ADAPTIVE RHYTHMIC COMPONENT EXTRACTION WITH REGULARIZATION FOR EEG DATA ANALYSIS

Yuki Saito, Toshihisa Tanaka, and Hiroshi Higashi

Department of Electrical and Electronic Engineering
Tokyo University of Agriculture and Technology (TUAT)
2–24–16, Nakacho, Koganei-shi, Tokyo 184–8588, Japan
Emails: tanakat@cc.tuat.ac.jp, saito@sip.tuat.ac.jp, higashi@sip.tuat.ac.jp

ABSTRACT

Rhythmic component extraction (RCE) is a method for extracting a signal oscillating at a certain frequency from multi-channel sensor signals. This method can be effectively used for detecting rhythmic signals such as alpha and beta waves, which are the feature signals in brain computer/machine interfaces (BCI/BMI). We are addressing a problem in developing an on-line adaptive algorithm for RCE. Since a rhythmic signal in the brain slowly varies, the signals extracted in adjacent frames should not largely different. We propose introducing a regularization term that evaluates the correlation between the signal extracted in the last step and the one to be extracted to achieve this. We show that the maximization of the cost function with the proposed regularization term is reduced to a generalized eigenvalue problem and experimental results from practical EEG data support this analysis.

Index Terms— Electroencephalogram (EEG), signal extraction, brain computer interface, multi-channel signal processing

1. INTRODUCTION

It is crucial to extract the brain activity of humans from measured brain signals in a brain-computer/machine interface (BCI/BMI) and clinical diagnosis. To observe brain activity, non-invasive measurement devices, such as electroencephalogram (EEG), magnetroencephalography (MEG), and functional magnetic response imaging (fMRI), are widely used. Among these, because of its simplicity and low cost, EEG is the most practical measurement device for engineering applications. In general, signals measured by EEG have high resolution times, but low spatial resolution. In addition, the amplifier gain is very large and then the obtained signal is highly affected by measurement noise. Therefore, signal processing techniques need to be developed to effectively extract the desired features from the measurements [1].

In BCI, which is a challenging signal processing and neuroscience application, rhythmically oscillating components, such as mu rhythm and beta wave, and/or event related potential, such as P300, are widely used as feature values. To extract these features, methods based on frequency analysis, such as linear filtering and Fourier analysis, and/or ones based on statistical signal processing, such as independent component analysis (ICA) [2, 3], have been used [1]. Moreover, the so-called common spatial pattern (CSP) [4], which is a method based on learning, is also known as a method for extracting features in BCI applications.

Frequency filtering is a classical and simple single channel processing method for extracting a specific frequency component of interest. However, when we measure data in a noisy environment[1], it is difficult to differentiate a component generated by brain activity from noise-related components. In addition, a filter spanning a narrow frequency range like a notch filter, can extract an alpha wave-like signal from random noise. ICA is a multi-channel signal processing method that exploits the statistical nature of signals. ICA is theoretically well-established and there have been several reports where ICA can be used to extract rhythmic signals like alpha waves [2, 3]. However, in practical settings, a component of interest cannot always be obtained by using ICA, because there is no guarantee that the desired component meets the assumption of ICA, that is, the original sources are statistically independent and the number of originals is equal to the number of observations. Also, the CSP needs to learn the data, which is strongly dependent on a subject, a task, and the conditions of the experiment.

Rhythmic component extraction (RCE) [5] was developed to extract a rhythmically oscillating signal, Like other methods for multichannel signal processing, RCE extracts the target signal by using a linear combination of the observed signals with weight coefficients. This method uses physically well-established information, that is, frequency. However, unlike frequency filtering, this method does not discard or filter out frequency components, but actually "enhances" frequency components of interest by using a simple linear combination of the channel signals. This way, the RCE successfully extracts a signal that has energy mostly in the frequency of interest. Moreover, RCE is independent of the subjects or methods of learning.

However, Tanaka and Saito [5] have developed the only batch-type algorithm for RCE. To analyze EEG signals or to develop BCI systems, we need to adapt the weight coefficients. This adaptation should be made so that there is no *discontinuity* in the weight coefficients or the extracted signal over time. For instance, when we attempt to extract a component with frequency of 10 Hz, if the estimated weight coefficients change a lot in sequence, the frequency of the extracted signal is no longer the desired 10 Hz. Therefore, the objective of this paper is to develop an adaptive RCE algorithm that avoids the discontinuity effect. Therefore, we propose to regularize the cost function of the weight coefficients originally proposed for RCE in [5] with the correlation between rhythmic components extracted in the previous and current updates. Then this regularized cost function is maximized by solving a generalized eigenvalue

Tanaka is also affiliated with RIKEN Brain Science Institute, Saitama, Japan. This work was supported in part by Tateisi Science and Technology Foundation, 2008.

[1] Even slight moves of a human head lead to significant noise.

problem. In the experiments, the proposed adaptive algorithm is successfully applied to extract a rhythmic signal from practical EEG data measured in an unshielded environment.

2. RHYTHMIC COMPONENT EXTRACTION (RCE) METHOD

Rhythmic component extraction (RCE) is a method for extracting a component that concentrates its energy within a certain frequency range by using a weighted sum of the channel signals. This section reviews the theory of RCE proposed in [5]. Let $x_i[k]$ ($k = 0, \ldots, N-1$) be an observed signal in the ith channel, where $i = 1, \ldots, M$. We extract a signal by using a linear combination of the channel signals as follows:

$$\hat{x}[k] = \sum_{i=1}^{M} w_i x_i[k], \tag{1}$$

where w_i, $i = 1, \ldots, M$, is a weight coefficient to be determined by a certain criterion. The RCE determines the weight coefficients in such a way that the energy in specific frequency components of $\hat{x}[k]$ is as large as possible while the energy in the other frequency $\hat{x}[k]$ is as small as possible. This idea is formulated in the following way.

Let $\hat{X}(e^{-j\omega})$ be the discrete-time Fourier transform (DTFT) of $\hat{x}[k]$, that is, $\hat{X}(e^{-j\omega}) = \sum_{k=0}^{N-1} \hat{x}[k]e^{-j\omega k}$, and let $\Omega_1 \subset [0, \pi]$ and $\Omega_2 \subset [0, \pi]$ be the frequency ranges of interest and those to be suppressed, respectively. It is sufficient to use positive frequencies because the EEG signal is real-valued. Then, the RCE cost function to be maximized is given as follows [5]:

$$J_1[w_1, \ldots, w_M] = \frac{\int_{\Omega_1} |\hat{X}(e^{-j\omega})|^2 d\omega}{\int_{\Omega_2} |\hat{X}(e^{-j\omega})|^2 d\omega}. \tag{2}$$

The maximization of the above cost function is reduced to a generalized eigenvalue problem in the following way. Define $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ as $[\boldsymbol{X}]_{ik} = x_i[k]$, and matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ as

$$[\boldsymbol{W}_1]_{l,m} = \Re \int_{\Omega_1} e^{-j\omega(l-m)} d\omega, \tag{3}$$

$$[\boldsymbol{W}_2]_{l,m} = \Re \int_{\Omega_2} e^{-j\omega(l-m)} d\omega, \tag{4}$$

respectively, where $l, m = 0, \ldots, N-1$ and $\Re\cdot$ takes the real-part of the complex value. Then, $J_1[\boldsymbol{w}]$ in (2) can be described in the matrix-vector form as

$$J_1[\boldsymbol{w}] = \frac{\boldsymbol{w}^T \boldsymbol{X} \boldsymbol{W}_1 \boldsymbol{X}^T \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{X} \boldsymbol{W}_2 \boldsymbol{X}^T \boldsymbol{w}}, \tag{5}$$

where $\boldsymbol{w} = [w_1, \cdots, w_M]^T$ ($\cdot^T$ describes the transpose). The maximizer of $J_1[\boldsymbol{w}]$ is given by the eigenvector corresponding to the maximum eigenvalue of the following generalized eigenvalue problem:

$$\boldsymbol{X} \boldsymbol{W}_1 \boldsymbol{X}^T \boldsymbol{w} = \lambda \boldsymbol{X} \boldsymbol{W}_2 \boldsymbol{X}^T \boldsymbol{w}. \tag{6}$$

The problem can be solved by using a matrix square root of $\boldsymbol{X}\boldsymbol{W}_2\boldsymbol{X}^T$. Since $\boldsymbol{X}\boldsymbol{W}_2\boldsymbol{X}^T$ is symmetric, a matrix square root, $\boldsymbol{S}$, exists such that $\boldsymbol{X}\boldsymbol{W}_2\boldsymbol{X}^T = \boldsymbol{S}\boldsymbol{S}^T$. Note that $\boldsymbol{S}$ is not uniquely determined. Then, the optimal solution, $\boldsymbol{w}^*$, is given by

$$\boldsymbol{w}^* = \boldsymbol{S}^{-T} \hat{\boldsymbol{w}}, \tag{7}$$

where $\hat{\boldsymbol{w}}$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{W}_1\boldsymbol{X}^T\boldsymbol{S}^{-T}$, where $\cdot^{-T} = (\cdot^{-1})^T$.

3. ADAPTATION AND REGULARIZATION OF RCE

We discuss how to adapt RCE, and illustrate that a *discontinuity* problem arises by introducing the adaptation. Then, it is suggested that a solution to this problem is to consider the correlation between the extracted signals in successive two frames.

3.1. Adaptation

Adaptation of RCE can be made by using frame processing. Let $x_i[n]$ be the observed signal in the ith channel with time index n. In this case, the time index can be either finite or infinite. In a way similar to (1), a rhythmic component is extracted as follows:

$$\hat{x}[n] = \sum_{i=1}^{M} w_i^{(n)} x_i[n], \tag{8}$$

where $w_i^{(n)}$ is the weight for the nth sample. The weight, $w_i^{(n)}$, is obtained by replacing sample matrix $\boldsymbol{X}$ in (5) with matrix $\boldsymbol{X}^{(n)}$ defined by

$$\left[\boldsymbol{X}^{(n)}\right]_{ik} = a[k]x_i[n+k-d], \tag{9}$$

where $a[k]$, $k = 0, \ldots, N-1$, is an appropriate window function with a length of N and d is a time delay. For instance, when $d = (N-1)/2$ where N is odd, the time index n corresponds to the center of the frame. In summary, we obtain the weight at time index n as

$$\boldsymbol{w}^{(n)} = \arg\max_{\boldsymbol{w}} J_1[\boldsymbol{w}]\Big|_{\boldsymbol{X}=\boldsymbol{X}^{(n)}}. \tag{10}$$

Moreover, $J_1[\boldsymbol{w}^{(n)}]$ represents a variation of the energy of the rhythmic component in the frequency of interest, Ω_1. This implies that $J_1[\boldsymbol{w}^{(n)}]$ can lead to a time-frequency analysis.

3.2. Regularization for Adaptive RCE

In the adaptive RCE introduced above, the weights obtained in the previous and current frames are independently derived. Therefore, it is highly possible that the rhythmic components extracted in the $n-1$th frame and in the nth frame significantly differ from each other. However, the waveform and/or phase of a rhythmic brain signal that is generated from the synchronization of electrical activity of neurons cannot suddenly change[2]. Therefore, in this subsection, we propose a regularization method that takes into consideration the correlation between the signals extracted in the previous frame and to be extracted in the current frame.

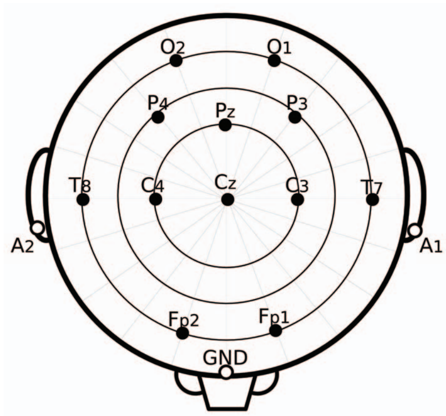
Let $\boldsymbol{w}^{(n-1)}$ and $\boldsymbol{w}^{(n)}$ be the weight coefficients in the $n-1$th and the nth frames, respectively. Notice that $\boldsymbol{w}^{(n-1)}$ has already been obtained and $\boldsymbol{w}^{(n)}$ is an unknown vector to be determined. Then, the correlation of rhythmic components extracted in the previous and current frames is described as

$$r^{(n)} = \boldsymbol{w}^{(n)T} \boldsymbol{X}^{(n)} \boldsymbol{P}_1 \boldsymbol{P}_0{}^T \boldsymbol{X}^{(n-1)T} \boldsymbol{w}^{(n-1)}, \tag{11}$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_0$ are the matrices of size $l \times (l-1)$ that take the part overlapping between $\boldsymbol{X}^{(n)}$ and $\boldsymbol{X}^{(n-1)}$. In the case of a one-sample shift of frames, with zero matrix $\boldsymbol{0}_{1\times(l-1)}$, they are described as

$$\boldsymbol{P}_0 = \begin{bmatrix} \boldsymbol{I}_{l-1} \\ \boldsymbol{0}_{1\times(l-1)} \end{bmatrix}, \boldsymbol{P}_1 = \begin{bmatrix} \boldsymbol{0}_{1\times(l-1)} \\ \boldsymbol{I}_{l-1} \end{bmatrix}. \tag{12}$$

[2] When sampling EEG signal with rate 1kHz, the time interval between time indexes $n-1$ and n is 1/1000 seconds.



Ch No. i	Electrode
1	Cz
2	Pz
3	C4
4	T8
5	C3
6	T7
7	O2
8	O1
9	Fp2
10	Fp1
11	P4
12	P3

Fig. 1. Location of electrodes

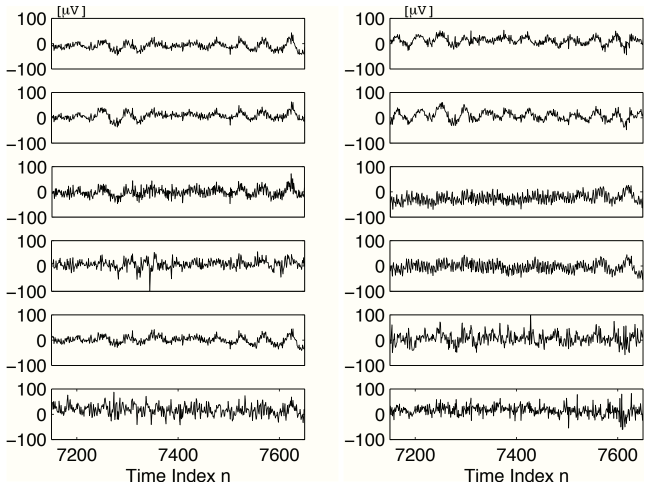
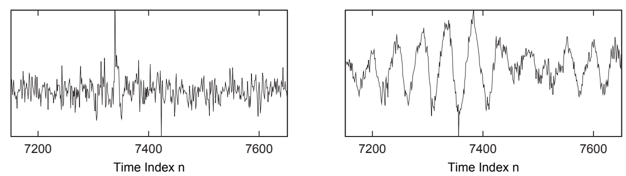


Fig. 2. Raw observed EEG signals. $7150 \leq n \leq 7650$. From top to bottom, left: Cz, Pz, C4, T8, C3, T7; right: O2, O1, Fp2, Fp1, P4, P3.



(a) ICA (b) Proposed method

Fig. 3. Extracted signals using Fast ICA and adaptive RCE from channel signals shown in Fig. 2.

Under the assumption that a rhythmic signal from a brain varies slowly with time, we should estimate the $\boldsymbol{w}$ in such a way that $|r^{(n)}|$ remains large while adapting. Therefore, we propose to regularize the cost function with $|r^{(n)}|^2$, and find the $\boldsymbol{w}$ that maximizes the new cost function given as

$$J_2[\boldsymbol{w}] = \frac{\boldsymbol{w}^T \boldsymbol{X}^{(n)} \boldsymbol{W}_1 \boldsymbol{X}^{(n)^T} \boldsymbol{w} + \epsilon |r^{(n)}|^2}{\boldsymbol{w}^T \boldsymbol{X}^{(n)} \boldsymbol{W}_2 \boldsymbol{X}^{(n)^T} \boldsymbol{w}}, \tag{13}$$

that is, $\boldsymbol{w}^{(n)}$ is given by

$$\boldsymbol{w}^{(n)} = \arg\max_{\boldsymbol{w}} J_2[\boldsymbol{w}], \tag{14}$$

where ϵ is a regularization coefficient. When $\epsilon = 0$, $J_2[\boldsymbol{w}]$ coincides with $J_1[\boldsymbol{w}]$.

It can be shown that the maximization of $J_2[\boldsymbol{w}]$ can also be reduced to a generalized eigenvalue problem. By defining rank-1 matrix

$$\boldsymbol{C}^{(n-1)} = \boldsymbol{P}_1 \boldsymbol{P}_0{}^T \boldsymbol{X}^{(n-1)^T} \boldsymbol{w}^{(n-1)} \boldsymbol{w}^{(n-1)^T} \boldsymbol{X}^{(n-1)} \boldsymbol{P}_0 \boldsymbol{P}_1{}^T, \tag{15}$$

$|r^{(n)}|^2$ can be rewritten as

$$|r^{(n)}|^2 = \boldsymbol{w}^{(n)^T} \boldsymbol{X}^{(n)} \boldsymbol{C}^{(n-1)} \boldsymbol{X}^{(n)^T} \boldsymbol{w}. \tag{16}$$

Therefore, we obtain

$$J_2[\boldsymbol{w}] = \frac{\boldsymbol{w}^T \boldsymbol{X}^{(n)} (\boldsymbol{W}_1 + \epsilon \boldsymbol{C}^{(n-1)}) \boldsymbol{X}^{(n)^T} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{X}^{(n)} \boldsymbol{W}_2 \boldsymbol{X}^{(n)^T} \boldsymbol{w}}. \tag{17}$$

In the way similar to that for $J_1[\boldsymbol{w}]$, the maximizer of this cost function, $\boldsymbol{w}^{(n)}$, is given as

$$\boldsymbol{w}^{(n)} = \boldsymbol{S}^{(n)^{-T}} \hat{\boldsymbol{w}}^{(n)}, \tag{18}$$

where $\boldsymbol{X}^{(n)} \boldsymbol{W}_2 \boldsymbol{X}^{(n)^T} = \boldsymbol{S}^{(n)} \boldsymbol{S}^{(n)^T}$, and $\hat{\boldsymbol{w}}^{(n)}$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{S}^{(n)^{-1}} \boldsymbol{X}^{(n+1)} (\boldsymbol{W}_1 + \epsilon \boldsymbol{C}^{(n-1)}) \boldsymbol{X}^{(n+1)^T} \boldsymbol{S}^{(n)^{-T}}$.

4. EXPERIMENTAL RESULTS

Our experiments proved that the proposed adaptive RCE with regularization works well in extracting a rhythmic signal. We measured a set of EEG signals by using a multi-channel bio-amplifier, (MEG-6116) by NIHON KOHDEN, with an A/D converter (AIO-163202F-PE) by Contec, in a normal unshielded room. The following are the experimental settings and measured data information:

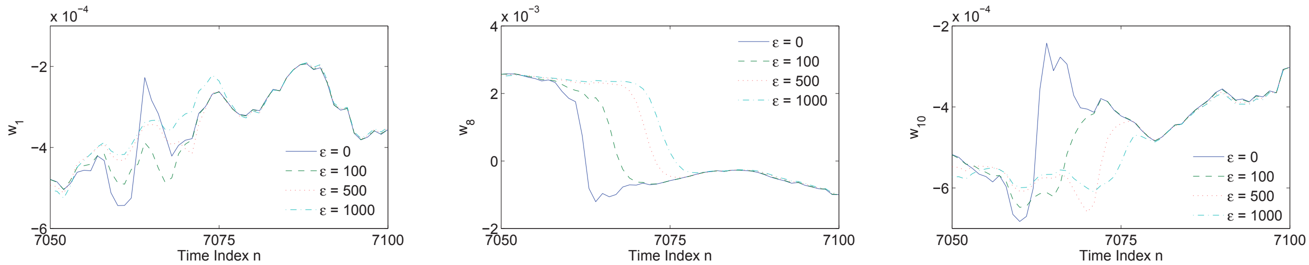
- Sampling frequency: $F_s = 500$ Hz
- Number of channels: $M = 12$ (Cz, Pz, C4, T8, C3, T7, O2, O1, Fp2, Fp1, P4, P3)
- Electrode location: See Fig. 1.
- Task of subject: Eye opened (0 to 10 sec.), eye closed (10 to 20 sec.), and eye opened (20 to 30 sec.).

A part of the observed signals for all the channels are illustrated in Fig. 2. It shows that the task of closing one's eyes yields a visible alpha wave that appears for some channels, but many of them are strongly contaminated by noise. Moreover, the amplitude range of the signals varies.

For simplicity, we extracted the alpha wave-type signals, and then set Ω_1 to $[7, 13](2\pi/F_s)$ and Ω_2 to $[0, \pi] - \Omega_1$. In addition, the frame size, N, was set to 512, and the frame window was shifted by one sample.

4.1. Adaptive RCE vs Fast ICA

The first example shows the comparison of the proposed adaptive RCE with the well-known Fast ICA [2]. Remember that the adaptive RCE algorithm is given by (8) and (18). For comparison, the Fast ICA is also applied in an adaptive manner. In other words, the weight coefficients obtained in the nth frame are used for extracting the nth



(a) Cz (i = 1)

(b) O1 (i = 8)

(c) Fp1 (i = 10)

Fig. 5. Variation of $w_i^{(n)}$, $7050 \leq n \leq 7100$.

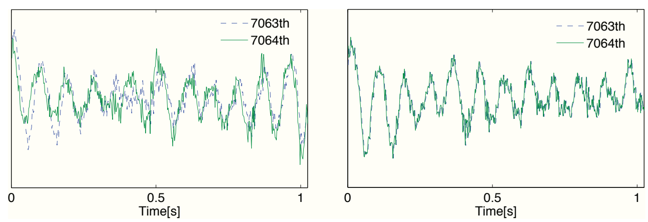


Fig. 4. Results from using $\boldsymbol{w}^{(n-1)}$ and $\boldsymbol{w}^{(n)}$ for two successive frames (the $n-1$th and nth frames). Left: Without regularization ($\epsilon = 0$). Right: With regularization ($\epsilon = 1000$).

sample of the independent component in the same way as in (8). For the RCE, the regularization coefficient was fixed to $\epsilon = 1000$ in this comparison.

Figure 3 shows the signals for a time index from 7150 to 7650 extracted by using both adaptive RCE and Fast ICA. It is clear from the figure that the ICA fails to extract the dominant signal that we can see in the raw data, and it is difficult to give a physical meaning to the extracted component. However, the proposed RCE algorithm successfully extracts an alpha-like wave with less noise than the observed data illustrated in Fig. 2.

4.2. Effect of Regularization

The next comparison is to show the effect of regularization. If we apply the weights for windowed frame signals, when we obtain weight coefficient vectors in the $n-1$th and nth frames, the extracted signals with $\boldsymbol{w}^{(n-1)}$ and $\boldsymbol{w}^{(n)}$ should be very similar, Figure 4 shows how the regularization stabilizes the extraction of a signal. When there is no regularization term ($\epsilon = 0$), the extracted signals in the $n-1$th and nth frames are totally different, as shown in Fig. 4. However, by imposing regularization, a small difference between signals extracted by $\boldsymbol{w}^{(n-1)}$ and $\boldsymbol{w}^{(n)}$ can be seen in Fig. 4.

This effect is clearly observed when we track the values in $\boldsymbol{w}^{(n)}$. Figure 5 depicts the variation of components in $\boldsymbol{w}^{(n)}$ corresponding to channels CZ, O1, and Fp1. We found that there is a big "jump" when $\epsilon = 0$ (without regularization). However, when the cost function is regularized, we can see that the values of $w_i^{(n)}$ slowly change.

5. CONCLUSION AND FUTURE WORK

We proposed an adaptive RCE algorithm with regularization that takes into consideration the correlation between the extracted signals in the previous and current frames. It was proven that the maximization of the proposed cost function can be maximized by solving a generalized eigenvalue problem. Then, a method of adaptation was introduced. It was clarified by experimentation using practical EEG data that the proposed adaptive RCE can avoid *discontinuity* or *jump* in the weight coefficients and/or the extracted signal.

Our future work will be to use the proposed technique for practical BCI/BMI. In particular, extracting the mu rhythm related to the motor cortex and/or the steady state visual evoked potential (SSVEP) is of interest. Moreover, a fast algorithm to compute adaptive weight coefficients is under development.

6. REFERENCES

[1] S. Sanei and J. Chambers, *EEG Signal Processing*. Hoboken, NJ: John Wiley & Sons, 2007.

[2] A. Hyvaärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. England: John Wiley & Sons, 2001.

[3] J. Cao, N. Murata, S. Amari, A. Cichocki, and T. Takeda, "A robust approach to independent component analysis of signals with high-level noise measurements," *IEEE Trans. Neural Networks*, vol. 14, pp. 631–645, May 2003.

[4] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabilitation Engineering*, vol. 8, pp. 441–446, Dec. 2000.

[5] T. Tanaka and Y. Saito, "Rhythmic component extraction for multi-channel EEG data analysis," in *Proc. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008)*, (Las Vegas, NV), pp. 425–428, Apr. 2008.