CLASSIFICATION OF MOVEMENT-RELATED SINGLE-TRIAL MEG DATA USING ADAPTIVE SPATIAL FILTER

Futoshi Asano[1,2], *Masahiro Kimura*[1,3], *Tatsuhiko Sekiguchi*[1], *Yukiyasu Kamitani*[3]

1) Honda Research Institute Japan (HRI-JP), Japan
2) National Institute of Advanced Industrial Science and Technology (AIST), Japan
3) Advanced Telecommunications Research Institute International (ATR-CNS),Japan
Email: f.asano@aist.go.jp, sekiguchi@jp.honda-ri.com, {kmr/kmtn}@atr.jp

ABSTRACT

In this paper, a method for extracting and classifying movement-related brain signals is proposed. A single-trial MEG observation is first processed with a pre-whitening filter so that strong stationary interference is eliminated. Next, a brain signal effective for classification is extracted using an adaptive spatial filter. The extracted signal is then classified with a support vector machine. From the experimental results, it is shown that the classification rate of 62.6 % is obtained for the brain signals related to the three types of hand movements ("scissors-paper-rock").

***Index Terms*—** MEG, GEVD, adaptive spatial filter, SVM

1. INTRODUCTION

Much research has been done on the extraction and classification of a brain signal of interest from a single trial data obtained using noninvasive sensors such as those of magnetoencephalogram (MEG) or electroencephalogram (EEG) [1]. One of the problems in extracting these target signals from a single trial observation is that the target signal is usually buried in various spontaneous activities such as alpha rhythm, and the signal-to-noise ratio (SNR) is extremely low. To overcome this problem, the authors have previously proposed a method of pre-whitening of the sensor observation using generalized eigenvalue decomposition (GEVD) [2]. In this method, the spatial statistics (covariance between the sensors) of the interference signal are obtained in the period in which the target signal is not active. By using this interference covariance, the interference signals are effectively removed from the observation.

In this paper, an adaptive spatial filter (ASF) approach is proposed to extract brain signals effective for classification from the observation data purified by the GEVD pre-whitening filter. Next, features are extracted from the ASF output and are fed to the support vector machine (SVM) classifier. The results of the classification are reported and discussed.

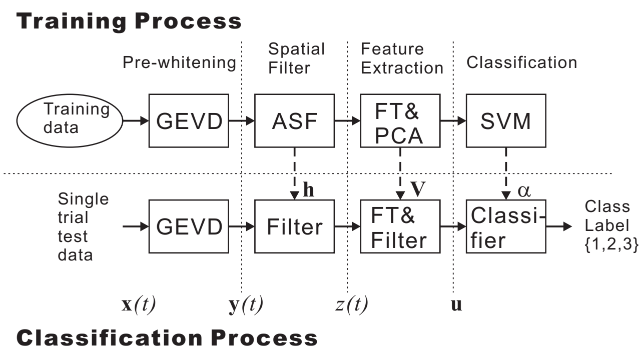

Fig. 1. Overview of the proposed method.

2. OVERVIEW OF THE METHOD

The aim of this study is to develop a method for classifying single-trial MEG observation related to hand movement (the game of "scissors-paper-rock").

Fig.1 shows a block diagram of the entire method. In the training process, MEG observations belonging to the training data set are processed with a pre-whitening filter using GEVD to eliminate stationary interference first. Next, the ASF is trained using the pre-whitened training data so that it can extract the signals containing key information for classification. In the feature extraction step, the filtered signal is transformed into the frequency domain by the Fourier transform (FT) to extract the features. Principal component analysis (PCA) is then performed for dimension reduction and normalization. In the classification step, a kernel-SVM classifier is trained so that the obtained features can be classified according to the corresponding hand movements.

In the classification process, a single trial MEG observation of unknown class (movement) is processed with the same steps as in the training process with the fixed coefficients obtained in the training process.

3. FILTERING

3.1. Pre-whitening using GEVD

The purpose of pre-whitening is to eliminate stationary interferences such as alpha rhythm from the observations. In this section, the GEVD pre-whitening filter that was originally proposed in the field of acoustics [3] and was introduced to brain signal processing [2] is summarized to facilitate understanding of the following sections.

Let us denote the MEG sensor observation as

$$\mathbf{x}(t) = [x_1(t), \cdots, x_M(t)]^T \tag{1}$$

where $x_m(t)$ denotes the observation at the mth sensor. The symbol M denotes the number of MEG sensors. Noise covariance and signal+noise covariance used for GEVD is then defined respectively as

$$\begin{aligned} \mathbf{K} &= E\left[\mathbf{x}(t)\mathbf{x}^T(t)\right], \text{ for } t \in \Psi_I & (2) \\ \mathbf{R} &= E\left[\mathbf{x}(t)\mathbf{x}^T(t)\right], \text{ for } t \in \Psi_{I+S} & (3) \end{aligned}$$

where Ψ_I denotes the period of time in which only stationary interference exists while Ψ_{I+S} denotes the period in which both a target brain signal and interference exist. It is assumed that the stationary interference in Ψ_I and that in Ψ_{I+S} have the same spatial characteristics.

GEVD of $\mathbf{K}$ and $\mathbf{R}$ is defined as

$$\mathbf{R}\mathbf{e}_m = \lambda_m \mathbf{K}\mathbf{e}_m \tag{4}$$

where λ_m and $\mathbf{e}_m$ denote the eigenvalue and eigenvector, respectively. Using the eigenvectors of GEVD, the pre-whitening filter $\mathbf{W}$ is obtained as

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{W}\mathbf{x}(t) \\ \mathbf{W} &= \mathbf{E}^{-T}\mathbf{G}\mathbf{E}^T & (5) \end{aligned}$$

where $\mathbf{E} = [\mathbf{e}_1, \cdots, \mathbf{e}_M]$ and the gain matrix $\mathbf{G}$ is defined as

$$\begin{aligned} \mathbf{G} &= \mathrm{diag}[\mathrm{g}_1, \cdots, \mathrm{g}_\mathrm{M}] & (6) \\ \mathbf{g} &= [g_1, \cdots, g_M] = [\overbrace{1, \cdots, 1}^{L}, 0, \cdots, 0] & (7) \end{aligned}$$

The symbol L denotes the dimension of the signal subspace. By using this pre-whitening filter $\mathbf{W}$, interference common in Ψ_I and Ψ_{I+S} is eliminated from the observation $\mathbf{x}(t)$.

3.2. Spatial Filtering using ASF

The purpose of ASF is to extract a brain signal effective for classification from the pre-whitened observation. Figure 2 shows a block diagram of ASF.

In the training step, pre-whitened observation $\mathbf{y}(t)$ is classified according to the class label $\{C1, C2, C3\}$ (C1:rock, C2:scissors. C3:paper). The classified observation is then averaged to obtain a "standard" waveform pattern common to its class, i.e.,

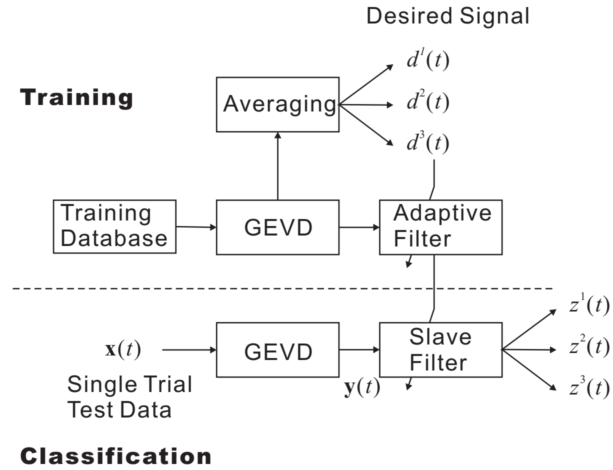

Fig. 2. Training and classification process of ASF.

$$\bar{\mathbf{y}}^i(t) = \frac{1}{N_i}\sum_n \mathbf{y}_n^i(t). \tag{8}$$

where $\mathbf{y}_n^i(t)$ denotes the nth sample waveform belonging to the ith class, and N_i denotes the total number of trials in the ith class in the training data. Next, the averaged waveform at one of the sensors is selected and is used as a desired signal for the adaptive filter training:

$$d^i(t) = \bar{y}_D^i(t) \tag{9}$$

where $\bar{y}_D^i(t)$ is the selected Dth component of $\bar{\mathbf{y}}^i(t)$. In a preliminary experiment, the averaged sensor observation was found to show a pattern peculiar to each class and can be classified when an appropriate sensor which is close to the brain area active during the movement is selected. The selected Dth sensor is termed the target sensor hereafter. The selection of the target sensor is discussed in Section 5. Figure 3 shows an example of the desired signal $d^i(t)$.

The coefficient vector for the adaptive filter is given by

$$\begin{aligned} \mathbf{h}^i &= \left(\mathbf{R}^i\right)^{-1}\mathbf{p}^i & (10) \\ \mathbf{R}^i &= \sum_n\sum_t \mathbf{y}_n^i(t)\left(\mathbf{y}_n^i(t)\right)^T & (11) \\ \mathbf{p}^i &= \sum_n\sum_t \mathbf{y}_n^i(t) d^i(t) & (12) \end{aligned}$$

In the filtering step, an unknown pre-whitened observation $\mathbf{y}(t)$ is filtered by each coefficient vector as

$$z^i(t) = \left(\mathbf{h}^i\right)^T \mathbf{y}(t) \tag{13}$$

Since the waveform common to the ith class $\bar{y}_D^i(t)$ is the desired signal $d^i(t)$, the adaptive filter is expected to yield waveform similar to $\bar{y}_D^i(t)$ when $\mathbf{y}(t) \in Ci$.

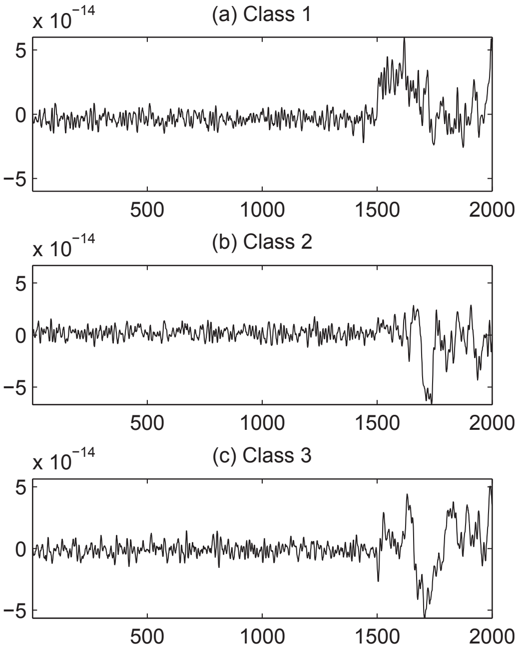

Fig. 3. Example of desired signal $d^i(t)$. The period [1,1500] and [1501,2000] correspond to Ψ_I and Ψ_{I+S}, respectively.

4. CLASSIFICATION

4.1. Feature Extraction

In this section, features for the classification are extracted from the filtered output $z^i(t)$. First, $z^i(t)$ is Fourier transformed as

$$Z^i(k) = F[z^i(t)] \tag{14}$$

where $Z^i(k)$ denotes a complex Fourier coefficient. Then, lower order coefficients are extracted from $\{Z^i(k)\}$, and the extracted features for each class are stacked as follows:

$$\mathbf{z} = \left[Z^1(1), \cdots, Z^1(K), \cdots, Z^3(1), \cdots, Z^3(K)\right]^T \tag{15}$$

The reason for extracting lower order Fourier coefficients up to K is that they represent the envelope of the time waveform which is considered to be effective for classification.

For efficient classification, the dimensions of the feature vector $\mathbf{z}$ are reduced and normalized by principal component analysis (PCA). In the training process, the following covariance matrix is calculated.

$$\mathbf{\Phi} = \sum_n \mathbf{z}_n \mathbf{z}_n^T \tag{16}$$

where $\mathbf{z}_n$ denotes the feature vector for the nth trial in the training data set. Let us denote the largest J eigenvalues and the corresponding eigenvectors of $\mathbf{\Phi}$, respectively, as

$$\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \cdots, \gamma_J) \tag{17}$$

$$\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_J] \tag{18}$$

Table 1. Parameters for extracting features.

Parameter	Value
Max order of FT coeffs. K	20
PCA order J	5
Signal subspace dimension L	60

Table 2. Classification rate for each session.

	Closed		Open		
Session	1	2	3	4	5
Score [%]	78.9	86.6	67.8	60.0	60.0

In the classification process, arbitrary feature vector $\mathbf{z}$ is processed by the following PCA filter $\mathbf{V}$:

$$\mathbf{u} = \mathbf{V}^T \mathbf{z} \tag{19}$$

$$\mathbf{V} = \mathbf{Q}\mathbf{\Gamma}^{-1/2} \tag{20}$$

Without the dimension reduction, the SVM classifier described in the next section tends to be overtrained, resulting in a poor classification rate for an open test. The parameters used for the feature extraction are summarized in Table 1.

4.2. Classification using SVM

The feature vector developed in the previous section is classified using SVM. A binary classifier, in which $C1$ and $\{C2, C3\}$ are classified first, and then $C2$ and $C3$ are classified was employed. As a kernel function of SVM, a radial basis function (RBF) with a Gaussian kernel was employed [4].

5. EXPERIMENT

5.1. Condition

The observation data were obtained using 208-channel MEG sensors. Brain activities during the movements of a hand of a single subject (the game of “scissors-paper-rock”) were measured. A single session consists of 90 trials (movements). Movements of ’paper’, ’rock’ and ’scissors’ were equally included in a session. There were five sessions, the first two sessions being used for training ASF, PCA and SVM, and the remaining three sessions being used as test data for the classification.

5.2. Results

Table 2 shows the classification rate for both closed and open tests when sensor #154 was used as the target sensor. A classification rate of 62.6 % was obtained for the open test.

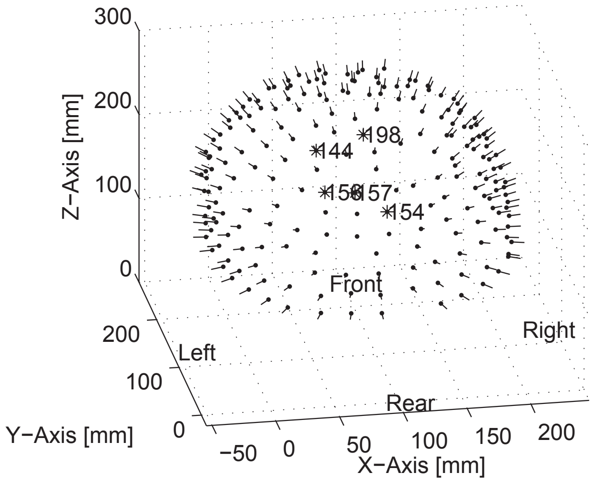

Fig. 4. Location of the target sensors which exhibited high classification rates.

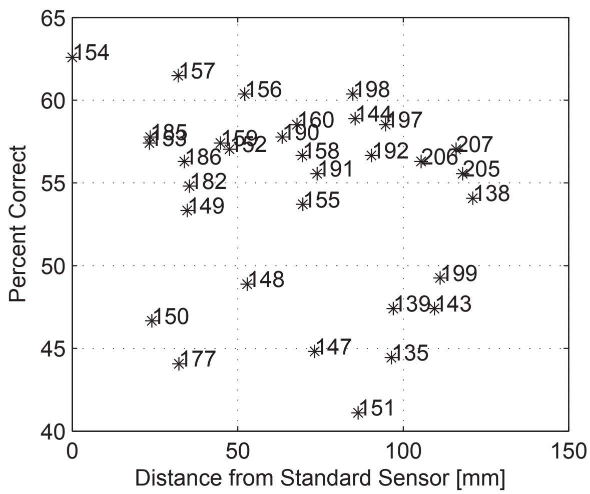

Fig. 5. Classification rate as a function of the distance between the target sensor and the standard sensor (#154).

The location of the target sensor in the case of the top five scores is depicted in Fig.4. From this, it can be seen that the target sensors located close to the rear upper part of the brain showed higher scores.

Figure 5 shows the classification rate when the target sensor was varied. The lateral axis of this figure is the distance between the tested target sensor and the standard sensor (sensor #154, which showed the highest rate). From this, it can be seen that the selection of the target sensor is sensitive to the classification.

5.3. Discussion

Figure 6 shows a schematic diagram of the signal sources and the sensors. When signal sources are distributed as depicted in Fig.6(a), a waveform pattern consisting of signals coming from various part of distributed sources appears at the sensors. This mixture of signals is similar to interference fringes of lights and reflects spatial information of the distributed sources. Due to this spatial information included in the waveform extracted by the proposed method, it is considered that the classification shows relatively good results. In the preliminary experiment in which a brain signal from a single dipole source was extracted by the adaptive beamformer [2] (the case of Fig.6(b)), on the other hand, the classification failed. In this case, no spatial information is included in the extracted waveform. From these, spatial information of the distributed sources is considered to contribute to the classification of movement-related brain signals.

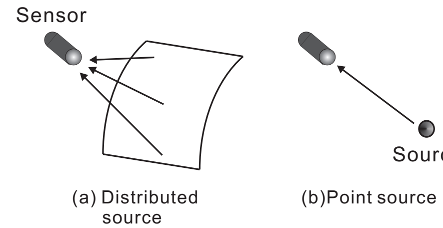

Fig. 6. Schematic diagram of distributed source and point source. For the sake of simplicity, the source signal is depicted as it reaches the sensor directly. In an actual physical phenomenon, however, the signal travels according to the Biot-Savart low.

6. CONCLUSION

In this paper, a method of extracting and classifying a movement-related brain signal from a single trial MEG observation is proposed. Observation was pre-whitened by GEVD and the information effective for classification was extracted using a spatial adaptive filter. From the results of classification, it was shown that a classification rate of 62.6 % was obtained.

7. REFERENCES

[1] M. Guimaraes, *et al.*, "Single-trial classification of MEG recordings," *IEEE Transaction Biomedical Engineering*, vol. 54, no. 3, pp. 436–443, 2007.

[2] F. Asano, *et al.*, "Localization and extraction of brain activity using generalized eigenvalue decomposition," in *Proc. ICASSP 2008*, 2008.

[3] S. Doclo and M. Moonen, "GSVD-based optimul filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2230–2245, Sep. 2002.

[4] B. Schölkopf and A. Smola, *Learning with kernels*, MIT Press, 2002.