Annals of Biomedical Engineering, Vol. 33, No. 8, August 2005 (© 2005) pp. 1053–1070
DOI: 10.1007/s10439-005-5772-1

Recognition of Motor Imagery Electroencephalography Using Independent Component Analysis and Machine Classifiers

CHIH-I HUNG,[1,2] PO-LEI LEE,[2] YU-TE WU,[1,2,3] LI-FEN CHEN,[2,4] TZU-CHEN YEH,[2,5] and JEN-CHUEN HSIEH[2,3,5,6]

[1]Institute of Radiological Sciences, National Yang-Ming University, Taipei, Taiwan; [2]Laboratory of Integrated Brain Research, Department of Medical Research and Education, Taipei Veterans General Hospital, Taipei, Taiwan; [3]Institute of Health Informatics and Decision Making, School of Medicine, National Yang-Ming University, Taipei, Taiwan; [4]Center for Neuroscience, National Yang-Ming University, Taipei, Taiwan; [5]Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan; and [6]Institute of Neuroscience, School of Life Science, National Yang-Ming University, Taipei, Taiwan

(Received 19 April 2004; accepted 14 March 2005)

Abstract—Motor imagery electroencephalography (EEG), which embodies cortical potentials during mental simulation of left or right finger lifting tasks, can be used to provide neural input signals to activate a brain computer interface (BCI). The effectiveness of such an EEG-based BCI system relies on two indispensable components: distinguishable patterns of brain signals and accurate classifiers. This work aims to extract two reliable neural features, termed contralateral and ipsilateral rebound maps, by removing artifacts from motor imagery EEG based on independent component analysis (ICA), and to employ four classifiers to investigate the efficacy of rebound maps. Results demonstrate that, with the use of ICA, recognition rates for four classifiers (fisher linear discriminant (FLD), back-propagation neural network (BP-NN), radial-basis function neural network (RBF-NN), and support vector machine (SVM)) improved significantly, from 54%, 54%, 57% and 55% to 70.5%, 75.5%, 76.5% and 77.3%, respectively. In addition, the areas under the receiver operating characteristics (ROC) curve, which assess the quality of classification over a wide range of misclassification costs, also improved from .65, .60, .62, and .64 to .74, .76, .80 and .81, respectively.

Keywords—Brain computer interface (BCI), Rebound maps, Fisher linear discriminant (FLD), Back-propagation neural network (BP-NN), Radial-basis function neural network (RBF-NN), Support vector machine (SVM).

INTRODUCTION

In recent years, significant progress in neuroscience has inspired studies on the development of a brain computer interface (BCI),[10,22,29–31,34–36] a novel technique in assisting people to communicate with external environments or trigger surrounding devices by means of brain signals. These systems are particularly useful for individuals who are unable to produce motor activity, e.g., patients with amyotrophic lateral sclerosis or locked-in syndrome whose cognitive or sensor functions may be intact. Such patients can be trained to perform mental tasks in simulating right or left hand movements without any overt motor output. The feasibility of BCI systems counts on two important ingredients: distinguishable neural patterns and effective classifiers. This study aims to 1) extract a reliable featured pattern from each single-trial of the motor imagery EEG using independent component analysis and 2) employ various machine classifiers to investigate efficacy of the featured pattern.

Rhythmic activities elicited from an imagined hand movement and from an actual hand movement in the primary sensorimotor areas present similar EEG patterns.[26,28] Performing a specific or an imagined finger movement composes of three phases: planning, execution and recovery. The planning and execution phases can be viewed as an EEG correlate of an activated cortical motor network, resulting in alpha and lower beta bands amplitude attenuation, namely, event-related desynchronization (ERD).[29] The recovery phase may reflect deactivation/inhibition in the underlying cortical network, producing focal mu and amplitude enhancement, namely, event-related synchronization (ERS).[26,28]

Motor imagery right and left movement can be discriminated by identifying the reactivity patterns of ERD, ERS or both in the alpha (8–13 Hz) and beta (18–28 Hz) bands.[29,30] Since the most relevant ERD/ERS frequency bands can be substantially different between subjects, Pfurtscheller *et al*. have applied distinctive sensitive learning vector quantization to determine subject-specific frequency components. The reactivity patterns and associated frequency bands of imagery for left and right hand movements for four subjects in their study were beta ERD (9–13 Hz and 21–26 Hz) for subject 1, beta ERD (10–12 Hz and 21–23 Hz) for subject

Address correspondence to Yu-Te Wu, Ph.D., Institute of Radiological Sciences, National Yang-Ming University, No. 155, Li-Nong Street, Section 2, Pei-Tou, Taipei, 112, Taiwan, ROC. Electronic mail: ytwu@ym.edu.tw

2, beta ERD and beta ERS (18–26 Hz) for subject 3 and beta ERS (13–19 Hz) for subject 4.[29] On-line error was between 10.0 and 38.1% when the learning vector quantization method was applied to imagery movement in driving a cursor to a target on a computer screen. Using an adaptive autoregressive model and linear discrimination analysis to analyze the ERD/ERS signal off-line, error was improved to between 5.8 and 32.8%. Muller-Gerking *et al.* applied a common spatial filter to detect actual (not imagined) left and right hand or right foot movements in single trial, reporting 84%, 90% and 94% accuracies for three subjects, respectively.[22] Wolpaw *et al.* have demonstrated that most subjects (i.e. about 80%) were able to manage their 8–12 Hz EEG activity in controlling cursor movements on a computer screen after four to six practice sessions over 3 weeks. In a study from 1994 to 2000, Wolpaw *et al.* used cursor control as a prototype BCI application for answering simple questions or spelling words, achieving an information transfer rate of about 20–25 bits/min.[34,35,36]

ICA, a blind source separation method, has been used extensively to remove artifacts for the analysis of EEG.[1,2,14–16,20–21,25,32] Makeig *et al.* have applied ICA to EEG data for source identification and localization,[20] and on the 29-channel data from a visual selective attention task to demonstrate that extracted μ-components exhibited much stronger motor-response related spectral reactivity than that measured at optimally placed single scalp channels.[21] Tang *et al.* showed that blind source separation can find neurophysiolgically and neuroanatomically meaningful neuronal components without assuming any prior physic models.[32] ICA was also used to separate temporally distinct but spatially overlapping time-locked, phase-locked, EEG evoked brain with distinct relationships to task events,[15,16] or used to extract and analyze of oscillatory time-locked, but non-phase-locked ERS activities for the study of subtle brain dynamics.[18] Recently blind source separation has been formulated to eigenvalue decomposition which provides a rather general solution to verify the underlying statistical assumption.[25] Delorme *et al.* used ICA to decompose 64-channel EEG data from a highly trained subject when he regulated EEG (12 Hz) to move a screen cursor upward or downward to hit target boxes. Independent components with large variances have been selected and further analyzed using the time-frequency technique. They found that one of the components displayed a right-hemisphere μ-rhythm pattern with spectral peaks at 11 and 22 Hz, and some of components were interpreted to be independent posterior alpha activities with 5- and 8-Hz spectral peaks in the vicinity of frontocentral cortex.[9] Both works have suggested that ICA can serve as a promising tool to better understand the EEG rhythms which can in turn optimize the performance of BCI.

In this study, EEG recordings of imagined right and left index finger lifting were analyzed and classified. We focused on the ERS, rather than ERD, since it appeared more persistently per single trial for each subject and has higher task and movement specificity.[27,28] Previous studies also showed an ERD phenomenon insufficiently prominent to make it an effective feature in the classification.[17] Because each EEG trial was inevitably contaminated by artifacts and oscillatory reactivity may express itself independently of artifacts,[19] we employed ICA to decompose each EEG epoch into a set of temporally independent components and spatial maps. These spatial maps represented the weights of the corresponding temporal component at each EEG sensor and can be categorized into task-related and task-unrelated groups, respectively. The spatial maps correlating highly to pre-defined movement-related templates were selected as task-related components, and corresponding temporal components were conjointly used for subsequent data reconstruction. After filtering, the reconstructed data for each single trial in a reactive frequency band, an envelope of task-related oscillation, can be extracted using the Hilbert transform which squares and sums real and imagery parts of the data.[3] The maximal amplitude of an envelope on the sensorimotor area during post movement is referred to as a rebound. Similar method has been developed and successfully applied on our previous single-trial magnetoencephalography measurements to remove artifacts and improve the signal-to-noise level of measured data, leading to better quantification of beta activity.[13,18,37–39] Since bilateral activation manifested in the vicinity of sensorimotor areas with stronger and earlier activity in the hemisphere contralateral to the side of imagined finger lifting,[26] we extracted two topographical maps constituted of values of envelopes across all channels at the two distinct instants at which the rebounds occurred in contralateral prior to ispilateral sensorimotor areas. The paired topographical maps, termed as contralateral and ispilateral rebound maps, were concatenated as a featured vector and fed into four classifiers, FLD, BP-NN, RBF-NN and SVM, for training and testing. By comparing with the performance of without-ICA rebound maps, the advantages of ICA in artifacts removal and the superiority of ICA-based rebound maps in classification will be demonstrated.

EXPERIMENTAL PARADIGM FOR MOTOR IMAGERY

Four right-handed healthy subjects (two males and two females), aged between 20 and 28, participated in this study. Each subject was naive to the experiment and watched a short video clip which displayed the actual finger lifting. He or she has practiced to conduct mental simulation of motor task for 20 min prior to the first session. During each session, subjects were asked to perform 100 trials of imagery right index finger lifting, followed by another 100 trials of imagery left index finger lifting. The length of each trial was 10 s. Each trial began with 1 s presentation of random noise during which subjects were allowed to blink

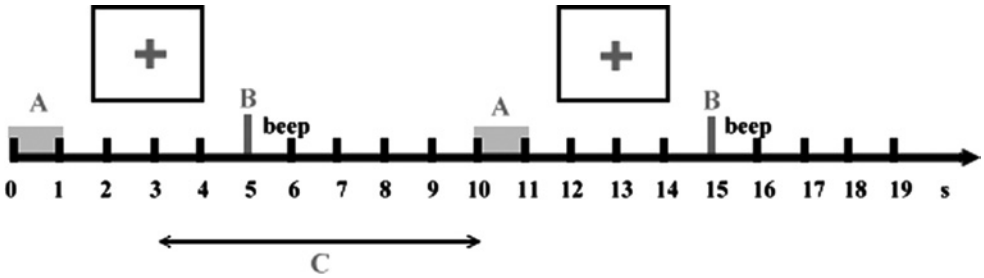

FIGURE 1. Timing of two consecutive 10-s trials of the motor imagery task. Each trial began with 1-s presentation of random noise during which subjects were allowed to blink his/her eyes (A). The subject was then instructed to stare at the fixation cross in the center of the monitor at 2 s and started to image right or left index finger lifting right after he/she heard an acoustic cue "beep" (with frequency 1 k Hz and 10 ms duration) at 5 s (B). Signals from 3 s to 10 s (C) in each trial (excluding bad epochs) were extracted to construct paired contralateral and ipsilateral rebound maps.

his/her eyes (Fig. 1A). The subject was then instructed to stare at the fixation cross in the center of the monitor at 2s and to begin to imagine right or left index finger lifting upon hearing an acoustic cue (frequency 1 k Hz, 10 ms duration) at 5s (Fig. 1B.). Inter-stimulus interval was 10 s.

The whole scalp of each subject was covered with 64 EEG electrodes (left and right ear references were used) placed onto anatomical locations according to the extended 10–20 international system. The configuration of the extended 10–20 system is shown in Fig. 2 in which C3 and

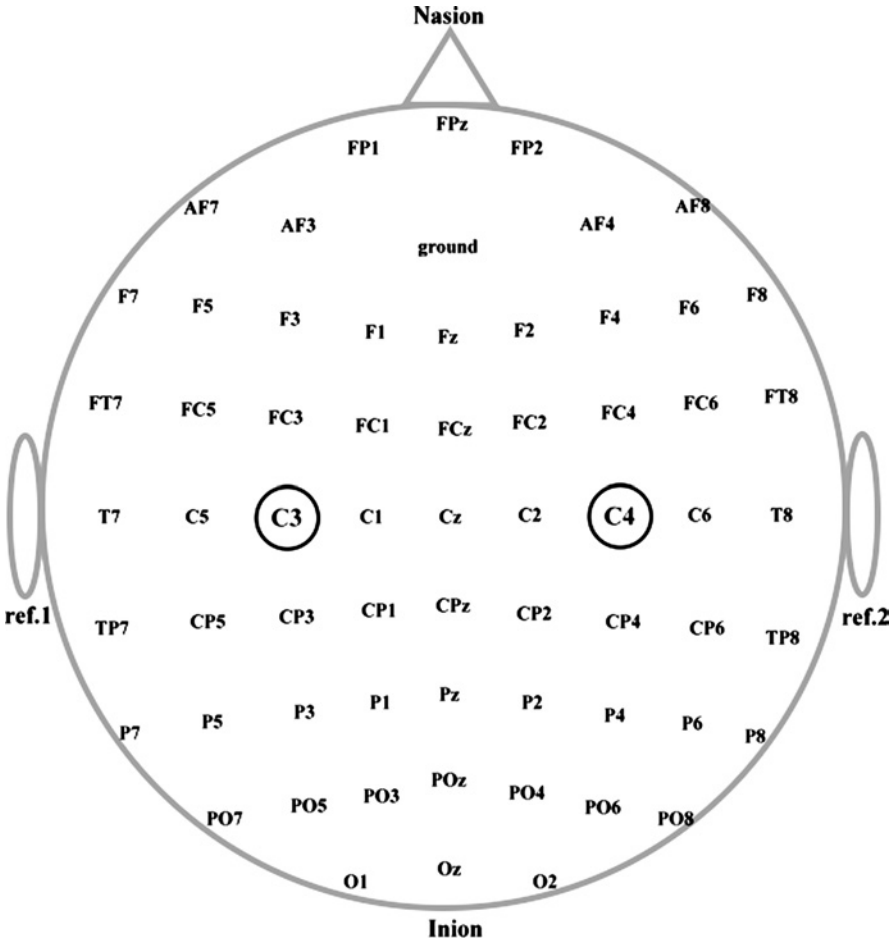

FIGURE 2. The configuration of the extended 10–20 system used in this study in which C3 and C4 corresponds to sensorimotor areas. The whole scalp of each subject was covered with 64 EEG electrodes placed onto anatomical locations according to this extended 10–20 international system.

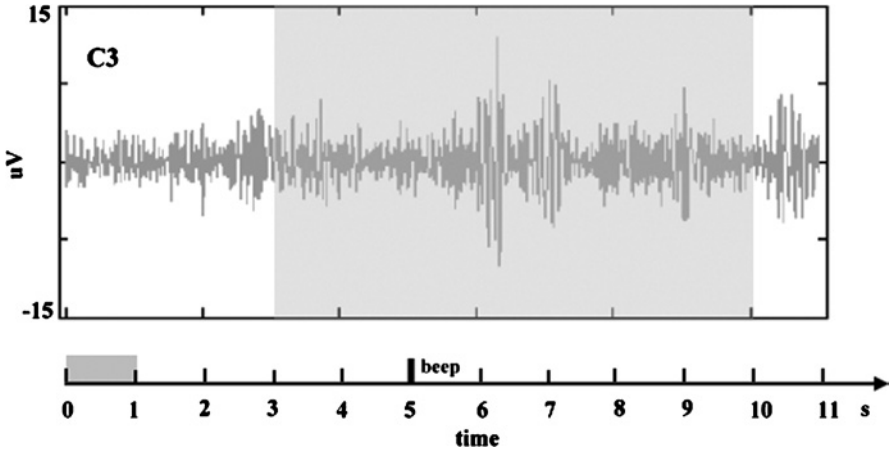

FIGURE 3. A pre-processed epoch recorded from sensorimotor area C3 when an imagery right index finger lifting task was performed. Signal in shaded area was segmented for subsequent processes.

C4 corresponds to sensorimotor areas. Vertical and horizontal electro-oculograms (VEOG and HEOG) were applied to reject bad epochs induced by eye blinking during the recording. The data were digitized at 250 Hz. Recorded signals were further bandpass-filtered at 6–50 Hz to remove dc drifts and 60 Hz noise. Throughout the recordings, surface electromyogram (EMG) monitoring was done from the m. extensor digitorum communis (digitized at 2 KHz) to detect motion status. Data were collected in four sessions for each subject. Signals from 3 s to 10 s (C in Fig. 1) in each trial (excluding bad epochs) were used for the subsequent feature extraction process. Figure 3 shows a pre-processed epoch from the sensorimotor area during an imagery right index finger lifting task (channel C3 in 10–20 system).

CONSTRUCTION OF CONTRALATERAL AND IPSILATERAL REBOUND MAPS WITH AND WITHOUT USING ICA

Pfurtscheller *et al.*[26,28] have reported that the beta reactivity mainly originates in the primary sensorimotor cortex and observed the similarity of circuitry involved in mental simulation of motor task and movement execution. Their movement-related ERS works have motivated us to construct the contralateral and ipsilateral rebound maps as salient features for the subsequent classification. The paired rebound maps have distinguished temporal, spatial and spectral characteristics since, after data were filtered in the optimally reactive frequency band, they were generated at the time points when the spatial maps exhibit peaked activities in the vicinity of sensorimotor area contralateral to the side of imagined finger lifting. An ICA-based method was employed to remove the artifacts from EEG motor imagery data and extract the paired rebound maps from the envelopes of reconstructed oscillatory activity. The extraction procedure is as follows:

Step 1: Signal Decomposition Using ICA

Each pre-processed epoch was arranged across m channels ($m = 62$) and n sampled points ($n = 1750$) into a $m \times n$ matrix X. The *ith* row contains the observed signal from *ith* EEG channel, and the *jth* column vector contains the observed samples at the *jth* time point across all channels. In the present study, all calculations were performed using the FastICA algorithm because of its fast convergence.[14] The FastICA technique first removes means of the row vectors in the X matrix and then uses a whitening procedure to transform the covariance matrix of the zero-mean data into an identity matrix. The whitening process was implemented using Principal Components Analysis.[14] Only the first N most significant eigenvectors ($N = 15$ in our analysis) were preserved in the subsequent ICA calculation. In the next step, FastICA searched for a matrix to further separate the whitened data into a set of components which were as mutually independent as possible. In combination with previous whitening process, the matrix X is transformed into a matrix S via an un-mixing matrix W, i.e.,

$$S = WX \tag{1}$$

so that rows of S are mutually independent. Each column in W^{-1}, the mixing matrix, represents a spatial map describing the relative projection weights of the corresponding temporal components at each EEG channel. These spatial maps will hereinafter be referred to as IC spatial maps. Figure 4 shows 12 IC spatial maps for 12 independent temporal components (not shown) decomposed from a single-trial imagery right hand movement. Maps IC2, IC3, IC5, IC7 and IC9 were highly related to the motor imagery task and categorized as task-related components, while the IC4 and IC6 maps were associated with the occipital alpha rhythm, and IC1 was noise from a bad channel.

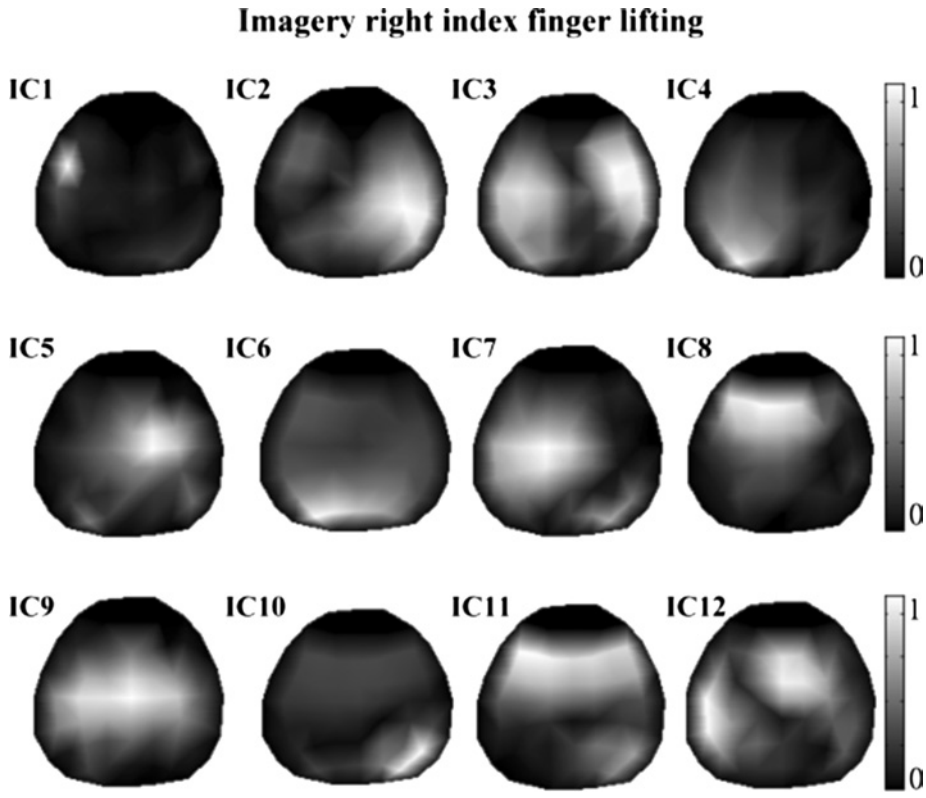

FIGURE 4. Decomposed IC spatial maps from a single EEG epoch of imagery right index finger lifting task. The maps IC2, IC3, IC5, IC7 and IC9 were highly related to motor task and categorized as task-related components, while the IC4 and IC6 maps were associated with the occipital alpha rhythm, and IC1 map was identified as noise emanated from a bad channel.

Step 2: Selection of Task-Related Components by Matching IC Spatial Maps to Pre-Defined Spatial Templates

Since the motor imagery task elicits bilateral activation in the vicinity of sensorimotor areas, four spatial patterns encompassing C3, C4, Cz/FCz (close to supplementary motor area) and both C3 and C4 areas, respectively, were used as spatial templates (see Fig. 5) to select task-related spatial maps. The spatial templates were generated prior to motor imagery experiments in the following manner. Subjects were asked to perform actual right and left finger lifting repeatedly, and each EEG epoch was decomposed using ICA. Movement-related spatial maps were identified manually and similar patterns were averaged to yield four spatial templates. Four spatial templates were used rather than a single template covering C3 and C4 because task-related activities can be separated by ICA and exhibited in multiple IC spatial maps. Each template was correlated with all the IC spatial maps of a single trial with the best two matches were selected. For example, spatial maps IC2, IC3, IC5, IC7 and IC9 in Fig. 4 were selected automatically due

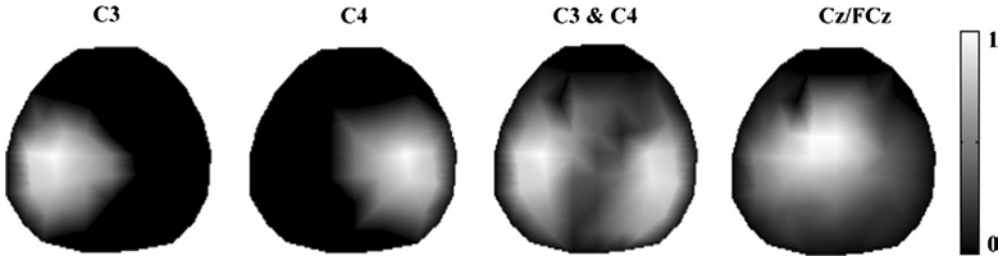

FIGURE 5. Spatial templates used to select task-related components from decomposed IC spatial maps. Prior to imagery motor experiments, subjects were asked to perform actual right and left finger lifting repeatedly and each EEG epoch was decomposed by ICA. Movement-related spatial maps were identified manually and similar patterns were averaged to yield four spatial templates which were used to select the task-related IC spatial maps.

to their high correlation values. The task-related IC spatial maps and the corresponding temporal components were used to reconstruct the signal X based on Eq. (1).

Step 3: Envelope Reconstruction of Oscillatory Reactivity from Reconstructed Signals Using the Amplitude Modulation Method

The optimal frequency band encompassing the relevant brain activities may vary between individuals and sessions. To tackle this problem, the reactive frequency band was divided into five passbands, 8–12, 12–16, 16–20, 20–24, and 24–28 Hz, with these used to band-pass filter the reconstructed signals. The Amplitude Modulation (AM) method based on the Hilbert transform was applied to detect the envelope of the filtered EEG signals and quantify event-related oscillatory activities.[3] Each envelope, referred to as an AM waveform, was computed based on (see Fig. 6 (a))

$$m(t) = \sqrt{M_{\mathrm{BP}}(t)^2 + H(M_{\mathrm{BP}}(t))^2} \tag{2}$$

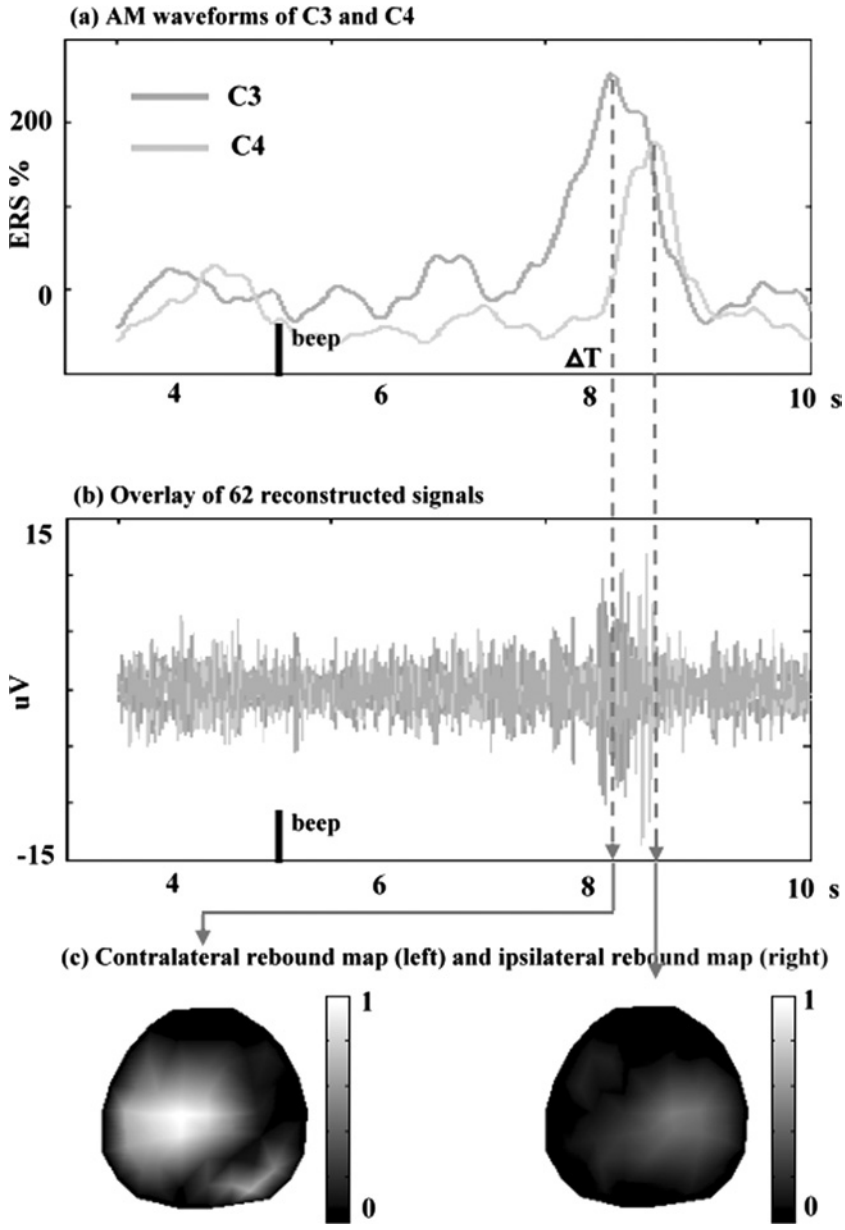

FIGURE 6. Construction of the paired contralateral and ipsilateral rebound maps of imagery right index finger lifting task. (a) The AM waveform of C3 and C4. ΔT is the time lag between rebounds at C3 and C4. (b) Reconstructed signals of 62 channels (overlaid together) which were used to calculate the AM waveforms. (c) The contralateral rebound map (left) and ipsilateral rebound map (right) created from reconstructed signals on 62 channels indexed to the time points of rebounds at C3 and C4. The amplitudes of rebound maps were rescaled to between 0 and 1.

in which $M_{BP}(t)$ is the single-trial band-passed EEG signal, and $H(M_{BP}(t))$ is its Hilbert transform. Following the method used in,[3] ERS reactivity was represented as a relative percentage calculated using

$$\frac{\text{maximum amplitude} - \text{initial baseline}}{\text{initial baseline}} \times 100\%. \quad (3)$$

Step 4: Extraction of Contralateral and Ipsilateral Rebound Maps

The imagery finger lifting task, like actual finger movement, induced larger and earlier rebound in the contralateral sensorimotor area than in the ipsilateral one. Rebounds occurring in contralateral followed by ispilateral sensorimotor areas with a constrained time lag (less than 0.5 s) suggest that the topographical maps encompassing rebounds at C3 and C4 can be considered reliable features. Specifically, two distinct time points were sought such that both the AM waveforms of C3 and C4 have maximum peaks but with a time lag (ΔT in Fig. 6(a) for imagery right index finger lifting). The amplitudes of reconstructed signals across 62 channels at these two time points (Fig. 6(b)) yielded two topographical maps, i.e., contralateral and ipsilateral rebound maps, respectively (Fig. 6(c)). These were concatenated into a feature vector 124×1. Figures 7(a) and 7(c) display an example of extracted rebound maps of imagery right and left index finger lifting, respectively. Both topographical maps manifest stronger activities in the vicinity of sensorimotor areas contralateral to the side of imagery finger lifting.

To show that performance of classifiers benefited from ICA-based rebound maps, the same data sets were processed without ICA decomposition and template matching. That is, for this comparison, only step 3 and the time points used for the ICA rebound maps were used to construct rebound maps. Compared with ICA-based rebound maps which show evident motor patterns in Figs. 7(a) and 7(c), the non-ICA rebound maps, shown in Figs. 7(b) and 7(d), appear to be severely contaminated by noise and artifacts. Note that, although ICA-based method is effective in reducing the noise contamination, complex brain dynamics in a single trial may limit ICA to produce perfect separation and residual of task-unrelated activations (bright spot in the occipital region in the contralateral rebound map in 7(a)) is sometimes inevitable. Procedures for extracting rebound maps with and without the use of ICA are summarized in Figs. 8 and 9, respectively. Each pair of rebound maps were concatenated as a feature vector and fed into four classifiers for subsequent training and testing processes.

TWO-CLASS SUPERVISED CLASSIFICATION

In this section, four two-category classifiers used in our study are briefly reviewed. They were the Fisher linear discriminant (FLD), back-propagation neural network (BP-NN), radial basis function network (RBF-NN) and support vector machine (SVM). Paired rebound maps for imagery right and left hand movement are denoted by $\overrightarrow{x}_i$. Each is a 124×1 column vector. They were divided into two data

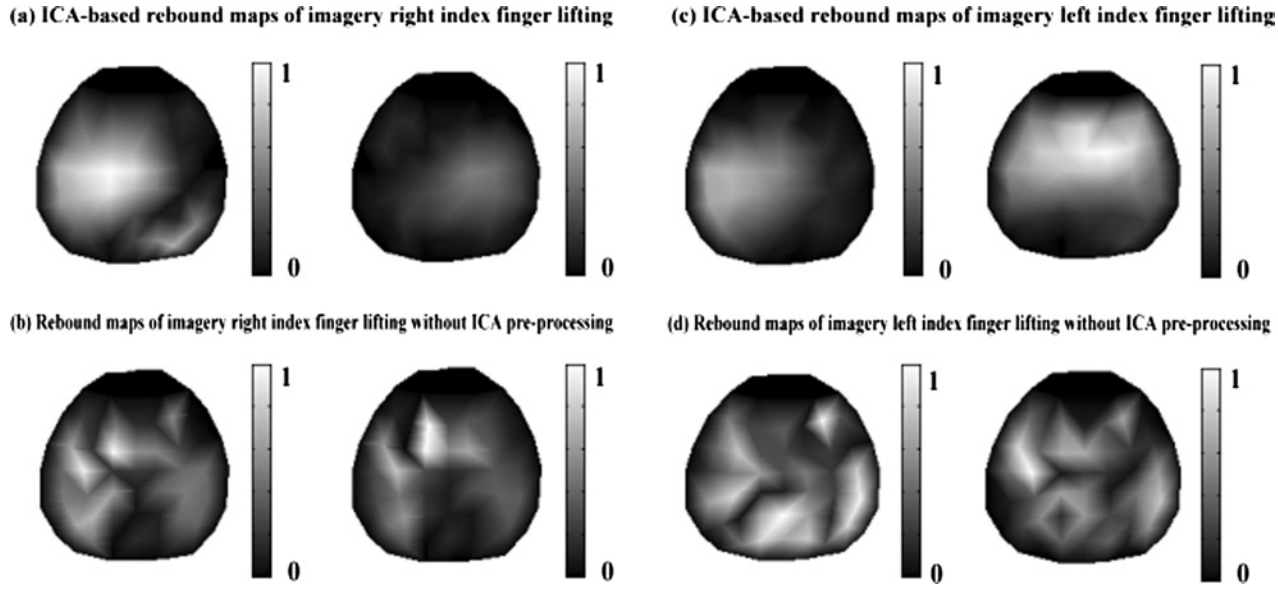

FIGURE 7. A comparison of rebound maps with and without using ICA pre-processing and template matching on a single EEG epoch. (a) and (c) are ICA-based rebound maps of imagery right and left index finger lifting, respectively, which manifest lucid motor patterns. (b) and (d) are non-ICA rebound maps of imagery right and left index finger lifting (using step 3 only), respectively, which show severely noise contamination.

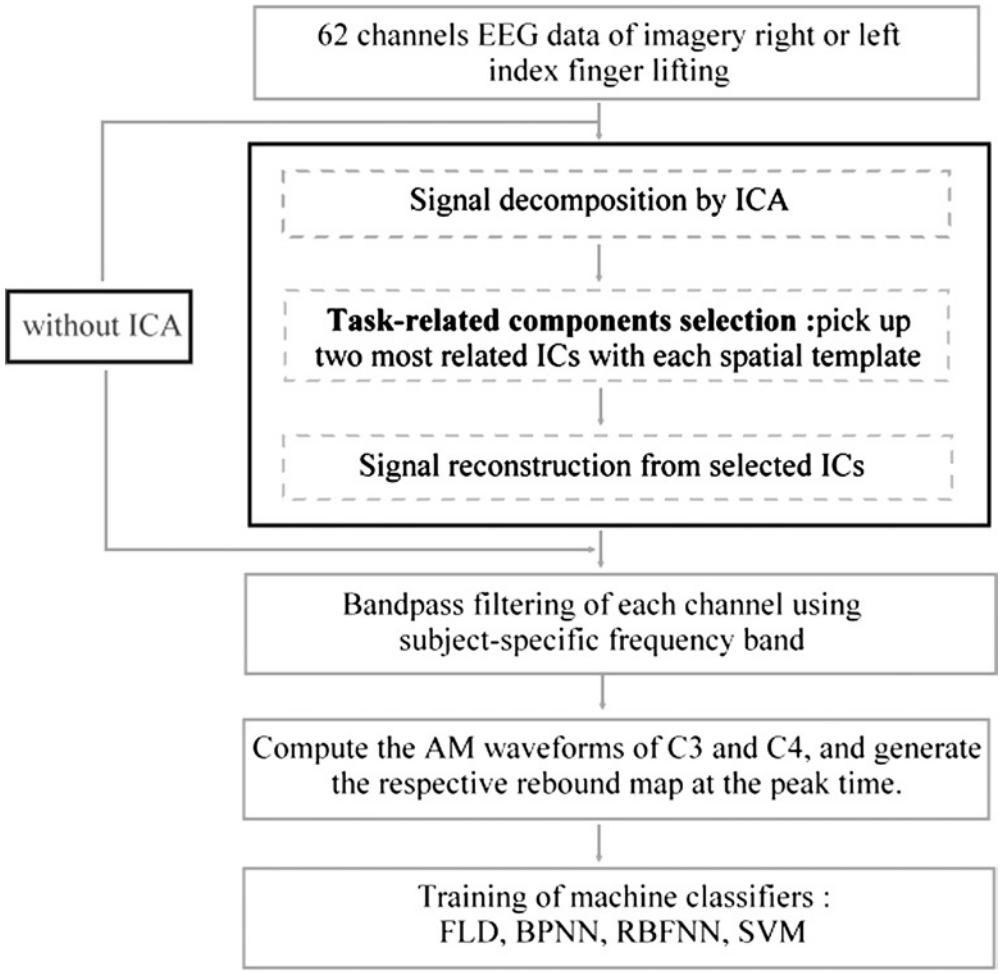

FIGURE 8. The flowchart of rebound maps extraction and training process. The number of rebound maps used in the training stage for each subject at each session was 60.

sets, one for training and the other for classifier testing. The number of rebound maps used in the training and testing phases for each subject at each session were 60 and 30 respectively. These rebound maps were randomly shuffled before separation into training and testing data sets. For the sake of simplicity, the notations R and L are used to denote imagery right and left index finger movement, respectively, in the following discussion.

FLD

The goal of FLD is to seek a vector $\vec{w}$ such that two projected clusters of R and L feature vectors $\vec{x}_i$ on $\vec{w}$ can be well separated from each other while maintaining a small variance for each cluster. This can be done by maximizing Fisher's criterion.[11]

$$J(w) = \frac{w' S_{\mathrm{b}} w}{w' S_{\mathrm{w}} w}$$

with respect to $\vec{w}$, where S_{b} is the between-class scatter matrix defined by:

$$S_b = (m_R - m_L)(m_R - m_L)'$$

and S_{w} is the within-class scatter matrix:

$$S_{\mathrm{w}} = \sum_{x \in R} (\vec{x} - m_R)(\vec{x} - m_L)' + \sum_{x \in L} (\vec{x} - m_R)(\vec{x} - m_L)'$$

in which two summations run over all the training samples of classes R and L, respectively. m_R and m_L represent the group mean of classes R and L, respectively. The optimal $\vec{w}$ is the eigenvector corresponding to the largest eigenvalue of $S_w^{-1} S_b$. After $\vec{w}$ is obtained using the training data, the test samples were projected on it, and projected points were classified using the k-nearest-neighbor decision rule.

BP-NN

The BP-NN was trained in a supervised manner based on the error-correction learning rule. The hierarchy of a

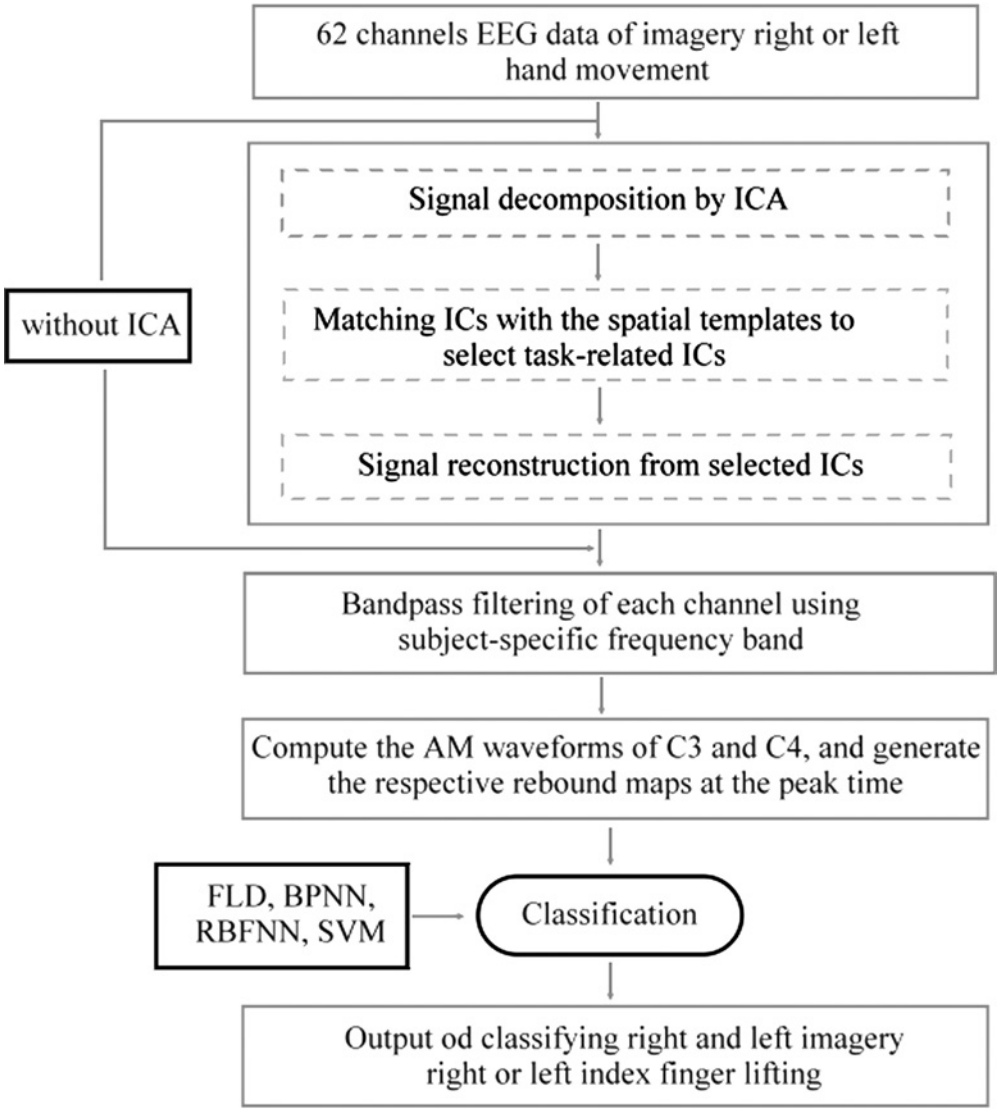

FIGURE 9. The flowchart of rebound maps extraction and classification process. The number of rebound maps used in the testing phase for each subject at each session was 30.

BP-NN in our implementation is depicted in Fig. 10, and consists of one input layer, one hidden layer, and one output layer. The training phase was accomplished by iterating two passes, forward and backward. In the forward pass of the back-propagation learning, as shown in Fig. 10, the output of the BP-NN at iteration n was calculated by

$$y(n) = \varphi(v(n))$$

in which $\varphi(\cdot)$was the activation function and $v(n)$ was the induced local field of output neuron

$$v(n) = \sum_{j=1}^{m} w_j(n) o_j(n),$$

in which m was the total number of the inputs applied to the output neuron, w_j was the weight connecting neuron j to the output neuron, and $o_j(n)$ was the output signal of neuron j. The error signal, $e(n)$, between $y(n)$ and the desired output, $d(n)$, was computed at each iteration. If the error met the stopping criterion, the training procedure was terminated. Otherwise, it was minimized in the subsequent backward pass to update the synaptic weighting $w_j(n)$

$$w_j(n+1) = w_j(n) + \alpha[w_j(n+1)] + \eta\delta(n) o_j(n)$$

in which α was the momentum constant, and $\delta(n)$ local gradients of the output layer in the network, given by $\delta(n) = e(n)\varphi'(v(n))$. During the testing phase, input feature vectors, $\vec{\mathrm{X}}$, can be linearly classified according to the value of $y(n)$ in the output layer.

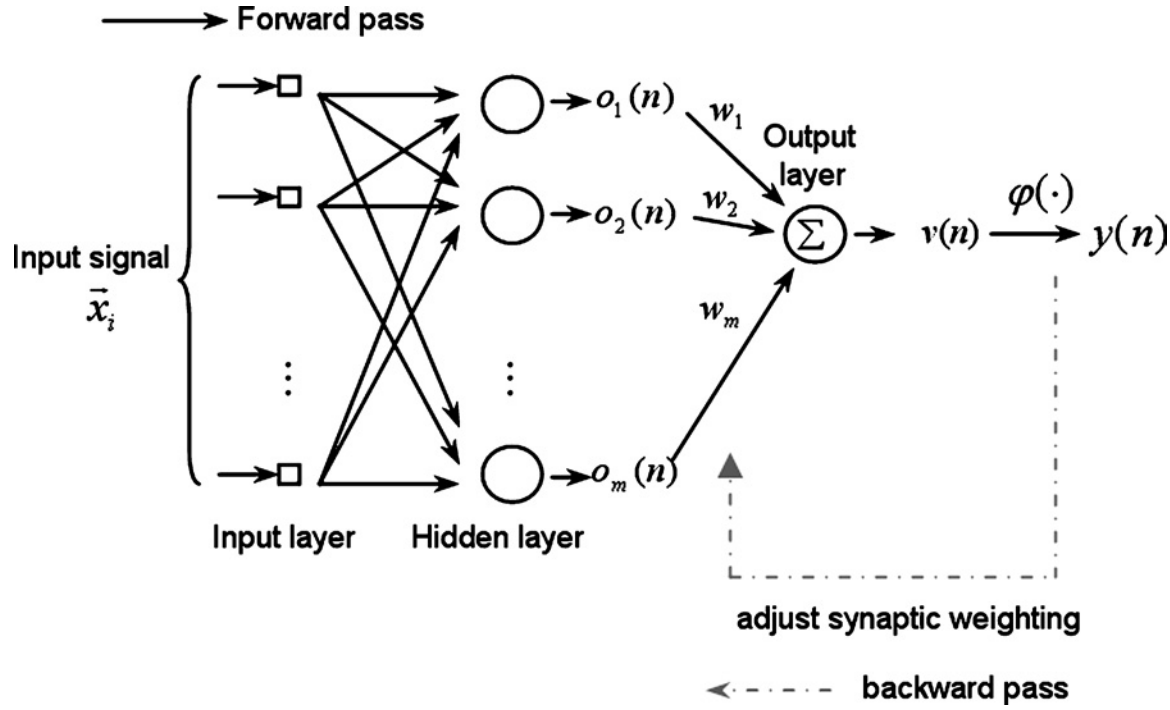

FIGURE 10. The hierarchy of BP neural network used in this study, which consisted of one input layer ($\vec{x}_i$ represents the *ith* feature vector with dimension 1 × 124), one hidden layer (10 neurons), and one output layer.

RBF-NN

The RBF neural network[12] uses a nonlinear function to map the input data into high-dimension space so that they are more likely to be linearly separable than in low-dimension space.[4,5,6]. The hierarchy of (regularization) RBF neural network is shown in Fig. 11, and consists of one input layer, one hidden layer, and one output layer. Each RBF network is designed to have a nonlinear transformation from the input layer to the hidden layer, followed by a linear mapping from the hidden layer to the output layer. The

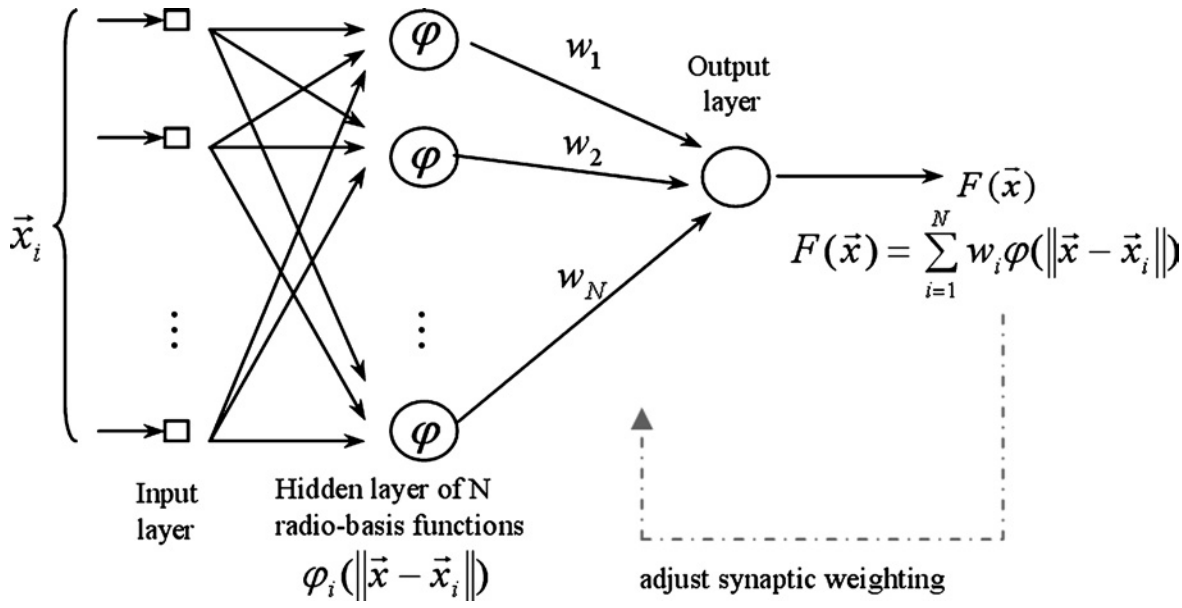

FIGURE 11. The hierarchy of RBF neural network used in this study, which consisted of one input layer ($\vec{x}_i$ represents the *ith* feature vector with dimension 1 × 124), one hidden layer ($N = 60$), and one output layer.

mapping between the input and output space is expressed by:

$$F(\vec{x}) = \sum_{i=1}^{N} w_i \varphi(||\vec{x} - \vec{x}_i||) \tag{4}$$

in which $\varphi(||\vec{x} - \vec{x}_i||) = e^{-||\vec{x}-\vec{x}_i||^2}$ and w_i represent the weighting from the *ith* hidden neuron to output neuron, and $\vec{x}_i$ represents the *ith* known feature vector with dimension m ($m = 124$), $i = 1, 2, \ldots N$ ($N = 60$). The distance between input vector $\vec{x}$, and center $\vec{x}_i$, is mapped into high-dimension space by means of a Gaussian function $\varphi(||\vec{x} - \vec{x}_i||)$. During supervised learning, training feature vectors $\vec{x}_i$, i(1, 2, ... N, and output desired output $F(\vec{x}_i) = d_i$, which is either 1 or -1 in our design, are given. For the sake of simplicity, the training feature vectors are used as centers. With the known N input feature vectors and the corresponding designed outputs, the weighting w_i can be computed from the input-output relationship in Eq. (4):

$$Gw = d \tag{5}$$

in which

$$G = \begin{bmatrix} \varphi(||\vec{x}_1 - \vec{x}_1||) & \varphi(||\vec{x}_1 - \vec{x}_2||) & \cdots & \varphi(||\vec{x}_1 - \vec{x}_N||) \\ \varphi(||\vec{x}_2 - \vec{x}_1||) & \varphi(||\vec{x}_2 - \vec{x}_2||) & \cdots & \varphi(||\vec{x}_2 - \vec{x}_N||) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(||\vec{x}_N - \vec{x}_1||) & \varphi(||\vec{x}_N - \vec{x}_2||) & \cdots & \varphi(||\vec{x}_N - \vec{x}_N||) \end{bmatrix}, \; w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}, \; d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

By solving the linear system (5), the resultant weighting w vector is

$$w = G^{+}d \tag{6}$$

in which $G^{+} = (G^T G)^{-1} G^T$ is the pseudoinverse matrix of G. Compared with other neural networks which use gradient-based optimization processes to estimate weightings, for example, the back-propagation recurrent neural network, the RBF neural network solves for a set of linear equations to avoid entrapment in a local minimum, thus greatly reducing the training time. During the testing phase, input feature vectors, $\vec{x}$, can be linearly classified based on the values of $F(\vec{x})$'s.

SVM

The concept of the support vector machine hinges on two mathematical operations: (1) under an appropriate nonlinear mapping $\varphi(\cdot)$ of an input vector into a high-dimensional feature space, linear separation of data from two categories via a hyperplane,[4,7,12,33] and (2) construction of an optimal hyperplane for separating the features in (1). Let $\vec{x}$ denote a vector drawn from the input space, assumed to be of dimension m_0 and let $\{\varphi_j(\vec{x})\}_{j=1}^{m_1}$ denote a set of nonlinear transformations from the input space to the feature space; m_1 is the dimension of the feature space. Given such a set of nonlinear transformations, we may define a hyperplane acting as the decision surface as follows:

$$\sum_{j=0}^{m_1} w_j \varphi_j(\vec{x}) = 0 \tag{7}$$

in which $w = \{w_0, w_1, ..., w_{m_1}\}$ denotes a set of linear weights connecting the feature space to the output space. Assume that $\varphi_0(\vec{x}) = 1$ for all $\vec{x}$, so that w_0 denotes the bias. Eq. (8) defines the decision surface computed in the feature space in terms of the linear weights of the machine. Defining the vectors $\varphi(\vec{x}) = [\varphi_0(\vec{x}), \varphi_1(\vec{x}), \ldots, \varphi_{m_1}(\vec{x})]^T$ and $w = [w_0, w_1, \ldots, w_{m_1}]^T$, the decision surface can be expressed in compact form as:

$$w^T \varphi(\vec{x}) = 0 \tag{8}$$

Given that the training feature samples $\varphi(\vec{x}_i)$ correspond to the input pattern $\vec{x}_i$, and the corresponding desired response $d_i, i = 1, \ldots, N$, which is either 1 or -1 in our design, the optimal weight vector w can be expressed as[12]

$$w = \sum_{i=1}^{N} \alpha_i d_i \varphi(\vec{x}_i) \tag{9}$$

in which $\{\alpha_i\}_{i=1}^{N}$ are the optimal Lagrange multipliers resulting from maximizing the subject function

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \varphi^T(\vec{x}_i) \varphi(\vec{x}_j) \tag{10}$$

subject to two constraints: (1) $\sum_{i=1}^{N} \alpha_i d_i = 0$, and (2) $0 \leq \alpha_i \leq C$, in which C is a user-specified constant.

Substituting Eq. (9) into (8), we obtain the optimal hyperplane

$$\sum_{i=1}^{N} \alpha_i d_i \varphi^T(\vec{x}_i) \varphi(\vec{x}) = 0 \tag{11}$$

which will be used to linearly separate the testing data. That is, for any testing sample x, if

$$\sum_{i=1}^{N} \alpha_i d_i \varphi^T(\vec{x}_i) \varphi(\vec{x}) 0,$$

x is classified into the subset having the training response $d_i = 1$, otherwise it is classified into the other with $d_i = -1$. In our implementation, the radial basis function was chosen in defining the inner-product kernel $\varphi^T(\vec{x}_i)\varphi(\vec{x})$ as follows:

$$K(\vec{x}, \vec{x}_i) \equiv \varphi^T(\vec{x}_i)\varphi(\vec{x}) = \exp(0.0005||\vec{x} - \vec{x}_i||^2). \tag{12}$$

According to Eq. (11), once the number of nonzero Lagrange multipliers, α_i, is determined, the number of radial-basis functions and their centers are determined automatically. This differs from the design of the conventional neural network, for example, the BP-NN or RBF-NN,[12] where the numbers of hidden layers or of hidden neurons are usually determined heuristically. In other words, the complexity of the SVM is characterized by the number of a subset of training data rather than the dimensionality of the transformed feature space. As a consequence, SVM tends to be less prone to problems of overfitting than BP-NN and RBF-NN.

RESULTS AND DISCUSSION

To examine whether ICA preprocessing improved the recognition rate or not, we generated the rebound maps with and without ICA preprocessing. The continuous EEG signals were first segmented followed by band-pass filtered between 6 Hz and 50 Hz. In the with-ICA preprocessing procedure, these EEG epochs were decomposed into spatial weightings and temporally independent components trial by trial and reconstructed based on the task-related components. Procedure of selecting task-related components was described in step 2 of Section 3. Finally, the paired rebound maps were generated at the time points when the beta AM waveforms of reconstructed data in C3 or C4 exhibited maximum amplitudes. The without-ICA rebound maps, on the other hand, were generated directly from AM waveforms of band-pass filtered EEG epochs. Data collected in each session were separated into 60 and 30 trials for training and testing classifiers, respectively. Parameters in the ICA and classifiers were defined as follows. Each EEG trial was decomposed into 12 independent components in the ICA pre-processing, both BP-NN and RBF-NN have three layers with 10 and 60 hidden neurons, respectively, and the coefficient in radial basis of SVM was 0.0005 (see Eq. (12)).

Table 1 reports the averaged recognition performances of four classifiers based on two types of EEG rebound maps, obtained with and without the use of ICA, from four healthy subjects (s1, s2, s3 and s4) performing two motor-related mental tasks, namely imagined right and left index finger lifting. When the rebound maps constructed from selected IC maps were used as feature vectors, performance of the four classifiers was superior. The last column of Table 1 shows that average recognition rates (mean $\pm$ std in percentages) of FLD, BP-NN, RBF-NN and SVM over four subjects improved from 55.3 (3.1 to 70.5 $\pm$ 7.9, 54.0 $\pm$ 6.1 to 75.5 $\pm$ 7.5, 57.3 $\pm$ 6.9 to 76.5 $\pm$ 8.3 and 55.0 $\pm$ 7.4 to 77.3 $\pm$ 6.8, respectively, and the last two rows show that the averaged scores over four classifiers for subject s1 to s4 improved from 63 $\pm$ 3.8 to 71 $\pm$ 5.2, 55 $\pm$ 3.1 to 84 $\pm$ 3.1, 53 $\pm$ 3.4 to 77 $\pm$ 2.4 and 51 $\pm$ 0.5 to 68 $\pm$ 3.6 respectively. Overall, the performance of classifiers improved significantly, from 55.5% to 75.0%

Another way to evaluate the performance of binary detection classifiers is based on the receiver operating characteristics (ROC) curve, i.e., a plot of true-positive (imagery right/left rebound map being classified as imagery right/left rebound map) rate versus false-positive (imagery right/left rebound map being classified as imagery left/right rebound map) rate. Since the area under the ROC curve (AROC) interprets the probability of a random sample being assigned to the positive class rather than to the negative class, it can be used as a measure of the percent correct classifications by classifiers. Table 2 summaries the mean AROC for all subjects and four classifiers. As seen in the last column and last two rows of Table 2, the use of ICA improved the average AROC for FLD, BP-NN, RBF-NN and SVM over four subjects from .65 to .74, .60 to .76, .62 to .80 and .64 to .81, respectively, and the average AROC over four classifiers for subject s1 to s4 from .68 to .73, .60 to .87, .60 to .78 and .63 to .73, respectively. The overall mean percentage of classification for four classifiers increased 19% from 0.63 to 0.78.

TABLE 1. Averaged recognition rates (in percentages) of four subjects, s1, s2, s3 and s4, over four sessions resulted from different classifiers. Paired contralateral and ipsilateral rebound maps were extracted as feature vectors with and without using ICA.

Classifier	ICA	s1	s2	s3	s4	Mean
FLD	Without	58 $\pm$ 3.6	55 $\pm$ 10.4	57 $\pm$ 12.1	51 $\pm$ 10.2	55.3 $\pm$ 3.1
	With	63 $\pm$ 8.4	80 $\pm$ 20.2	74 $\pm$ 18.5	65 $\pm$ 3.4	70.5 $\pm$ 7.9
BP-NN	Without	63 $\pm$ 4.9	52 $\pm$ 15.2	50 $\pm$ 11.3	51 $\pm$ 13.9	54.0 $\pm$ 6.1
	With	72 $\pm$ 3.2	84 $\pm$ 20.6	79 $\pm$ 18.4	67 $\pm$ 2.4	75.5 $\pm$ 7.5
RBF-NN	Without	66 $\pm$ 8.2	59 $\pm$ 7.1	54 $\pm$ 16.2	50 $\pm$ 10.2	57.3 $\pm$ 6.9
	With	75 $\pm$ 13.7	86 $\pm$ 19.7	79 $\pm$ 16.3	66 $\pm$ 6.9	76.5 $\pm$ 8.3
SVM	Without	66 $\pm$ 10.2	53 $\pm$ 11.4	50 $\pm$ 10.8	51 $\pm$ 10.5	55.0 $\pm$ 7.4
	With	72 $\pm$ 5.6	87 $\pm$ 13.6	77 $\pm$ 11.0	73 $\pm$ 5.6	77.3 $\pm$ 6.8
Mean	Without	63 $\pm$ 3.8	55 $\pm$ 3.1	53 $\pm$ 3.4	51 $\pm$ 0.5	55.5 $\pm$ 5.3
	With	71 $\pm$ 5.2	84 $\pm$ 3.1	77 $\pm$ 2.4	68 $\pm$ 3.6	75.0 $\pm$ 7.1

TABLE 2. Averaged ROC areas over four sessions resulted from different classifiers with and without using ICA for feature extraction. The numbers of paired rebound maps used for training and testing for each subject at each session were 60 and 30.

Classifier	ICA	s1	s2	s3	s4	Mean
FLD	Without	.71	.64	.58	.67	.65
	With	.68	.86	.74	.68	.74
BP	Without	.65	.56	.61	.58	.60
	With	.68	.83	.80	.73	.76
RBF	Without	.73	.60	.54	.62	.62
	With	.75	.91	.77	.76	.80
SVM	Without	.64	.61	.66	.63	.64
	With	.81	.87	.82	.73	.81
mean	Without	.68	.60	.60	.63	.63
	With	.73	.87	.78	.73	.78

To quantify the reliability of classification, we computed the information transfer rate, B, in bits per trial based on the formula.[24]

$$B = \log_2 N + P \cdot (\log_2 P) + (1 - P) \cdot \log_2 \frac{(1 - P)}{(N - 1)}$$

where $N = 2$ is the number of different mental tasks, i.e., left and right motor imagery tasks, and P is the accuracy of classification. Results with and without using ICA are listed in Table 3. The mean information transfer rate have been greatly increased from 0.017 to 0.208 after the data were processed using ICA.

Among four classifiers, BP-NN (75.5 ± 7.5), RBF-NN (76.5 ± 8.3) and SVM (77.3 ± 6.8) performed comparably but were superior to FLD (70.5 ± 7.9). This was not surprising, and indicated a trade-off between recognition rate and implementation complexity. In FLD, it is advantageous to solve the linear projecting function analytically and efficiently, rather than using iterative nonlinear minimization (such as gradient descent) in the learning process for BP-NN and SVM. Though training and classifying projected data on a line is much simpler, linear projections from two classes may result in a confused mixture when variance of training data is large, producing inferior recognition performance. BP-NN, RBF-NN and SVM, on the contrary, assume and learn nonlinear mappings between training feature vectors and desired outputs, which are more general than the linear projection for pattern-classification problems, so that superior performances are expected. As a compromise between recognition accuracy and computational efficiency, the RBF-NN would be preferable for future on-line implementation since its learning procedure requires only the inversion of a matrix of moderate size (60×60), which is less computationally expensive than the learning of BP-NN and SVM, yet achieves performance comparable to that of SVM.

TABLE 3. The information transfer rate over four sessions resulted from different classifiers with and without using ICA for feature extraction. The numbers of paired rebound maps used for training and testing for each subject at each session were 60 and 30.

Classifier	ICA	s1	s2	s3	s4	Mean
FLD	Without	0.019	0.007	0.014	0.0002	0.010
	With	0.049	0.278	0.173	0.066	0.146
BP-NN	Without	0.049	0.001	0	0.0003	0.013
	With	0.145	0.366	0.259	0.085	0.214
RBF-NN	Without	0.075	0.024	0.005	0	0.026
	With	0.189	0.416	0.259	0.075	0.235
SVM	Without	0.075	0.003	0	0.0003	0.020
	With	0.145	0.443	0.222	0.159	0.242
Mean	Without	0.055	0.009	0.005	0.0002	0.017
	With	0.132	0.376	0.228	0.096	0.208

Rhythmic ERS activities, which are time-locked but not phase-locked to external stimuli or tasks, can vary from trial to trial contingent upon variations in a subject's performance and state, linking to fluctuations in expectation, attention, arousal, and task strategy. Most of the single-trial EEG methods were developed for analyzing time-locked, phase-locked, evoked brain activities.[15,16,32] However, approaches to single-trial movement-related oscillatory changes are less explored. To our knowledge, ICA has not been applied to extract the task-related components in past imagery motor studies. The novelty of current method is that the use of ICA with template matching permits the extraction and analysis of oscillatory ERS activities on a single trial base for the study of brain dynamics in BCI and consequently generates reliably paired rebound maps for classification. Existing methods for extraction of ERS signals essentially measure power changes of corresponding frequency bands as derived from the average of dozens trials, and thus is not applicable to BCI which demands trial-by-trial analysis. The significance of using ICA with template matching has been demonstrated in Figs. 7 and 12 and in the with- and without-ICA classification results. Figure 12 shows four typical examples of unselected IC spatial maps (upper-right-hand map in each plot), AM envelopes for corresponding signal time curves (dark curves) and time-frequency maps (lower-right-hand map in each plot) for corresponding signal time curves obtained via the Morlet-wavelet transform.[8] The reactive frequency band was chosen as 20–24 Hz since this band was also used to bandpass filter an averaged EEG epoch at C3 (from 100 trials) of actual right index finger lifting in creating the AM waveform (gray dashed curve in each plot) for comparison. In all plots, none of the IC spatial maps presents prominent activities in the vicinity of C3 or C4, and none exhibit AM waveforms similar to those of actual movement. Time-frequency maps in plots (a) and (b) show transient power changes 2 s prior to and 4–5 s after auditory cuing, respectively, with narrow spectral contents around 8 Hz, and almost total silence for the rest of trial. These two components were considered as noises unrelated to brain activities since movement-related

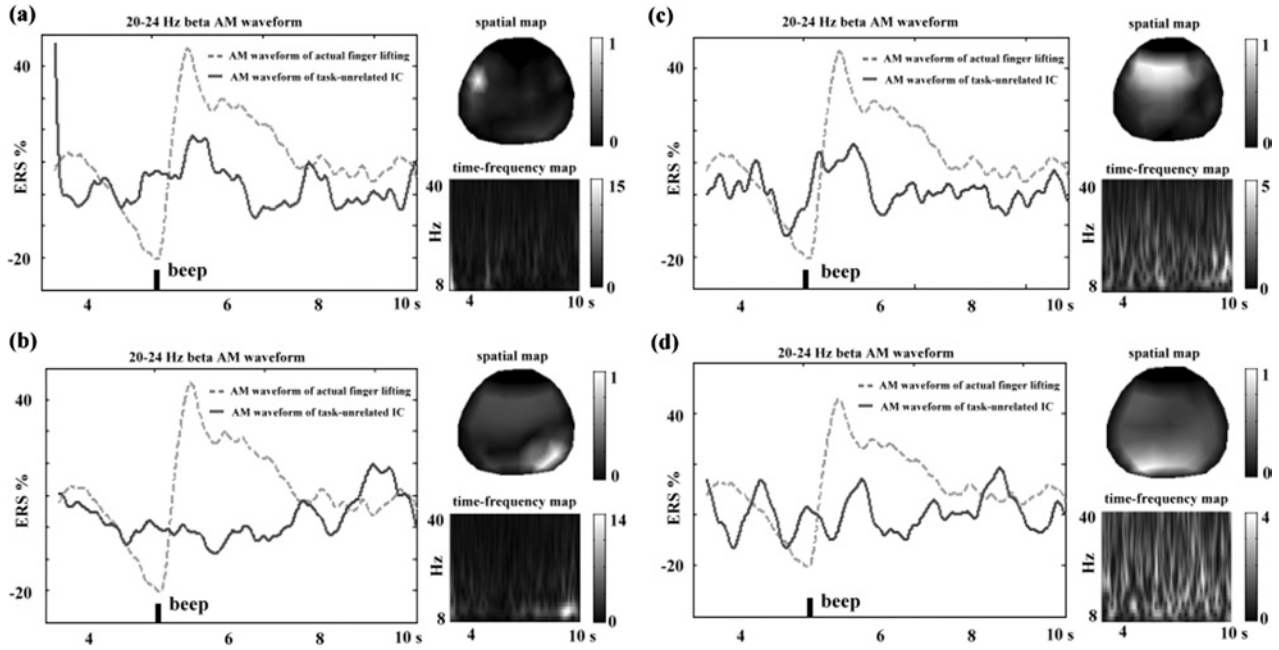

FIGURE 12. Four typical examples of unselected IC spatial maps, their beta-band (20–24 Hz) AM envelope and the Morlet-wavelet time-frequency maps of corresponding signal time curves. The beta-band (20–24 Hz) AM envelope of actual right index finger lifting at C3 (a grand average from 100 trials) was computed and plotted for comparison. (a) and (b) were considered as noises unrelated to brain activities since (1) signal powers changed transiently with narrow spectral contents and remained silence during most of trial, (2) the AM waveforms were dissimilar to that of actual movement and (3) spatial maps exhibit focal activities. (c) and (d) show prominent power change in broad frequency bands and suggest that the decomposed components might be related to cognitive activities, such as anticipation elicited from prefrontal area, or physiologically signals, such as surface electromyogram near occipital region, or occipital oscillation.

brain oscillations usually contain alpha- and beta-band spectral components and induce power changes before and after movement onset. The time-frequency maps in plots (c) and (d) evidence power changes in broad frequency bands across the trial and suggest that the decomposed components might be related to certain cognitive activities such as anticipation elicited from the medial frontal area, or physiological signals, such as surface electromyogram near the occipital region, or occipital oscillation. On the other hand, the resultant AM waveforms of single-trial imagery index finger lifting resemble those of average actual movement (see Fig. 13(a)). Note that the rebounds for imagery movements have longer latency relative to the auditory cue than actual movements, since mental simulation required longer preparation and was more difficult to perform. Figure 13(c) shows time-frequency maps of averaged actual movements in which signal power changes prior to and after auditory cuing, exhibiting alpha ERD (8–12 Hz) and beta ERS (12–20 Hz). Compared to Fig. 13(c), the prominent beta ERS (bright regions around 12–30 Hz) can be clearly identified in Fig. 13(b), while the alpha ERD is much less discernible in single-trial cases. The improvement of signal-to-noise ratio has also been addressed in previous work on ICA-based single-trial analysis of post-movement MEG beta synchronization, in which much circumscribed source localizations congruent to neuroanatomical representations and larger amplitudes of beta rebound were reported.[18]

Since reactive spectral components vary between individuals, the most relevant frequency bands should be determined for each subject and used in the computation of AM waveforms to yield a substantial desynchronization and synchronization in pre- and post-movement periods, respectively. Several frequency bands, such as 8–10, 10–12, 16–20, 20–24 Hz, have been reported to be related to sensorimotor oscillations.[23] To achieve better results in a single-trial classification, past studies have examined relevant spectral components at different frequencies to determine the subject-specific frequency band in motor imagery experiments.[31] Based on these studies, the present study divided the reactive bands into five passbands for simplicity, i.e., 8–12, 12–16, 20–24, 20–24 and 24–28 Hz, and used each to bandpass filter the reconstructed signals from which the rebound maps were extracted. The resultant paired rebound maps from each frequency band were used as feature vectors in the training and testing procedures of classifiers. Optimal frequency bands not only differed between subjects but also between sessions with a single subject when data was recorded on different days. For example, Table 4 shows the optimal frequency band for each subject in each session under SVM classification. Values in parentheses

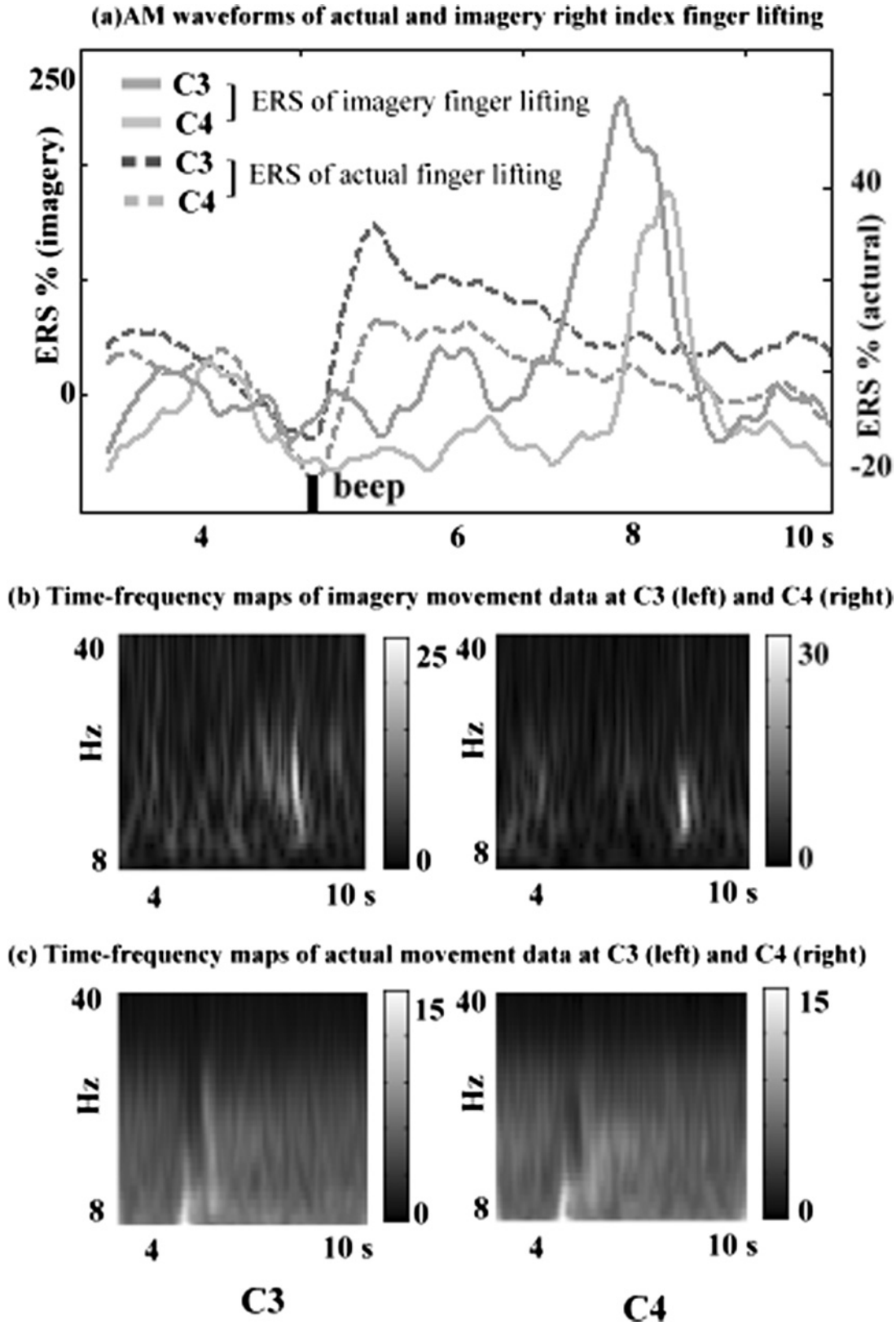

FIGURE 13. Validation of ICA-based AM waveforms (filtered in beta band 20–24 Hz) of motor imagery task. (a) shows that the ICA-based AM waveforms of single-trial imagery right index finger lifting resemble AM waveforms of actual movement (averaged trials), where ERS % denotes 100 × (peak amplitude-baseline)/baseline (b) demonstrates the apparent ERS (bright regions around 12–30 Hz) at C3 and C4 which are also shown in (c) of actual movements.

TABLE 4. The optimal frequency passband used to filter the reconstructed EEG from which paired rebound maps were extracted for each subject in each session. Values in the parenthesis represents the recognition rates (in percentage) resulted from the SVM classifier.

	Session 1	Session 2	Session 3	Session 4
S1	16–20-Hz (75%)	16–20-Hz (78%)	20–24-Hz (65%)	16–20-Hz (71%)
S2	20–24-Hz (69%)	8–12-Hz (84%)	12–16-Hz (98%)	16–20-Hz (97%)
S3	16–20-Hz (93%)	20–24-Hz (73%)	20–24-Hz (74%)	24–28-Hz (68%)
S4	12–16-Hz (66%)	8–12-Hz (72%)	24–28-Hz (76%)	24–28-Hz (79%)

are the mean recognition rates. In the case of subject s1, 16–20 Hz was optimal in sessions 1, 2 and 4 while 20–24 Hz was optimal in session 3. In the case of subject s2, the optimal frequency band was different for each session (20–24, 8–12, 12–16, and 16–20 Hz for session 1, 2, 3 and 4, respectively). Similarly, in the cases of subject s3 and s4, different reactive frequency bands yielded the best recognition rates on different days. The change of relevant frequency components at different sessions might be caused by a learning effect as subjects adjusted their strategies in optimizing task performance, or could be due to different mental states during the experiments, such as attention, arousal, expectation, etc.

A commonly used BCI technique, the common spatial pattern (CSP) method,[22] was also applied to the same data sets. EEG epochs were reduced to 56 channels encompassing sensorimotor and frontal areas as in.[22] During the training phase, the CSP method computed the sum of covariance for averaged EEG from two conditions (R and L) and extracted common spatial patterns carrying the most discriminative information between the paired conditions. Projecting each trial of m by n data of imagery right (left) index finger movement onto the CSP directions ensured that the variance of first row of the transformed data was maximal for condition R (L) and at the same time minimal for condition L (R).[22] On the other hand, the variance of last row was minimal for condition R (L) but maximal for condition L (R). Log-normalized variances of the first two and last two rows of each projected trial were used as a feature vector (4×1). The numbers of feature vectors used in the training and testing phases for each subject per session were 60 and 30, respectively, as with ICA-based rebound maps. Comparing the CSP results with those from rebound maps (see Table 5), the CSP method was overall 5%–10% inferior to the method proposed in this study. This is because the CSP method used variances as features, and these are sensitive to noise or artifacts. In particular, when the signals were severely contaminated by noise or artifacts, variances from major noise or artifacts conveyed none or less discriminative information in two conditions and thus degraded the performance of CSP. The ICA-based method, on the contrary, alleviated the noise and artifact contamination by selecting task-related components from IC spatial maps, leading to superior classification results.

CONCLUSIONS

The current study presents a novel ICA-based method for extraction of event-related oscillatory activities from motor imagery EEG data and constructed paired contralateral and ipsilateral rebound maps, which were used as feature vectors for training and testing four classifiers. With minimum training for each subject (20 min), satisfactory classification rates (more than 70%) from four classifiers have been achieved. Results suggest that ICA-based analysis is effective for artifact removal and extraction of reliable

TABLE 5. A comparison of recognition performance between CSP and the ICA-based method.

Classifier	ICA	s1	s2	s3	s4	Mean
FLD	CSP	68	57	70	63	62.5
	Rebound-map	63	80	74	65	70.5
BP	CSP	62	59	69	63	63.3
	Rebound-map	72	84	79	67	75.5
RBF	CSP	69	58	63	59	62.3
	Rebound-map	75	86	79	66	76.5
SVM	CSP	68	58	69	62	64.3
	Rebound-map	72	87	77	73	77.3
Mean	CSP	66.8	58.0	67.8	61.8	63.6
	Rebound-map	70.5	84.3	77.3	68.3	75.1

neural features, which in turn facilitate classification in BCI systems.

ACKNOWLEDGMENTS

The study was funded by the National Science Council (93-2218-E-001) and the Ministry of Education of Taiwan (89BFA221401).

REFERENCES

[1]Belouchrani, A., K. Abed-Meraim, J. F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* (see also *IEEE Trans. Acoustics, Speech,Signal Process.*) 45(2):434–444, 1997.

[2]Cichock, A., and S. I. Amari. Adaptive blind signal and image processing. England: Wiley, 2002.

[3]Clochon, P., J. M. Fontbonne, N. Lebrun, and P. Etevenon. A new method for quantifying EEG event-related desynchronization: amplitude envelope analysis. *Electroencephalogr. Clin. Neuro-Physiol.*. 98:126–129, 1996.

[4]Cover, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE trans. Electron. Comput.* EC-14:326–334, 1965.

[5]Cover, T. M., and J. A. Thomas. Elements of Information Theory. New York: Wiley, 1991.

[6]Cover, T. M. Capacity Problems for Linear Machines. Washington, DC: Thompson Book, Pattern Recognition, 1988, pp. 293–289.

[7]Cristianini, N., and J. Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press, 2000.

[8]Daubechies, I. Ten Lectures on Wavelets. Philadelphia: Society for Induced and Applied Mathematics, 1992.

[9]Delorme, A., and S. Makeig. EEG changes accompanying leraned regulation of 12-Hz EEG activity. *IEEE Trans. Neural Syst. Rehab. Eng.* 11(2):133–137, 2003.

[10]Ebrahimi T., J. M. Vesin, and G. Garcia. Brain-computer interface a new frontier in multimedia communication. *IEEE Signal Process. Mag.* 20(1):14–24, 2003.

[11]Guger, C., A. Schlogl, C. Neuper, D. Walterspacher, T. Strein, and G. Pfurtscheller. Rapid prototyping of an EEG-based brain-computer interface (BCI). *IEEE Trans. Neural Syst. Rehab. Eng.* 9(1):49–58, 2001.

[12]Haykin, S. Neural Network: A Comprehensive Foundation. New York: Macmillan. 1994.

[13]Hung, C. I., P. L. Lee, Y. T. Wu, L. F. Chen, T. C. Yeh, and J. C. Hsieh. Single-trial quantification of EEG imagery Beta-band post-movement rebound in finger lifting task using independent component analysis (ICA). In: *Proceeding of the 2003 World Congress on Medical Physics and Biomedical Engineering*. Sydney, Australia, 2003.

[14]Hyvarinen, A., J. Karhunen, and E. Oja. Independent Component Analysis. New York: Wiley, 2001.

[15]Jung, T. P., S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski. Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping.* 14:166–185, 2001.

[16]Jung, T. P., S. Makeig, M. J. Mckeown, A. J. Bell, T. W. Lee, and T. J. Sejnowski. Imaging Brain Dynamics Using Independent Component Analysis. *Proc. IEEE* 89(7):1107–1122, 2001.

[17]Kelly, S., D. Burke, P. de Chazal, and R. Reilly. Parametric models and spectral analysis for classification in brain-computer interfaces. In: *Proceedings of the 14th International Conference on Digital Signal Processin.* Greece, July 2002.

[18]Lee, P. L., Y. T. Wu, L. F. Chen, Y. S. Chen, C. M. Cheng, T. C. Yeh, L. T. Ho, M. S. Chang, and J. C. Hsieh. ICA-based spatiotemporal approach for single-trial analysis of post-movement MEG beta synchronization. *NeuroImage*. 20:2010–2030, 2003.

[19]Lins, O., T. Picton, P. Berg, and M. Scherg. Ocular artifacts in EEG and event-related potentials. I: Scalp topography. *Brain Topogr*. 6:51–63, 1993.

[20]Makeig, S., S. Enghoff, T. P. Jung, and T. J. Sejnowski. A natural basis for efficient brain-actuated control. *IEEE Trans. Rehab. Eng*. 8(2):208–211, 2000.

[21]Makeig, S., A. J. Bell, T. P. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems 8*. 145–151, 1996.

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

[23]Niedermeyer, E., and F. H. Lopes da Silva. Electroencephalography : Basic Principles, Clinical Applications, and Related Fields. Baltimore, Md. : Williams (Wilkins, 1999, pp. 958–967.

[24]Obermaier, B., C. Neuper, C. Guger, and G. Pfurtscheller. Information Transfer Rate in a Five-Classes Brain-Computer Interface. *IEEE Trans. Rehab. Eng.* 9(3):283–288, 2001.

[25]Parra, L., and P. Sajda. Blind source separation via generalized eigenvalue decomposition. *J. Machine Learning Res.* 4:1261–1269, 2003.

[26]Pfurtscheller, G., and F. H. Lopes da Silva. Event-related desynchronization. Handbook of Electroencephalography and Clinical Neurophysiology, Revised Series, Vol. 6. Amsterdam: Elsevier Science, 1999, pp. 303–325.

[27]Pfurtscheller, G., and A. Aranibar. Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movements. *Electroencephalogr. Clin. Neurophysiol.* 46:138–146, 1979.

[28]Pfurtscheller, G., A. Stancak Jr, and C. Neuper. Post-movement beta synchronization. A correlate of an idling motor area? *Electroencephalogr. Clin. Neuro- Physiol*. 98:281–293, 1996.

[29]Pfurtscheller, G., C. Neuper, A. Schlogl, and K. Lugger. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans. Rehab. Eng*. 63:316–325, 1998.

[30]Pfurtscheller, G., C. Guger, G. Muller, G. Krausz, and C. Neuper. Brain oscillations control hand orthosis in a tetraplegic. *Neurosci. Lett*. 292:211–214, 2000.

[31]Pregenzer, M., and G. Pfurtscheller. Frequency component selection for an EEG-based brain to computer interface. *IEEE Trans. Rehab. Eng*. 7:413–419, 1999.

[32]Tang A. C., B. A. Pearlmutter, M. Zibulevsky, and S. A. Carter. Blind source separation of multichannel neuromagnetic responses. *Neurocomput*. 32–33:1115–1120, 2000.

[33]Vapnik, V. N. The Nature of Statistical Learning Theory, 2nd ed., New York: Springer-Verlag, 2000.

[34]Wolpaw, J. R., D. J. McFarland, G.W. Neat, and C. A. Forneris. An EEG-based brain–computer interface for cursor control. *Electroenceph. Clin. Neurophysiol*. 78:252–259, 1991.

[35]Wolpaw, J. R., D. J. McFarland, and T. M. Vaughan. Brain–computer interface research at the Wadsworth Center. *IEEE Trans. Rehab. Eng*. 8:222–226, 2000.

[36]Wolpaw, J. R., N. Birbaumer, D. J. McFarlanda, G. Pfurtschellere, and T. M. Vaughan. Brain–computer interfaces for communication and control (Invited Review). *Clin. Neurophysiol*. 113:767–791, 2002.

[37]Wu, Y. T., H. Y. Chen, P. L. Lee, Y. S. Chen, L. F. Chen, and J. C. Hsieh. Classifying MEG Data of Left, Right Index Finger Movement and Resting State Using Support Vector Machine (SVM). In: *Proceedings, BioMag 13th International Conference on Biomagnetism*, 2002, pp. 1042–1044.

[38]Wu, Y. T., P. L. Lee, L. F. Chen, T. C. Yeh, and J. C. Hsieh. Single-trial quantification of imagery beta-band Murhythm in finger lifting task using independent component analysis (ICA). In: *Proceedings, BioMag 13th International Conference on Biomagnetism*, 2002, pp. 1045–1047.

[39]Wu, Y. T., P. L. Lee, L. F. Chen, T. C. Yeh, and J. C. Hsieh. Quantification of movement-related modulation on beta activity of single-trial magnetoencephalography measuring using independent component analysis (ICA). In: *Proceedings, 1st International IEEE EMBS Conference on Neural Engineering*, 2003, pp. 396–398.