical Report MBE*, 2013.

[15] Steven J Luck and Emily S Kappenman. *The Oxford handbook of event-related potential components*. Oxford university press, 2011.