Journal of Signal Processing, Vol. 15, No. 2, pp. 133-144, March 2011

PAPER

Common Spatial Pattern with Error-Correcting Output Code and Preprocessing on Neural-Network-Based Brain-Computer Interface

Prapong Prechaprapranwong[1], Kenji Nakayama[2] and Akihiro Hirano[3]

Graduate School of Natural Science and Technology, Kanazawa University
Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan
E-mail: prapong@leo.ec.t.kanzawa-u.ac.jp[1], nakayama@t.kanazawa-u.ac.jp[2], hirano@t.kanazawa-u.ac.jp[3]

Abstract In previous works, Fourier-transform-based electroencephalogram (EEG) feature extraction with effective preprocessing and multilayer neural networks (MLNNs) has been applied in classifying cognitive mental tasks in a brain-computer interface (BCI). However the EEG signals have some shortcomings of noise and spatial resolution. In this paper, spatial filtering techniques are employed. The use of common spatial pattern (CSP) filter is one of the well-known spatial filtering methods in the BCI framework. The CSP filters aim to find the most discrimination between two classes. Also, the CSP filter is mostly applied to extracting features of motor-imagery-based EEGs. In this work, we apply the CSP filter to cognitive mental task classification. A feature extraction method is proposed for the CSP filter, in which Fourier-transform-based features are used. An error-correcting output code (ECOC) is combined with the CSP filters to realize a multiclass BCI system, and also to correct some misclassifications. The proposed BCI system combines several binary classifiers, called an 'elemental classifier' in a parallel form. An elemental classifier consists of the CSP filters, FFT, preprocessing and an MLNN. The output of the MLNN in one elemental classifier is a 1-bit code, that is, 1 or 0. The output of the parallel form becomes a multibit code that is generated by the ECOC. All elemental classifiers receive the same EEG signal. An elemental classifier classifies two groups of mental tasks. The assignment of mental tasks to two groups is accomplished by the ECOC. The CSP filters and the MLNN in each elemental classifier are optimized so as to classify two groups of mental tasks. The ECOC is further decoded into the final code, which expresses a single estimated mental task. We also develop a nonuniform resolution sampling technique in preprocessing to approximate spectral information in useful frequency bands and suppress noise in the high-frequency bands. The experimental results of 4 subjects show that the proposed BCI system can boost the correct classification rates from 66%~88% to 84%~96% and suppress the misclassification rates from 4%~26% to 4%~12%.

Keywords: brain-computer interface (BCI), electroencephalogram (EEG), common spatial filter, error-correcting output code (ECOC), multilayer neural network (MLNN)

1. Introduction

In a noninvasive brain-computer interface (BCI) framework, the EEG-based BCI has recently been approved as one of the most interesting topics. However EEG signals have some shortcomings of noise and spatial resolution. The voltage potentials from sources can contribute within a small radius through the scalp toward each electrode, making the observed signals obscure and noisy [1]. Moreover EEGs are also inherently nonstationary owing to changes in the individual subject's brain processes through out an experimental session [2], [3]. Therefore, several spatial filtering techniques are applied to remedy the noise and obtain more localized signals, or signals corresponding to single sources. Some examples of prominent techniques that are applied in the BCI framework are bipolar filtering, the common average reference method, Laplace filtering, and finally, the statistical linear-transformation-based spatial filtering for linearly transforming raw EEGs to new feature spaces [1]. The well-known spatial filtering algorithms in this family are principle component analysis (PCA), independent component analysis (ICA) and common spatial pattern (CSP). The aforementioned algorithms are recently being used in many BCI studies [1], [4]-[10] as an effective preprocessing process that provides feature extraction and dimension reduction at the same time. With a suitable feature extraction method, the unnecessary dimensions of signals, which are determined as noise, could be reduced while still retaining necessary information.

In this study, we focus on CSP [11] spatial filtering. We concentrate on the concept of CSP aiming to find the greatest discrimination between two data sets by optimizing the ratio between within-class scatter and between-class scatter of those two data sets. This is unlike the PCA [12], which is based on the scatter of the whole data set and decomposes it into the principle components that are ranked by variance. The principle of ICA [13] is to solve the blind source separation (BSS) problem by finding the most mutual statistical independence between source signals. ICA has also been widely used for analyzing brain signals and to remove artifacts in brain signals [14], [15]. However, in ICA, there exists a permutation problem, that is, the order of extracting features may change data by data. The features are arranged as a one-dimensional vector that is used in a classifier. Different orders cause different input patterns for the classifier, resulting in insufficient classification performance [16], [17].

CSP has been successfully applied to classify motor-imagery-based EEG [1], [4]-[10]. Furthermore, CSP has also been applied to slow cortical potential (SCP)-based EEG. In this work, the CSP method will be applied to different behaviors of EEG data sets, which include nonmotor-imagery mental tasks, e.g., mathematical multiplication, letter composition, 3D object rotation, and visual number counting.

These EEG data sets were provided by the Colorado State University and have been used in many studies [2], [3], [16]-[23]. In previous research, many feature extraction methods have been employed, e.g., min-max amplitude, mean, variance, standard derivation, power spectral density, frequency band powers, asymmetry ratios, autoregressive (AR) coefficients, eigenvalues of correlation matrix, and wavelet packet entropy [24].

Nakayama and coworkers [17]-[19] used the Fourier-transform-based features and some optimized preprocessing with the multilayer neural network (MLNN) to successfully classify five classes of mental tasks for a single subject at an accuracy of 78~88%. In this study, in order to confirm efficiency in general, four subjects are considered. We follow that method by adding CSP filtering into the BCI system, as expressed in Sec. 4.1. We also propose a modified method for the sampling reduction procedure, elaborated in Sec. 4.3.

Several kinds of classification algorithms for EEG-based BCI systems have been compared in [25]. MLNN, denoted MLP in [25], has good classification ability. It does not always provide the best performance in all cases. In some cases, the support vector machine (SVM) can provide more efficient performance than MLNN owing to its regularization. However, by using the generalization method, the classification ability of MLNN can be drastically improved [19]. Therefore, in this study, we employ the MLNN with the generalization [19] as a classifier.

This work engages a multiclass classification problem. However, originally, CSP was designed for binary-class problems. To deal with multiclass problems, Dornhege and coworkers [6], [7] proposed many multiclass extension approaches to extend two-class CSP to multiclass application. Those methods are called, CSP-IN (CSP within multiple-binary classification) [4]-[5], CSP-OVR (binary CSP combination with one versus the rest strategy) [6], [7], [10] and CSP-SIM (CSP with simultaneous diagonalization method) [6], [7], [9]. We investigate those approaches that are based on MLNNs and for the CSP-IN approach, we propose an ensemble of binary MLNNs with the error-correcting output code (ECOC) framework.

ECOC was originally used in signal transmission approaches for correcting bit distortion cause by noisy communication. Later, Dietterich and Bakiri [26] proposed ECOC as a general framework for solving multiclass problems by reducing the multiclass problems into several binary class problems with an error-correcting property. We merge this property with the CSP, which theoretically makes the classes most discriminant. We intend to make the CSP-ECOC combination raise the accuracy rates for mental task classification by using the EEG data.

The remainder of the paper is organized as follows. CSP analysis and multiclass extensions are described in Sec. 2. The ECOC framework is described in Sec. 3. A proposed BCI system is described in Sec. 4 using a block diagram consist of CSP filtering, feature extraction, and preprocessing. The experimental setup is described in Sec. 5. In Sec. 6, experimental results and discussions are presented. Finally, our work is concluded in Sec. 7.

2. Common Spatial Pattern Method

The CSP algorithm [11] is aimed at finding spatial filters that project the original signals to the most different spaces between two data sets in term of variance. The variance of a data set is maximized and the variance of another data set is minimized, simultaneously.

2.1 CSP algorithm

1. Let 2 EEG data sets be denoted $\mathbf{X}_1$ and $\mathbf{X}_2$, the dimension of which is $P \times K$, where P is the number of electrodes of the EEG data, and K is the number of samples in the time domain.
2. Compute the normalized auto-covariance matrices $\mathbf{S}_i$ for each class.
$$\mathbf{S}_i = \frac{\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}}{trace(\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}})}, \quad i \in \{1,2\} \qquad (1)$$
3. The whitening transformation is performed by computing the sum of all class auto-covariance matrices.
$$\mathbf{S}_{\mathbf{sum}} = \mathbf{S}_1 + \mathbf{S}_2 \qquad (2)$$
4. Then, decompose $\mathbf{S}_1$ the eigenvector matrix and the eigenvalues of the matrix $\mathbf{S}_{\mathbf{sum}}$ as follows:
$$\mathbf{S}_{\mathbf{sum}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}} \qquad (3)$$
$\mathbf{U}$ and $\mathbf{\Lambda}$ are the eigenvector matrix and eigenvalue matrix of $\mathbf{S}_{\mathbf{sum}}$, respectively. Then the whitening transformation matrix becomes
$$\mathbf{V} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{T}} \qquad (4)$$
5. Apply the whitening process to both the auto-covariance matrices:
$$\hat{\mathbf{S}}_1 = \mathbf{V}\mathbf{S}_1\mathbf{V}^{\mathrm{T}} \qquad (5)$$
$$\hat{\mathbf{S}}_2 = \mathbf{V}\mathbf{S}_2\mathbf{V}^{\mathrm{T}} \qquad (6)$$
6. CSP analysis is given by the simultaneous diagonalization of these two covariance matrices. Thus, $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$ should share a common eigenvector matrix, and the corresponding eigenvalues for the sum of $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$ should be one. Therefore, $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$ can be decomposed as follows:
$$\mathbf{W}^{\mathrm{T}}\hat{\mathbf{S}}_1\mathbf{W} = \mathbf{D} \qquad (7)$$
$$\mathbf{W}^{\mathrm{T}}(\hat{\mathbf{S}}_1 + \hat{\mathbf{S}}_2)\mathbf{W} = \mathbf{I} \qquad (8)$$
$\mathbf{W}$ is a common eigenvector matrix of $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$, $\mathbf{D}$ is an eigenvalue matrix of $\hat{\mathbf{S}}_1$ and $\mathbf{I}$ is an identity matrix
7. Practically, we choose only a few of the most important eigenvectors from $\mathbf{W}$ by sorting the

eigenvalues in $\mathbf{D}$ in descending order and choosing $2m$ eigenvectors ($2m < P$) that correspond to the m largest and the m smallest eigenvalues. Then we obtain

$$\mathbf{W}_m = [\mathbf{w}_1,\ldots,\mathbf{w}_m,\mathbf{w}_{P-m+1},\ldots,\mathbf{w}_P] \quad (9)$$

where $\mathbf{w}_i$ is an eigenvector corresponding to an eigenvalue in $\mathbf{D}$.

8. Then, we obtain the spatial filters of the CSP transformation matrix as

$$\hat{\mathbf{W}}_m = \mathbf{W}_m{}^{\mathrm{T}}\mathbf{V} \quad (10)$$

9. Finally, the projected signals are defined as

$$\mathbf{X}_{\mathrm{CSP}\,i} = \hat{\mathbf{W}}_m \mathbf{X}_i,\ i \in \{1,2\} \quad (11)$$

2.2 Extension to multiclass CSP

Originally, CSP was designed for 2-class problems. To extend this concept to multiclass approaches, some approaches of multiclass extensions have been proposed. Muller-Gerking and coworkers [4] proposed a multiclass extension of CSP based on pairwise classification and voting. Then Dornhege and coworkers [6], [7] proposed the other two algorithms based on the concept of CSP. These approaches are summarized below.

1) CSP-IN: The problem is separated into several binary problems, and is found the optimum CSP for each binary problem.
2) CSP-OVR: Several binary CSPs are used for a single multiclass classifier.
3) CSP-SIM: CSP is designed by the joint approximate diagonalization (JAD) method and is applied to a single multiclass classifier.

CSP-IN and CSP-OVR approaches are similar in the concept of extension from a 2-class CSP by adding more binary CSP processes to form a multiple classifier. An important difference between them is the classification process. The CSP-IN uses several binary classifiers to deal with each binary problem. On the other hand, the CSP-OVR solves these binary problems by using a single multiclass classifier.

CSP-SIM derives from the concept of the 2-class CSP algorithm. The CSP algorithm will find a simultaneous diagonalization of both covariance matrices, in which the sum of eigenvalues is unity. Thus it can be extended to many classes if we can approximate a simultaneous diagonalization for the many-class problem. However, unlike the 2-class problem, there is no general strategy for choosing the appropriate CSP patterns for multiclass CSP, because the 2-class problem uses the strategy of the highest or the lowest eigenvalue.

Dornhege and coworkers [6], [7] have proposed a heuristic way of solving this problem using some score strategy. Given $\mathbf{D}$ is an approximate simultaneous diagonal matrix, computed by Joint Approximate Diagonalization (JAD), it is in a form of concatenated eigenvalue matrices i.e., $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2,\ldots, \mathbf{D}_N]$, N is the number of classes. An appropriate eigenvector is chosen for the i^{th} class, corresponding to the highest score of eigenvalue λ_{ij} in each submatrix $\mathbf{D}_i$, on the basis of the following criteria:

$$score(\lambda_{ij}) = \max(\lambda_{ij}, 1/(1+(N-1)^2\lambda_{ij}/(1-\lambda_{ij}))) \quad (12)$$

Note that if one eigenvector is selected more than once, it is replaced by the eigenvector with the next highest score.

3. Error-Correcting Output Codes

3.1 Multiclass classifier based on ECOC

Error-correcting output codes (ECOC) are approved as a general framework for combining several binary problems to address a multiclass problem [26]-[28]. Consider L binary classifiers. Each classifier outputs a single bit code. They are combined in a parallel form to produce an L-bit code, which is an output vector. This output vector is further decoded into the final code, which expresses the classification result.

The ECOC framework consists of two steps.

Step 1 (Coding):

A codeword is assigned to each class. For an N-class problem to be solved by using several binary classifiers, we required L binary classifiers. To supervise each L binary classifier, a set of N binary labels (1 or 0) is assigned. Then, combining those N-bit labels from L binary classifiers, a $N \times L$ coding matrix $\mathbf{M}$ is defined. A row vector of $\mathbf{M}$ is called a codeword, which used to determine the label of classes. A column vector of $\mathbf{M}$ corresponds to labels of subproblems for training each classifier.

Step 2 (Decoding):

The output vector for new EEG data will deviate from the target vector that is the codeword. In this case, the most similar codeword, which has the shortest distance from the output vector, is selected from the codeword table, and decoded into the final code, which expresses a mental task.

The performance of ECOC mostly relies on the codeword table that applied to the BCI system. The regulations of designing codewords have been discussed in many studies [27]-[29]. The method of generating codewords can be categorized into the following three types.

1) Algebraic coding theory methods
2) Randomization
3) Unique codewords for a particular data set

In this paper, we use the generalized algebraic coding theory, e.g., one-per-class coding (OPC), Hadamard-type coding, and exhaustive-ECOC (E-ECOC). For decoding, we use the L_1-norm distance. We prefer to the use L_1-norm distance instead of the hamming distance because it is more flexible for adjusting the threshold for rejecting the classification results due to small outputs.

3.2 Coding and decoding strategy

Coding strategy: one-per-class coding

One-per-class (OPC) coding is defined as one of standard output coding strategies. The simplest code defines a single binary value at the corresponding position for each class. OPC has also been used as the target vectors in a single MLNN, as shown in Table 1. C_j expresses the j^{th} classifier used in a binary classifier, and Ω_i indicates the i^{th} class. The number of binary classifiers, which are combined in a parallel form, is L=5, and the number of classes is also N=5.

Table 1 Codeword table of OPC coding for 5-class problem

Class	Classifier				
	C1	C2	C3	C4	C5
Ω_1	1	0	0	0	0
Ω_2	0	1	0	0	0
Ω_3	0	0	1	0	0
Ω_4	0	0	0	1	0
Ω_5	0	0	0	0	1

This table corresponds to the codeword matrix **M**. The row vectors indicate the classes, and the column vectors indicate the elemental classifiers. Namely, the i^{th} row and the j^{th} column mean the i^{th} class and the j^{th} elemental classifier, respectively.

Coding strategy: Hadamard-type coding

Hadamard-type coding [30] is created from the Hadamard matrix, which is a 2^L square matrix having elements of either +1 or -1 and mutually orthogonal rows. Hadamard coding has good points in two assessments:

1) Row separation: each codeword should be well separated in the Hamming distance from each of the other codewords.
2) Column separation: The elemental classifiers should be uncorrelated to each other, that is, they should not be redundant.

To design the Hadamard-type coding matrix, first we refer to a Hadamard matrix whose order is greater than the required number of classes. The two main methods of constructing the Hadamard matrices, Sylvester construction and Paley construction, have been discussed in [31].

For example, in the 5-class problem, the minimum order of the Hadamard matrix, which covers this problem, is 8 (2^3) as shown in Eq. (13).

$$H_8 = \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{bmatrix} \quad (13)$$

In the above equation, + and – are replaced by 1 and 0. The column that is identical (all ones or all zeros), is useless for a classifier, so the first column of the Hadamard matrix is neglected. In this case, we require 5 classes of codewords, so the last three rows of the Hadamard matrix are pruned. Finally, the Hadamard-type coding matrix is obtained, as shown in Table 2.

Table 2 Codeword table of Hadamard type for 5-class problem

Class	Classifier						
	C1	C2	C3	C4	C5	C6	C7
Ω_1	1	1	1	1	1	1	1
Ω_2	0	1	0	1	0	1	0
Ω_3	1	0	0	1	1	0	0
Ω_4	0	0	1	1	0	0	1
Ω_5	1	1	1	0	0	0	0

Coding strategy: exhaustive ECOC

Dietterich and Bakiri [26] have proposed a code and a procedure for generating a well-balanced hamming distance between rows and include all possible nontrivial and nonredundant $2^{(N-1)}$ -1 length codes for the N-class problem, called exhaustive ECOC (E-ECOC). These codes are recommended to be used for $3 \leq N \leq 7$. The procedure for generating E-ECOC is as follows. Assign the first row as all ones. The second row consists of $2^{(N-2)}$ zeros followed by $2^{(N-2)}$-1 ones. The third row consists of $2^{(N-3)}$ zeros followed by $2^{(N-3)}$ ones followed by $2^{(N-3)}$ zeros followed by $2^{(N-3)}$-1 ones. The i^{th} row consists of alternating $2^{(N-i)}$ zeros and ones. The last row contains 0, 1, 0, 1, 0, 1,…,0. For $8 \leq N \leq 11$, Dietterich and Bakiri suggested selecting a good subset of columns from the exhaustive code by means of the optimization algorithm. For N > 11, the random code generation with a hill-climbing procedure is recommend.

The E-ECOCs for 5 classes, obtained from the generation procedure detailed in [26], are shown in Table 3. In this study, we employ the E-ECOC.

Table 3 Codeword table of E-ECOC 5-class problem

	Classifier														
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
Ω_1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ω_2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
Ω_3	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1
Ω_4	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
Ω_5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Decoding strategy: L_1-norm distance

In the decoding process, the output vector from the parallel form of the elemental classifiers is compared with the patterns in the codeword table, and the most similar pattern, that is, the codeword, is selected, which corresponds to the real output. Similarity is evaluated using the L_1-norm-based distance. Let **y** be the output vector of the parallel form of the L elemental classifiers, $\mathbf{y} = [y_1, y_2,\ldots, y_L]^T$, where, y_j is the output of the j^{th} elemental classifier. The L_1-norm distance for the i^{th} class is defined by

$$d_i = \sum_{j=1}^{L} \left| c_{ij} - y_j \right| \quad (14)$$

where c_{ij} is a target value for the i^{th} class and the j^{th} elemental classifier. In other words, c_{ij} is an element at the i^{th} row and the j^{th} column in the codeword matrix.

3.3 Performance of correcting codes

The efficiency of ECOC is determined by the number of error bits that can be corrected. The capability of error correcting is $\lfloor (d-1)/2 \rfloor$ bits, where d is the distance between a row's codeword in the codeword matrix and $\lfloor \cdot \rfloor$ is the *floor* operation. For example, in a 5-class problem, a 15-length codeword, as shown in the Table 3, is applied, then distance d is 8; thus, the capability of error correcting $\lfloor (d-1)/2 \rfloor$ is 3 bits. In contrast, for the Hadamard-type, as shown in Table 2, the capability of error correcting is 1 (d=4). The OPC coding has no capability of error correcting (d=2). Because the distance *d* correspond to the code's performance, so it is possible to increase the performance of coding by expanding the codeword length. However, the longer codeword means more classifiers and more computation. The issues concerning the optimization of codeword length are discussed in [28], [29], but they are not in the scope of our study. The performance of different correcting codes applied to CSP-ECOC is also discussed in Sec. 6.

4. Proposed BCI System

4.1 Block diagram of proposed BCI system

The proposed BCI system combines the binary classifiers using the CSP filters and the ECOC framework; hence, it is called CSP-ECOC. We propose that combination of discriminative spatial filtering of CSP and error-correcting properties of ECOC can enhance the performance of EEG classification.

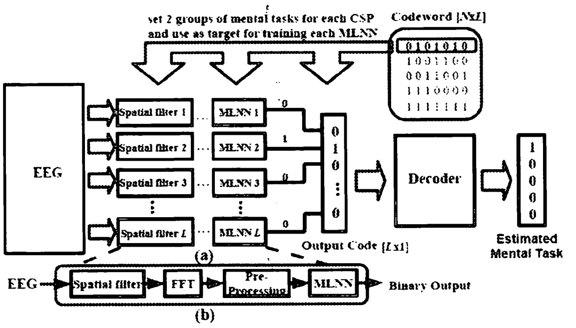


Fig.1 Block diagram: (a) proposed BCI system based on CSP filter and ECOC, and (b) elemental classifier

The block diagram of the proposed BCI system is shown in Fig. 1(a). Figure 1(b) shows the binary classifiers, called 'elemental classifiers' in this paper, that include the spatial filter, which is the CSP filter, FFT, preprocessing, and an MLNN in this order. They are arranged in a parallel form. All elemental classifiers receive the same EEG data, and classify two groups of mental tasks. The output of a single elemental classifier is 1 bit, that is, 1 or 0. A set of the output of all elemental classifiers corresponds to the codeword, which is used as the target pattern. The CSP filters and the MLNNs are optimized to classify two groups of mental tasks. This grouping is determined by the codeword matrix **M**. One example is shown here. Let the number of classes, that is, the mental tasks, be N=5 and the 1^{st} column of **M** be [0 1 0 1 1]. The 1^{st} elemental classifier, actually the CSP filter and the MLNN, is optimized so as to classify two groups of the mental tasks. The 1^{st} group contains the 1^{st} and 3^{rd} mental tasks, and the 2^{nd} group includes the 2^{nd}, 4^{th} and 5^{th} mental tasks, respectively.

The codeword matrix plays an important role in the CSP-ECOC approach. The i^{th} row indicates the i^{th} class, that is, the i^{th} mental task. It is also used as the target code of the parallel form of the elemental classifiers. The j^{th} column corresponds to the j^{th} elemental classifier. In other words, the i^{th} row and the j^{th} column element c_{ij} is the 1-bit target code for the i^{th} mental task and the j^{th} elemental classifier, respectively. Therefore, the numbers of rows and columns in the codeword matrix are equal to the numbers of mental tasks and elemental classifiers, respectively.

The output of the parallel block is corrected as follows. Let the output be $\mathbf{y}=[y_1, y_2,\ldots, y_L]$ and the target code, that is, the i^{th} row vector, be $\mathbf{c}_i=[c_{i1}, c_{i2},\ldots, c_{iL}]$ of **M**. The winner code is searched from a set of target codes, which minimizes the L_1-norm of $|\ \mathbf{c}_i - \mathbf{y}\ |$. The winner target code $\mathbf{c}_i$ is further decoded into the final output code, for example, [1 0 0 0 0], which means the 1^{st} mental task is estimated. One of the target codes uniquely corresponds to one of the final output codes. With this decoding system, the mental task estimation is more generalized by allowing some margin for binary misclassified output. The capability of error correcting has been described in Sec. 3.2.

Moreover, a rejection threshold is also employed in order to neglect vague classifications. If the L_1-norm distances for the entire target codes are larger than the threshold of rejection, then we cannot select a winner, and the mental task estimation will be rejected.

4.2 CSP filtering

The EEG data are measured at several points on the scalp. We use the EEG data sets, which are measured at 6 positions, i.e., C3, C4, P3, P4, O1, and O2 (according to the International 10-20 system) [20]. EOG is not included for computing the CSP filters. It is used for detecting eye artifacts in a different way. Six CSP filters are created from 6 positions of the EEG data sets. One CSP filter is used to emphasize the discrimination between the two considered mental tasks by linearly transforming the EEG data into a new signal. For optimization, a fewer number of effective CSP filters are selected for the most discrimination. The number of required CSP filters is variable. The number of features is also dependent on the number of chosen CSP filters.
For example, let the matrix $\mathbf{X}_{\mathbf{org}}$ represent 6 positions of the EEGs recorded using a 250 Hz sampling frequency during a 10 second interval. Therefore, $\mathbf{X}_{\mathbf{org}}$'s dimensions are 6×2500. If 4 CSP filters are chosen to form the transformation matrix $\hat{\mathbf{W}}_m$, then it becomes a 4×6 matrix. After transformation using Eq. (15), $\mathbf{X}_{\mathbf{CSP}}$'s dimensions are reduced to 4×2500.

$$\mathbf{X}_{\mathbf{CSP}} = \hat{\mathbf{W}}_m \times \mathbf{X}_{\mathbf{org}} \tag{15}$$

4.3 FFT and preprocessing

Data segmentation

The raw EEG signals have 2500 samples in a 10 s interval. In order to make the response of the BCI system faster, the EEG signal is segmented into short periods of 0.5-second length and 50% overlapping or shifted by 0.25 s. As a result, the BCI system can respond every 0.25 s. One segment includes 125 samples, and 39 segments are included in the original EEG signal. The effect of EEG's segmentation on classification performance will be examined in Sec. 6.

Fourier transformation of spatial filter outputs

The CSP analysis can be directly applied to extracting the variance-based features. However, we attempt to employ CSP filtering followed by Fourier transform to extract features that provide the spectral information [16]-[19].

The Discrete Fourier Transformation (DFT) of the transformed EEG signal $x_{CSP}(n)$, given by Eq. (15), is expressed by

$$X(k) = \sum_{n=0}^{K-1} x_{CSP}(n) \exp(-j\frac{2\pi}{K}kn),\ k = 0,1,2,\ldots,K-1 \quad (16)$$

where K is the number of EEG samples.

4.4 Frequency ranges and nonuniform resolution frequency ranges

It is generally believed that frequencies above 40 Hz carry little information. Keirn [2] and Keirn and Aunon [3] employed power values and asymmetry ratios from the four common frequency bands: delta (0-3 Hz), theta (4-7 Hz), alpha (8-13 Hz) and beta (14-20 Hz). Palaniappan [21], [22] improved the performance of the BCI system by using an additional low gamma band (24-37 Hz) spectral power and asymmetry ratio. However, some researchers have argued that in the high-frequency region (>40 Hz), there also exists some useful information. Graimann and coworkers [24] used higher frequency components (70-90 Hz) in the event-related potential (ERP)-based BCI system. Fitzgibbon and coworkers [32] investigated the EEG rhythm activity induced by eight cognitive tasks, i.e., visual checkerboard, reading, subtraction, music, expectancy, word learning, word recall, and video segment. They have proposed that sustained high-frequency EEG activity (30-100 Hz) connects to thinking processes. On the basis of the above investigations, we use all frequency ranges 0.1-100 Hz.

Uniform and nonuniform frequency resolution

When the EEG signal is sampled with a 250 Hz sampling frequency during 10 s, 2500 samples are generated. This data set is transformed by FFT, resulting in 2500 samples. Since, the EEG signal is a real number, half the number, that is, 1250 samples, is sufficient. 1250 is still a large number as MLNN input data. The number of samples after FFT is reduced by averaging successive samples. At the same time, the number of samples assigned to frequency bands is optimized.

Significant information on mental tasks appears in the low-frequency region. Therefore, sample intervals after averaging should be optimized for the BCI system. The conventional BCI system [18] arranges the samples uniformly, that is, 'uniform frequency resolution' which may cause some leakage of important information in the low-frequency region and inadequate noise reduction in the high-frequency region.

In this paper, to overcome these problems, 'non-uniform frequency resolution' is proposed. The samples after FFT are positioned uniformly, that is, the interval between the samples is given by 125 Hz/1250=0.1 Hz. The number of samples is further reduced by averaging the successive samples. The number of the samples to be averaged changes depending on the frequency bands: 40 samples during 0-4 Hz and 4-8 Hz, 80 samples during 8-16 Hz, 160 samples during 16-32 Hz, 320 samples during 32-64 Hz, and 610 samples during 64-125 Hz are averaged. This means only one sample is assigned to each band. Since the number of bands is 6, the total number of assigned samples is 6. This nonuniform band division, that is, nonuniform resolution, is shown in Table 4 and Fig. 2. Figure 2 also shows the uniform and non-uniform band divisions. In this figure, the samples on the left side are not used because the amplitude of FFT is symmetrical for real signals.

Table 4 Nonuniform division of frequency bands

Sample	1	2	3	4	5	6
Frequency (Hz)	0-4	4-8	8-16	16-32	32-64	64-125
Band	Delta	theta	alpha	Beta	gamma	omega

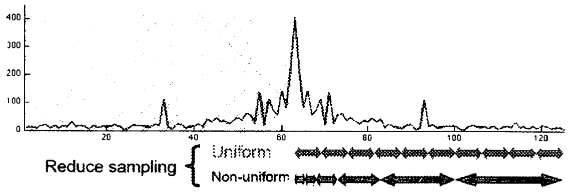


Fig.2 Uniform and nonuniform resolutions

4.5 Nonlinear normalization and MLNN input

Nonlinear normalization

The information of mental tasks may be widely distributed, not only in the peak's frequency band. Important information for classification may be included in small nonprominent frequency bands. Moreover, naturally in the neural networks, large inputs play an important role. To avoid the neural network's biased learning, this nonlinear normalization [18] is applied after the sample reduction.

$$f(x) = \frac{\log(x - x_{\min} + 1)}{\log(x_{\max} - x_{\min} + 1)} \quad (17)$$

Here, x means the sample value, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values in all data sets.

MLNN input patterns

The EEG data set for one mental task and one trial includes the following data: 2500 samples $\times$ 7 measuring points. The data of 6 measuring points, except for EOG, which are counted as 6 signal sets here, are further transformed though the CSP filters, resulting in a reduced number of signal sets, for example, 4 signal sets. In this case, 4 CSP filters are required. Each signal set is further processed through FFT, sample reduction by nonuniform resolution, and nonlinear normalization. At the same time, the EOG signal is also processed in the same way, except for the CSP filter. After that, the number of signal sets is 4+1=5 sets. Each signal set has 6 samples. Five signal sets, which include 6 samples, are ordered in a 1-dimensional vector, which is applied to the MLNN.

5. Experimental Setup

5.1 Data acquisition

In this study, we use the brainwave data sets that are available from the web site of Colorado State University [20]. The EEGs were recoded from 7 subjects, where 7 channels of electrodes were placed upon the scalp at the positions C3, C4, P3, P4, O1, and O2 according to the International 10-20 system. The last channel was the EOG recorded between the forehead above the left brow line and another on the left cheekbone. Recordings were made with reference to electrically linked mastoids A1 and A2. The EEG signals were recorded at a 250 Hz sampling rate for 10 seconds (total 2500 samples per channel). Recording was performed with a bank of Grass 7P511 amplifiers whose bandpass analog filters were set at 0.1 to 100 Hz. The experiment was divided into 5-trial sessions. Subjects were asked to perform 5 mental tasks and repeated 5 trials of each task in one day. They returned to do a second set of five trials on another day. Without any physical movement, the subjects were asked to perform the following mental tasks.

- *Baseline* The subjects were asked to do nothing, but relax.
- *Multiplication* The subjects were instructed to calculate a nontrivial multiplication problem.
- *Letter composition* The subjects were asked to mentally compose a letter.
- *Rotation* The subjects were asked to rotate a complex three-dimensional object in their mind.
- *Counting* The subjects were asked to write visualized numbers one by one, deleting the previous number before writing the next number.

5.2 Several BCI approaches

In order to evaluate the efficiency of the proposed method, we compare several conventional BCI systems. They are introduced in this section.

Conventional: The BCI system [18], [19] consists of FFT, preprocessing and a MLNN. 70 samples and 42 samples for uniform and nonuniform frequency resolution, respectively, are used as the MLNN input data.

CSP-OVR: 2 CSP-OVR-based spatial filters are used to extract Fourier-transformed features. The feature samples are uniformly reduced to 5 per electrode position. After the nonlinear normalization, the feature samples are ordered in a 55-dimensional vector. In the same way, by using the nonuniform frequency resolution, that is, 6 samples for 6 frequency bands, a 66-dimensional feature vector is generated. MLNNs are used for classifiers.

CSP-SIM: The CSP-SIM method is used to extract Fourier-transformed features. Furthermore, through sample reduction, nonlinear normalization and the ordering of all data sets, 30- and 36-dimensional feature vectors are generated for the uniform and the nonuniform frequency resolution, respectively. MLNNs are used for classifiers.

Table 5 Examples for dimension of feature vectors which are MLNN input data

Method	Number of spatial filters	Uniform resolution	Nonuniform resolution
Conventional	-	70	42
CSP-OVR	2	55	66
CSP-SIM	1	30	36
Conventional-ECOC	-	70	42
CSP-ECOC	4	25	30

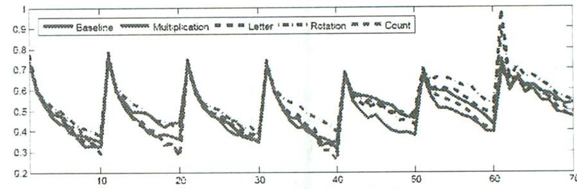


Fig.3 Examples of input patterns for MLNN by Conventional method [18]

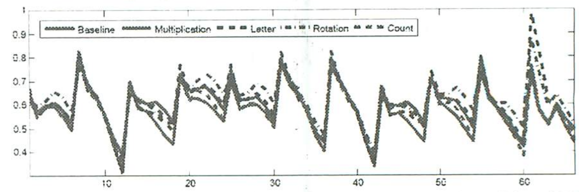


Fig.4 Examples of input patterns for MLNN by CSP-OVR method

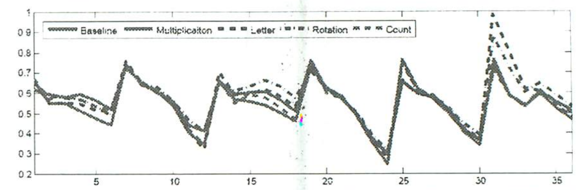


Fig.5 Examples of input patterns for MLNN by CSP-SIM method

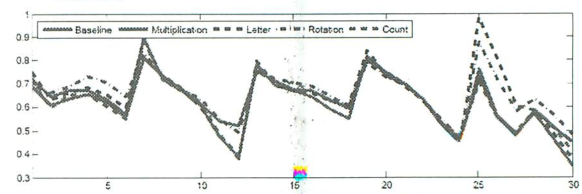


Fig.6 Examples of input patterns for MLNN by CSP-ECOC method for column-code $[1\ 0\ 1\ 0\ 1]^T$

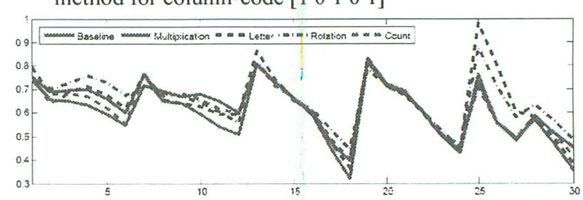


Fig.7 Examples of input patterns for MLNN by CSP-ECOC method for column-code $[1\ 0\ 1\ 1\ 0]^T$

Conventional-ECOC: The conventional method, which employs FFT and the preprocessing, is used to generate 70- and 42-dimensional feature vectors for uniform and non-uniform resolution, respectively. They are classified using MLNNs and the ECOC framework. This approach does not employ the CSP filters unlike the proposed method. Therefore, we can confirm the efficiency of using the CSP filters. Three types of codeword to the ECOC framework, i.e., OPC, Hadamard, and E-ECOC, are employed. MLNNs are used for classifiers.

CSP-ECOC (Proposed BCI System): The details have been described in Sec. 4. This method provides 25- and 30-dimemsional feature vectors with 4 CSP filters for uniform and nonuniform resolution, respectively. Coding is also varied to OPC, Hadamard, and E-ECOC.

Examples of the dimension of the feature vectors, which are the MLNN input data, are shown in Table 5. Some examples of MLNN input patterns for Subject 1, trial 1 and all mental tasks are shown in Figs. 3-7. In these figures, differences between the MLNN input patterns generated by various methods can be recognized. In Fig. 6, the EEGs are projected by the CSP filters, which are optimized to categorize three mental tasks, {*Baseline, Letter composition* and *Counting*} = 1, against the other mental tasks, {*Multiplication* and *Rotation*} = 0. In the case of Fig. 7, different CSP filters, which were optimized to categorize {*Baseline*, *Letter composition* and *Rotation*} = 1 against {*Multiplication* and *Counting*} = 0, were used.

5.3 Classification

MLNNs are used to perform both multiclass and binary classification modes. MLNN is optimized by the error back-propagation algorithm. Parameters for MLNN are determined and are shown in Tables 6 and 7.

Table 6 MLNN parameters for conventional method, CSP-OVR and CSP-SIM

Parameter	No Segmentation	Segmentation
Input node	Refer to Table 4	Refer to Table 4
Output node	5	5
Hidden node	20	20
Iteration	100000	5000
Learning Rate	0.01	0.1
Activation Function	Tanh – Logistic Sigmoid	Tanh – Logistic Sigmoid
Initial Weight Range	±0.1	±0.1
Threshold of Rejection	0.6	0.6
Generalization method: random noise	±0.05 - ±0.1	±0.05 - ±0.1

Table 7 MLNN parameters for conventional-ECOC and CSP-ECOC

Parameter	No Segmentation	Segmentation
Input node	Refer to Table 4.	Refer to Table 4.
Output node	1	1
Hidden node	10	10
Iteration	80000	5000
Learning Rate	0.01	0.1
Activation Function	Tanh – Logistic Sigmoid	Tanh – Logistic Sigmoid
Initial Weight Range	±0.1	±0.1
Threshold of Rejection: OPC Hadamard E-ECOC	 1.2 2.0 4.0	 1.2 2.0 4.0
Generalization method: random noise	±0.05 - ±0.1	±0.05 - ±0.1

5.4 Classification performance evaluation and validation

To evaluate the classification performance, the correct classification rate (P_c), error classification rate (P_e) and rate of correct and error classifications (R_c) are used.

$$P_c = \frac{N_c}{N_t} \times 100\% \qquad (18)$$

$$P_e = \frac{N_e}{N_t} \times 100\% \qquad (19)$$

$$R_c = \frac{N_c}{N_c + N_e} \qquad (20)$$

$$N_t = N_c + N_e + N_r \qquad (21)$$

N_c, N_e, and N_r are the numbers of correct classifications, error classifications, and rejections, respectively.

These experiments involve subject-specific classification. The EEG data are individually applied to the BCI system subject by subject. Each subject's EEG data consists of 5 trials in one experimental session. Each trial comprises 5 mental tasks hence, a total of 25 EEG data sets are obtained for one session. Four trials and 1 trial in each session are selected for training and testing data sets, respectively. Since Subjects 1 and 6 performed 2 sessions and Subjects 2 and 7 performed 1 session of EEG recording, there are 8 training trials and 2 testing trials for Subjects 1 and 6. On the other hand, there are 4 training trials and 1 testing trial for Subjects 2 and 7.

The experiments are validated by 5-fold cross-validation. The experiments of mental task classification are independently carried out 5 times by changing the combination of training and testing data sets. The classification performances are evaluated in each independent experiment. As a result, 5 independent classification results are obtained. They are averaged to evaluate the final classification performance.

6. Experimental Results and Discussions

The average, maximum and minimum values of the classification results are listed in Tables 8 and 9. In Conventional-ECOC and CSP-ECOC, the E-ECOC, which requires 15 elemental classifiers, is selected for the highest generalization.

6.1 Overall results

The classification results for nonsegmented data are demonstrated in Tables 8 (a), (b), (c), and (d) for Subjects 1, 2, 6 and 7, respectively.

For Subject 1's EEG data set, the conventional method [18], [19] works well; however, the proposed CSP-ECOC method can improve the correct classification rate P_c from 88% up to 92% for the uniform frequency resolution, and up to 90% for the nonuniform frequency resolution. The other methods provide roughly the same correct and error classification rates as the conventional method.

The CSP filter works well on Subject 2's EEG data set. The correct classification rates P_c are obviously increased by using any CSP filters. Moreover, the error classification rates P_e can also be decreased to 0% by CSP-OVR and CSP-SIM methods. The greatest improvement is provided by the proposed CSP-ECOC method with the nonuniform frequency resolution.

For Subject 6's data set, the results also indicate that the classification performances are obviously upgraded. The correct classification rates P_c are increased from 66% to 84%. In addition, the error classification rates P_e are also significantly decreased from 26% to 6% by the CSP-ECOC method with the nonuniform frequency resolution.

Table 8 Overall experimental results for nonsegmented data (maximum/minimum values are given in parentheses)

(a) Subject 1

Method	Uniform			Non-uniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	88.00 (90.00/86.00)	4.00 (6.00/2.00)	0.96 (0.98/0.94)	88.00 (90.00/86.00)	6.00 (6.00/6.00)	0.93 (0.94/0.93)
CSP-OVR	88.00 (88.00/88.00)	6.00 (8.00/4.00)	0.94 (0.96/0.92)	84.00 (88.00/82.00)	6.00 (10.00/0.00)	0.93 (1.00/0.89)
CSP-SIM	88.00 (90.00/86.00)	8.00 (12.00/4.00)	0.92 (0.96/0.88)	84.00 (86.00/82.00)	8.60 (10.00/8.00)	0.90 (0.89/0.91)
CON-ECOC	90.00 (92.00/88.00)	10.00 (14.00/8.00)	0.90 (0.92/0.86)	90.00 (92.00/88.00)	10.00 (14.00/8.00)	0.90 (0.92/0.86)
CSP-ECOC	**92.00 (92.00/90.0)**	**4.00 (6.00/0.00)**	**0.96 (1.00/0.94)**	90.00 (94.00/86.00)	6.00 (10.00/2.00)	0.94 (0.98/0.90)

(b) Subject 2

Method	Uniform			Non-uniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	80.00 (84.00/76.00)	4.00 (8.00/0.00)	0.95 (1.00/0.90)	84.00 (88.00/80.00)	0.00 (0.00/0.00)	1.00 (1.00/1.00)
CSP-OVR	86.70 (88.00/84.00)	0.00 (0.00/0.00)	1.00 (1.00/1.00)	88.00 (90.00/84.00)	0.00 (0.00/0.00)	1.00 (1.00/1.00)
CSP-SIM	88.00 (88.00/88.00)	0.00 (0.00/0.00)	1.00 (1.00/1.00)	92.00 (92.00/92.00)	0.00 (0.00/0.00)	1.00 (1.00/1.00)
CON-ECOC	88.00 (90.00/86.00)	4.00 (4.00/4.00)	0.96 (0.96/0.96)	88.00 (90.00/86.00)	4.00 (8.00/2.00)	0.96 (0.98/0.92)
CSP-ECOC	88.00 (90.00/86.00)	8.00 (10.00/4.00)	0.92 (0.96/0.90)	**96.00 (96.00/96.00)**	**4.00 (4.00/4.00)**	**0.96 (0.96/0.96)**

(c) Subject 6

Method	Uniform			Non-uniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	66.00 (70.00/60.00)	26.00 (30.00/24.00)	0.72 (0.74/0.67)	70.00 (74.00/66.00)	8.00 (10.00/6.00)	0.90 (0.92/0.88)
CSP-OVR	75.33 (76.00/74.00)	17.30 (20.00/14.00)	0.81 (0.84/0.76)	82.00 (84.00/80.00)	8.00 (10.00/6.00)	0.91 (0.93/089)
CSP-SIM	68.67 (70.00/66.00)	15.33 (18.00/14.00)	0.82 (0.83/0.79)	76.00 (78.00/72.00)	14.00 (16.00/12.00)	0.84 (0.87/0.82)
CON-ECOC	68.00 (72.00/64.00)	22.00 (24.00/20.00)	0.76 (0.77/0.74)	70.00 (74.00/68.00)	8.00 (10.00/6.00)	0.90 (0.92/0.88)
CSP-ECOC	64.00 (68.00/60.00)	22.00 (26.00/18.00)	0.74 (0.78/0.71)	**84.00 (86.00/82.00)**	**6.00 (8.00/4.00)**	**0.93 (0.95/0.91)**

(d) Subject 7

Method	Uniform			Non-uniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	80.00 (84.0076.00)	16.00 (20.00/12.00)	0.83 (0.87/0.79)	81.60 (84.00/78.00)	12.00 (20.00/12.00)	0.87 (0.87/0.79)
CSP-OVR	72.80 (76.00/68.00)	20.00 (24.00/16.00)	0.78 (0.83/0.74)	81.60 (84.00/80.00)	12.00 (20.00/8.00)	0.87 (0.91/0.80)
CSP-SIM	73.60 (76.00/72.00)	21.60 (24.00/20.00)	0.77 (0.79/0.75)	72.80 (76.00/68.00)	16.80 (20.00/16.00)	0.81 (0.83/0.79)
CON-ECOC	80.00 (84.00/76.00)	18.40 (20.0016.00)	0.81 (0.84/0.79)	81.60 (84.00/76.00)	15.20 (16.00/12.00)	0.84 (0.87/0.83)
CSP-ECOC	81.60 (84.00/80.00)	12.80 (16.00/12.00)	0.86 (0.88/0.83)	**84.00 (84.00/84.00)**	**12.00 (12.00/12.00)**	**0.88 (0.88/0.88)**

CON = Conventional Method, CON-ECOC = Conventional-ECOC

In the case of Subject 7's data set, the classification performance is increased only by the proposed CSP-ECOC method. The correct classification rate P_c is increased to 84% with the nonuniform frequency resolution.

The classification results for the segmented EEG data are shown in Table 9. In this case, we can summarize that the CSP approaches can enhance the classification performance. For instance, in the case of Subject 1, the proposed CSP-ECOC method can improve the correct classification rate from 82.1% to 86.7% with the nonuniform frequency resolution. In the case of Subject 2, the correct classification rate can also be enhanced from 71.5% to 83.74%. As well, in the case of Subject 6, the accuracy rate is upgraded from 47% to 66.97%. Lastly, in the case of Subject 7, P_c can be refined from 68.55% to 73.19%. Most CSP methods can improve the performance, except for the CSP-SIM method in the case of Subjects 1 and 7. P_c is a slightly degraded to around 2%.

The reason for the improvement may be some benefits of CSP spatial filtering that cause increases in discriminated information on input patterns.

Table 9 Overall experimental results for segmented data (maximum/minimum values are given in parentheses)

(a) Subject 1

Method	Uniform			Nonuniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	82.10 (83.30/80.90)	12.30 (14.30/10.30)	0.87 (0.89/0.85)	79.69 (80.40/78.90)	13.38 (14.70/11.40)	0.86 (0.89/0.85)
CSP-OVR	86.70 (86.90/86.40)	8.80 (8.90/8.70)	0.91 (0.91/0.91)	84.31 (85.6/82.3)	11.59 (13.20/9.79)	0.88 (0.90/0.87)
CSP-SIM	80.50 (81.20/80.20)	12.10 (12.70/11.60)	0.87 (0.87/0.87)	77.15 (78.05/75.65)	18.87 (19.15/18.57)	0.81 (0.81/0.81)
CON-ECOC	81.50 (81.90/81.10)	9.90 (10.30/9.50)	0.89 (0.89/0.89)	79.25 (79.50/79.00)	11.25 (12.80/10.40)	0.88 (0.88/0.86)
CSP-ECOC	**87.20 (87.75/86.65)**	**9.60 (10.60/8.60)**	**0.90 (0.91/0.89)**	86.70 (87.50/86.03)	10.60 (11.27/9.80)	0.89 (0.90/0.88)

(b) Subject 2

Method	Uniform			Nonuniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	71.50 (72.10/70.70)	17.60 (18.20/16.80)	0.80 (0.80/0.80)	75.69 (76.02/74.98)	15.86 (16.50/15.02)	0.83 (0.83/0.82)
CSP-OVR	78.00 (78.20/77.80)	15.60 (16.20/14.00)	0.83 (0.84/0.83)	78.68 (80.31/74.45)	15.29 (15.59/14.89)	0.84 (0.84/0.83)
CSP-SIM	74.70 (76.10/72.60)	17.30 (18.80/15.90)	0.81 (0.83/0.79)	77.79 (78.40/77.36)	18.56 (19.40/17.92)	0.81 (0.81/0.80)
CON-ECOC	78.20 (78.80/77.40)	20.10 (20.70/19.30)	0.80 (0.80/0.79)	78.50 (79.05/77.73)	16.70 (17.00/16.30)	0.82 (0.83/0.82)
CSP-ECOC	81.60 (82.60/80.60)	18.40 (19.40/17.40)	0.82 (0.83/0.81)	**83.74 (83.94/83.54)**	**16.21 (16.41/16.01)**	**0.84 (0.84/0.84)**

(c) Subject 6

Method	Uniform			Non-uniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	47.00 (47.80/46.40)	24.80 (25.10/24.40)	0.65 (0.66/0.65)	56.05 (56.9/55.12)	25.08 (26.06/24.23)	0.69 (0.70/0.68)
CSP-OVR	60.00 (61.10/59.00)	29.60 (30.90/28.40)	0.67 (0.68/0.65)	59.23 (60.03/58.56)	26.41 (27.31/25.71)	0.69 (0.69/0.69)
CSP-SIM	55.60 (56.00/55.00)	30.20 (30.60/29.60)	0.65 (0.65/0.64)	58.15 (59.08/57.30)	27.90 (28.57/27.10)	0.68 (0.68/0.67)
CON-ECOC	58.70 (59.50/58.00)	29.20 (30.10/28.50)	0.67 (0.67/0.66)	58.72 (59.75/57.77)	28.77 (29.54/27.87)	0.67 (0.67/0.67)
CSP-ECOC	63.70 (64.10/63.40)	30.70 (31.40/29.80)	0.67 (0.68/0.67)	**66.97 (68.20/65.34)**	**31.72 (33.60/30.00)**	**0.68 (0.69/0.66)**

(d) Subject 7

Method	Uniform			Nonuniform		
	P_c	P_e	R_c	P_c	P_e	R_c
CON	68.55 (69.60/66.80)	17.01 (18.30/15.90)	0.79 (0.81/0.78)	68.55 (69.60/66.80)	16.38 (18.00/15.00)	0.81 (0.82/0.79)
CSP-OVR	69.62 (71.60/67.50)	21.11 (23.10/17.20)	0.77 (0.80/0.75)	68.87 (70.00/68.20)	21.26 (24.80/18.30)	0.76 (0.79/0.74)
CSP-SIM	67.77 (71.18/64.51)	14.79 (21.64/10.67)	0.82 (0.87/0.77)	66.21 (68.10/64.82)	19.88 (20.21/19.28)	0.77 (0.78/0.76)
CON-ECOC	70.42 (71.00/69.90)	19.81 (20.02/19.40)	0.78 (0.78/0.78)	71.58 (72.00/71.20)	19.53 (20.30/18.7)	0.79 (0.79/0.78)
CSP-ECOC	72.29 (72.70/72.00)	16.49 (16.77/16.20)	0.81 (0.82/0.81)	**73.19 (74.67/71.28)**	**15.70 (16.50/15.00)**	**0.82 (0.83/0.82)**

CON = Conventional Method, CON-ECOC = Conventional-ECOC

6.2 Uniform versus nonuniform frequency resolution

Classification performances based on the two kinds of the sample reduction methods, that is, the uniform and nonuniform frequency resolution, are also listed in Tables 8 and 9, on the left and the right side, respectively. In most cases of nonsegmented data, the error classify-cation rates are depressed in the nonuniform frequency resolution mode. For examples, explicitly, in the case of Subject 6's nonsegmented data, not only are the error classification rates substantially decreased from 15.33%~26.00% to 6.00%~14.00%, but the correct classification rates are also increased from 66.00%~68.67% to 70.00%~84.00%. In the case of Subjects 2 and 7, the results also show the same trend as the case of Subject 6.

However, in the case of Subject 1, the classification performance cannot be improved by using nonuniform frequency resolution. Improvement upon using the nonuniform frequency resolution is dependent on noise in the high-frequency bands. The nonuniform frequency resolution method, which compresses data in the high-frequency bands more than in the low-frequency bands, can suppress the contaminated noise well in the high-frequency region. In the case of Subject 1, we assume that initially, EEGs are fairly clean, so an excellent classification performance can be obtained, no matter which method of sampling reduction is applied.

Unfortunately, the nonuniform frequency resolution for the segmented data, significant improvement cannot be obtained. The correct classification rates are insignificantly improved, but the error classification rates are also degraded.

6.3 Non-spatial filtering ECOC versus CSP filtering ECOC

In this paper, we also compare results between nonspatial filtering and CSP filtering when both are applied to the ECOC framework in order to prove that the improvement does not inherently come from the ECOC techniques, but also is a result of the benefits of CSP filtering. The experimental results shown in Tables 8 and 9 indicate that although both methods generally provide good performance, the CSP-ECOC method can provide more improved performance. The accuracy rates are improved from 58.70%~90.00% to 63.70%~96.00%.

From these experimental results, we can confirm the efficiency of using the CSP filtering.

Table 10 Classification results for different coding method for Subjects 1, 2, 6, and 7
(nonsegmented data /nonuniform frequency resolution)

(a) Conventional-ECOC

Coding	Subject 1			Subject 2		
	P_c	P_e	R_c	P_c	P_e	R_c
E-ECOC	90.00 (92.00/88.00)	10.00 (14.00/8.00)	0.90 (0.98/0.90)	88.00 (90.00/86.00)	4.00 (8.00/2.00)	0.96 (0.98/0.92)
Hadamard	88.00 (88.00/88.00)	12.00 (12.00/12.00)	0.88 (0.88/0.88)	84.00 (84.00/84.00)	16.00 (16.00/16.00)	0.84 (0.84/0.84)
OPC	81.20 (82.00/80.00)	18.80 (20.00/18.00)	0.81 (0.82/0.80)	83.60 (88.00/78.00)	15.20 (18.00/12.00)	0.85 (0.88/0.81)
Coding	**Subject 6**			**Subject 7**		
	P_c	P_e	R_c	P_c	P_e	R_c
E-ECOC	70.00 (74.00/68.00)	8.00 (10.00/6.00)	0.90 (0.92/0.88)	81.60 (84.00/76.00)	15.20 (16.00/12.00)	0.84 (0.87/0.83)
Hadamard	68.80 (72.00/66.00)	8.00 (12.00/6.00)	0.90 (0.92/0.85)	81.60 (84.00/76.00)	15.20 (16.00/12.00)	0.84 (0.87/0.83)
OPC	68.40 (72.00/66.00)	9.20 (14.00/8.00)	0.88 (0.9/0.84)	80.00 (88.00/76.00)	16.00 (20.00/12.00)	0.83 (0.88/0.79)

(b) CSP-ECOC

Coding	Subject 1			Subject 2		
	P_c	P_e	R_c	P_c	P_e	R_c
E-ECOC	90.00 (92.00/88.00)	6.00 (10.00/2.00)	0.94 (0.98/0.90)	96.00 (96.00/96.00)	4.00 (4.00/4.00)	0.96 (0.96/0.96)
Hadamard	85.00 (90.00/80.00)	6.00 (12/6)	0.93 (0.92/0.85)	90.00 (92.00/88.00)	6.00 (8.00/4.00)	0.94 (0.96/0.92)
OPC	84.00 (88.00/80.00)	4.00 (6.00/2.00)	0.95 (0.98/0.93)	84.00 (88.00/80.00)	6.00 (8.00/4.00)	0.93 (0.95/0.92)
Coding	**Subject 6**			**Subject 7**		
	P_c	P_e	R_c	P_c	P_e	R_c
E-ECOC	84.00 (86.00/82.00)	6.00 (8.00/4.00)	0.93 (0.95/0.91)	84.00 (84.00/84.00)	12.00 (12.00/12.00)	0.88 (0.88/0.88)
Hadamard	78.00 (80.00/76.00)	6.00 (8.00/4.00)	0.93 (0.95/0.91)	84.00 (84.00/84.00)	16.00 (16.00/16.00)	0.84 (0.84/0.84)
OPC	78.00 (80.00/76.00)	10.00 (8.00/4.00)	0.89 (0.95/0.91)	78.40 (80.00/76.00)	21.60 (24.00/20.00)	0.78 (0.80/0.76)

6.4 Comparison on coding in ECOC

The performance of ECOC depends on the length of the applied codewords, which is the same as the number of classifiers in the ensemble estimation. This is still an open issue for the ECOC framework in our study. In the experiments, 3 types of standard coding methods are performed and compared, i.e., OPC, Hadamard coding, and E-ECOC, in which 5, 7, and 15 classifiers are required, respectively. The theoretical performance of the correcting code has been mentioned in Sec. 3.

Table 10 shows results of using different coding methods with nonuniform sampling reduction and nonsegmented data. For the other modes, the results show the same trends as well. The E-ECOC with 15 classifiers overcomes both Hadamard coding with 7 classifiers and OPC with 5 classifiers because the increased length in the codewords can achieve a large capacity of error correction.

The increase in the codeword length in ECOC requires more classifiers and a long computation time in the learning process. The appropriateness of the coding method is dependent on the applications and computational tools.

The simulation results indicate that applying CSP filtering can improve the accuracy of the three coding methods.

7. Conclusions

We proposed a new BCI system, which combines CSP filtering and the ECOC framework to realize multiclass classification with an error-correcting capability. Also, a new sampling reduction method, that is, the nonuniform frequency resolution, was also proposed. The important information in the lower frequency region was precisely analyzed, and the contaminating noise in the high-frequency region was well suppressed. The classification performance of the proposed BCI system was compared with those of many conventional and related BCI methods. Furthermore, the effectiveness of using CSP filtering was examined, and several coding methods were compared. In order to confirm the general usefulness, the EEG data of 7 subjects, which are available on the web site of Colorado State University, were used. The experimental results for 4 subjects show that the proposed BCI system can boost the classification rate from 66%~88% to 84%~96% and suppress the error classification rates from 4%~26% to 4%~12%. Almost the same results were obtained for the other 3 subjects.

References

[1] G. Dornhege, J. d. r. Millán, T. Hinterberger, D. J. McFanland and K. R. Müller: Toward Brain-Computer Interfacing, The MIT Press, 2007.

[2] Z. A. Keirn: Alternative Modes of Communication between Man and Machine, Masters Thesis in Electrical Engineering, Purdue University, Dec., 1988.

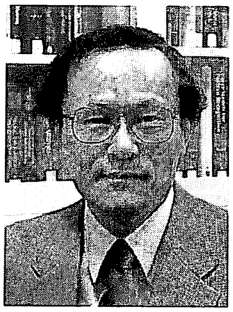
[3] Z. A. Keirn and J. I. Aunon: A new mode of communication between man and his surroundings, IEEE Trans. Biomed. Eng., Vol. 37, pp. 1209-1214, 1990.

[4] J. Muller-Gerking, G. Pfurtscheller and H. Flyvbjerg: Designing optimal spatial filters for single-trial EEG classification in a movement task, Clin. Neurophysiol., Vol. 110, pp. 787-798, 1999.

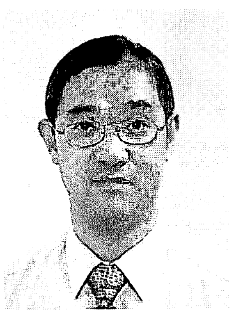
[5] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller: Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Trans. Rehabil.

Eng., Vol. 8, No. 4, pp. 441-446, 2000.
[6] G. Dornhege, B. Blankertz and G. Curio: Speeding up classification of multi-channel brain-computer interfaces: Common spatial patterns for slow cortical potentials, Proc. First Int. IEEE EMBS Conf., Neural Eng., pp. 595-598, 2003.
[7] G. Dornhege, B. Blankertz, G. Curio and K. R. Müller: Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms, IEEE Trans. Biomed. Eng., Vol. 51, No. 6, pp. 993-1002, 2004.
[8] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K. R. Müller: Optimizing spatial filters for robust EEG single-trail analysis, IEEE Signal Processing Magazine, Vol. 43, 2008.
[9] M. Grosse-Wentrup and M. Buss: Multi-class common spatial patterns and information theoretic feature extraction, IEEE Trans. Biomed. Eng., Vol. 55, No. 8, pp. 1991-2000, 2008.
[10] T. Y.an, T. Jingtian and G. Andong: Multi-class EEG classification for brain computer interfaces based on CSP, Int. IEEE BMEI Conf., Vol. 2, pp. 469-472, 2008.
[11] K. Fukunaga: Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, 1990.
[12] R. O. Duda, P. E. Hart and D. G. Stork: Pattern Classification, 2nd ed., Wiley-Interscience, 2000.
[13] A. Hyvärinen and E. Oja: Independent component analysis: Algorithms and applications, Neural Networks, Vol. 13, Nos. 4-5, pp. 411-430, 2000.
[14] C. A. Joyce, I. F. Gorodnitsky and M. Kutas: Automatic removal of eye movement and blink artifacts from EEG data using blind component separation, Psychophysiology, Vol. 41, No. 2, pp. 313-325, 2004.
[15] M. Crespo-Garcia, M. Mtienza and J. L. Cantero: Muscle artifact removal from human sleep EEG by using independent component analysis, Annals of Biomed. Eng., Vol. 36, No. 3, pp. 467-475, 2008.
[16] R. M. J. David: Brain Computer Interface Based on ICA and Neural Network, Masters Thesis, Kanazawa University, 2007.
[17] K. Nakayama, H. Horita and A. Hirano: Neural network based BCI by using orthogonal components of multi-channel brain waves and generalization, Lecture Notes in Computer Science, Vol. 5164, pp. 879-888, 2008.
[18] K. Nakayama and K. Inagaki: A brain computer interface based on neural network with efficient pre-processing, Proc. IEEE, ISPACS, pp. 673-676, 2006.
[19] K. Nakayama, Y. Kaneda and A. Hirano: A brain computer interface based on FFT and multilayer neural network-feature extraction and generalization, Proc. IEEE, ISPACS, pp. 101-104, 2007.
[20] EEG Pattern Analysis: http://www.cs.colostate.edu/eeg/
[21] R. Palaniappan: Brain computer interface design using band powers extracted during mental tasks, Proc. 2nd Int. IEEE EMBS Conf. on Neural Eng., Arlington, Virginia, pp. 312-324, Mar. 16-19, 2005.
[22] R. Palaniappan: Utilizing gamma band to improve mental task based brain-computer interface design, IEEE Trans. Neural Syst. Rehabil. Eng., Vol. 14, pp. 299-303, 2006.
[23] L. Zhang, W. He, C. He and P. Wang: Improving mental task classification by adding high frequency band information, J. Medical Systems, Vol. 34, No. 1, pp. 51-60, 2008.
[24] B. Graimann, J. E. Higgins, S. P. Levine and G. Pfurtscheller: Toward a direct brain interface based on human subdural recording and wavelet-packet analysis, IEEE Trans. Biomed. Eng., Vol. 51, pp. 954-962, 2004.
[25] F. Lotte, M. Congedo, A. Lécuyer, F Lamarche and B Arnaldi: A review of classification algorithms for EEG-based brain-computer interfaces, J. Neural Eng., Vol. 4, pp. R1-R13, 2007.
[26] T. G. Dietterich and G. Bakiri: Solving multiclass learning problems via error-correcting output codes, J. Artificial Intelligence Research, Vol. 2, pp. 263-286, 1995.
[27] E. L. Allweinnn, R. E. Schapire and Y. Singer: Reducing multiclass to binary: A unifying approach for margin classifiers, J. Machine Learning Research, Vol. 1, pp. 113-141, 2000.
[28] T. Windeatt and R. Ghaderi: Coding and decoding strategies for multi-class learning problems, Information Fusion, Vol. 4, pp. 11-21, 2003.
[29] S. Escalera, O. Puljol and P. Radeva: On the decoding process in ternary error-correcting output codes, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 32, No. 1, pp. 120-134, 2010.
[30] A. Zhang, Z. Wu, C. Li and K. Fang: On Hadamard-type output coding in multiclass learning, Lecture Notes, Comp. Sci. Vol. 2690, pp. 397-404, 2003.
[31] F. J. MacWillians and N. J. A. Sloane: The Theory of Error-Correcting Codes, Elsevier Science Publishers, 1977.
[32] S. P. Fitzgibbon, K. J. Pope, L. Mackenzie, C. R. Clark and J. O. Willoughby: Cognitive task augment gamma EEG power, Clin. Neurophysiol., Vol. 115, pp. 1802-1809, 2004.

Prapong Prechaprapranwong was born in Bangkok, Thailand in 1977. He received his B.E. and M.E. degrees in Computer Engineering from King Mongkut's University of Technology Thonburi (KMUTT). He is currently working towards his Ph.D. at Kanazawa University, under the supervision of Prof. Kenji Nakayama and Assist. Prof. Akihiro Hirano. His research interests include neural networks and brain-computer interfaces.



Kenji Nakayama received his B.E. and Dr. degrees in electronics engineering from the Tokyo Institute of Technology (TIT), Tokyo, Japan, in 1971 and 1983, respectively. From 1971 to 1972 he was engaged in research on classical network theory in TIT. He was affiliated with NEC Corporation from 1972 to 1988, where his research subjects were filter design methodology and digital signal processing. He joined the Department of Electrical and Computer Engineering at Kanazawa University in Aug 1988. He is currently a Professor in the Graduate School of Natural Science and Technology. He served as the Director of the International Student Center from April 2003 to March 2005. His current research interests include adaptive signal processing and neural networks. He is a senior member of IEEE and a member of INNS and RISP, Japan.



Akihiro Hirano received his B. Eng. M. Eng. and Dr. Eng. degrees from Kanazawa University, Kanazawa, Japan, in 1987, 1989 and 2000, respectively. He was affiliated with NEC Corporation, Kawasaki, Japan, from 1989 to 1998, where he has been a Research Engineer in the Research and Development Group. He joined the Faculty of Engineering, Kanazawa University in 1998. He is currently an Assistant Professor in the Division of Electrical Engineering and Computer Science, Graduate School of Natural Science and Technology, Kanazawa University. He has been engaged in research on adaptive signal processing and neural networks. He was awarded the 1995 Academic Encouragement Award by IEICE. Dr. Hirano is a member of IEEE.

(Received October 4, 2010; revised December 5, 2010)