



# An improved common spatial pattern combined with channel-selection strategy for electroencephalography-based emotion recognition

Mengmeng Yan<sup>a,b</sup>, Zhao Lv<sup>a,c,\*</sup>, Wenhui Sun<sup>a</sup>, Ning Bi<sup>d</sup>

<sup>a</sup>School of Computer Science and Technology, Anhui University, Hefei 230601, China

<sup>b</sup>Shan'xi Key Laboratory of Network and System Security, Xidian University, Xi'an 710071, China

<sup>c</sup>Institute of Physical Science and Information Technology, Anhui University, Hefei 230601, China

<sup>d</sup>School of Computer Science and Technology, Georgia Institute of Technology, 801 Atlantic Dr NW, Atlanta, GA 30332, USA

## ARTICLE INFO

### Article history:

Received 22 February 2019

Revised 24 April 2020

Accepted 13 May 2020

### Keywords:

Common spatial pattern (CSP)

Emotion recognition

Joint approximation diagonalization (JAD)

Channel selection

## ABSTRACT

Emotional human-computer interaction (HCI) has become an important research area in the fields of artificial intelligence and cognitive science, owing to the requirement for active emotion perception. To enhance the performance of electroencephalography (EEG)-based emotional HCI, this paper proposes an improved common spatial pattern combined with a channel-selection strategy (ICSPCS) for EEG-based emotion recognition. Specifically, we first use a common spatial pattern algorithm to design a spatial domain filter according to three different emotions (positive, neutral, and negative). The traditional joint approximation diagonalization method using the criterion of the “highest score eigenvalue” may be unable to solve multiple classifications of emotion representation. Therefore, we design three different eigenvalue selection methods in terms of the positions of the eigenvalues with the highest scores. Finally, to make the installation easier and reduce the computational load, we also develop a channel-selection strategy based on the weight values that individually reflect the degrees of influence of all the channels on emotion recognition. Experiments are conducted on a self-collected dataset and on the MAHNOB-HCI dataset. The average recognition accuracies for the three emotion tasks are found to be 85.85% and 94.13% on the self-collected and MAHNOB-HCI datasets, respectively, thus proving the effectiveness of the proposed ICSPCS method for emotion recognition.

© 2020 IPEM. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

Human-computer interaction (HCI) refers to the process of information exchange between a user and a computer, with the user using a certain “dialogue” language to interact with the computer to complete certain tasks [1]. Currently, intelligent HCI systems such as intelligent cars, intelligent voice navigation, intelligent medical equipment, and intelligent homes, are rapidly enriching our daily lives [2–4]. Such systems can achieve their corresponding functions according to the user commands well. However, adjusting their interaction mode based on the psychological state of a user is problematic, owing to poor emotion perception. It is difficult to realize a true “intelligent interaction”, which severely restricts the functions and applications of HCI systems. Therefore,

the development of an HCI system with emotional intelligence has become an important research area in the fields of artificial intelligence and cognitive science [5].

For the implementation of an emotional HCI system, the acquisition and recognition of human emotion information is a key step. To achieve this objective, researchers have conducted a series of studies in recent years. Based on the signal acquisition technique, emotion recognition methods can be classified into two categories: contact and contact-free. Currently, contact-free methods are mainly implemented based on human facial expressions or speech. Among them, a speech-based method perceives the emotional states of a user by extracting emotion-related features, such as the tone, energy, and spectrum [6,7]. Similarly, a facial expression method is primarily concerned with the emotional information corresponding to the variations in facial features [8,9]. These features generally consist of static information (e.g., skin color), slowly varying information (e.g., permanent wrinkles), or rapidly varying information (e.g., opening of the mouth or raising of the eyebrows) with respect to time. Contact-free methods have

\* Corresponding author at: Anhui University, School of Computer Science and Technology, School of Computer Science, Anhui...Hefei 230601, China.

E-mail addresses: [e17201017@stu.ahu.edu.cn](mailto:e17201017@stu.ahu.edu.cn) (M. Yan), [kjlz@ahu.edu.cn](mailto:kjlz@ahu.edu.cn) (Z. Lv), [e17201044@stu.ahu.edu.cn](mailto:e17201044@stu.ahu.edu.cn) (W. Sun), [nbi3@gatech.edu](mailto:nbi3@gatech.edu) (N. Bi).

the advantages of simple signal acquisition and being easy for the users. However, when the users attempt to mask their emotion, the real emotional state may be inconsistent with the external presentation. In such a case, contact-free methods have difficulty in obtaining a correct recognition. Owing to the non-deceptiveness of bioelectrical signals, contact methods have received increasing attention for identifying the emotional states of human users [10–12]. In general, the peripheral bioelectrical signals collected from the autonomic nervous system refer to those of electrocardiogram, blood pressure, skin conductance, body surface temperature, and respiration, among other parameters.

In addition to the aforementioned bioelectrical signals, the signals collected from the central nervous system in the brain have been proven to provide informative characteristics under different emotional states. These signals can be recorded using the electroencephalogram (EEG) method, which is a noninvasive brain activity measurement method with a temporal resolution of milliseconds. In recent years, EEG signals have been employed to analyze the procedure of emotion response and to perceive emotional states. For instance, Bhardwaj et al. [13] extracted the power spectral density (PSD) and energy features of theta, alpha, and beta frequency bands to recognize seven emotional states. The accuracies under the two classifiers were 74.14% and 66.5%, respectively. In addition, Zheng et al. [14] used the PSD and differential entropy (DE) features extracted from five frequency bands; the accuracy of this method was 86.08% for three different emotional states. Hadjidimitriou et al. [15] proposed a time-frequency (TF) analysis and obtained an average accuracy of 86.52% for two classes. These methods mainly focus on a time or time/frequency-domain analysis, and ignore the spatial characteristics of the emotions related to the EEG signals.

The use of spatial features has already achieved remarkable results for a motor imagery brain-computer interface. For example, Wang and Hong et al. [16] used event related desynchronization and synchronization phenomena generated by the motion imaging of left and right hands and feet, to control a wheelchair movement in three directions. The accuracies for online and offline use were 79.48% and 85%, respectively. Allison et al. [17] used the spatial features of a EEG to control the movement of a cursor on a two-dimensional (2D) plane effectively. In addition, the application of spatial features has also achieved excellent results in a saccade signal recognition based on electrooculography (EOG) [18]. Motivated by these studies, we expect that spatial characteristics may provide more relevant emotion information to distinguish different emotional states. In this study, we investigate the feasibility of emotion recognition using the common spatial pattern (CSP) algorithm. Furthermore, to improve the recognition accuracy and decrease the computational load, we also design a channel optimization strategy.

The organization of this paper is as follows. Section 2 introduces the generation and collection of EEG signals, emotion expression model, as well as public emotional dataset. The details of the methodology are provided in Section 3. Section 4 presents the experimental results and compares the time-frequency domain approach. Section 5 summarizes the paper and proposes suggestions for future work.

## 2. Materials

### 2.1. EEG Generation

An EEG signal is an electrophysiological monitoring signal obtained from the cerebral cortex using electrodes placed at different points on the skin or scalp of the head. It is transmitted along the nerve fibers from the sensory organs to the brain, and different waves are produced for different activities based on the four

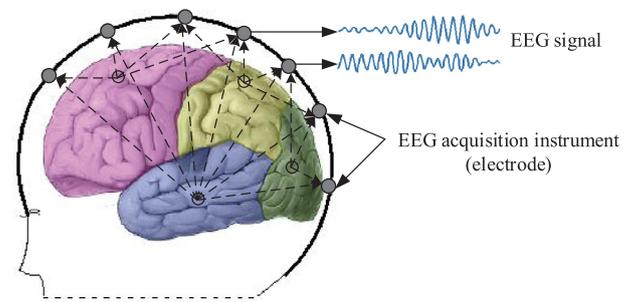


Fig. 1. Generation and detection of multi-channel scalp EEG signals.

main parts of the brain: the cerebrum, cerebellum, brain stem, and thalamus [19]. In general, EEG signals are considered to be composed of billions of neurons present in the human brain. When an emotional response is induced, some neurons located in the corresponding perception area are charged by the ions pumping across the membranes. Consequently, EEG signals can be used to record the spontaneous electrical activities of the brain between the electrode groups placed on the scalp, thus reflecting different human emotional states from the potential variations in thousands or millions of neurons. A schematic of the employed generation and detection procedure is shown in Fig. 1.

### 2.2. Emotion expression

In general, the existing emotion expression models can be classified into two categories: discrete and dimensional. The former focuses on the expression of “primary” emotions using certain discrete labels (e.g., joy, fear, or sadness), whereas, “non-primary” emotions can be considered as a mixture of different “primary” emotions [20]. By contrast, the latter is used to characterize emotions within a dimensional space (generally two or three-dimensions). A well-known dimensional model is the arousal-valence model [21], in which the valence is defined as the degree to which the users incorporate unpleasantness or pleasantness into their conscious emotional experience, and the arousal dimension represents the intensity of user excitement. Considering the convenience of the algorithm estimation and design, we used a discrete model to conduct the emotional analysis and recognition.

### 2.3. Datasets

In the present study, we used two datasets to evaluate our algorithm: a self-collected dataset and the public emotional MAHNOB-HCI dataset [22]. The MAHNOB-HCI dataset (which uses 24 subjects aged 19–40) includes EEG and physiological signals, facial expressions, audio and eye gazes. Nine types of emotions were expected to be elicited: fear, anger, sadness, disgust, anxiety, surprise, neutrality, happiness, and amusement. To achieve a quantitative assessment, discrete emotions were mapped to three types of emotional states based on the valence dimension, refer to [23] for more specific details.

#### 2.3.1. Stimuli selection

Regarding the emotional stimuli selection method, Uhrig et al. [24] and Zheng et al. [25] proved that the emotion elicitation from the stimulus of a movie is more effective than that of either still images or audio alone. Moreover, the elements of native culture (e.g., personal identity, family, educations, and values) play an important role in emotion elicitation. Therefore, we selected movies in the native language of a user and their cultural background as the stimuli. To determine the effectiveness of the emotional movie clips, we employed a clip optimization selection procedure using

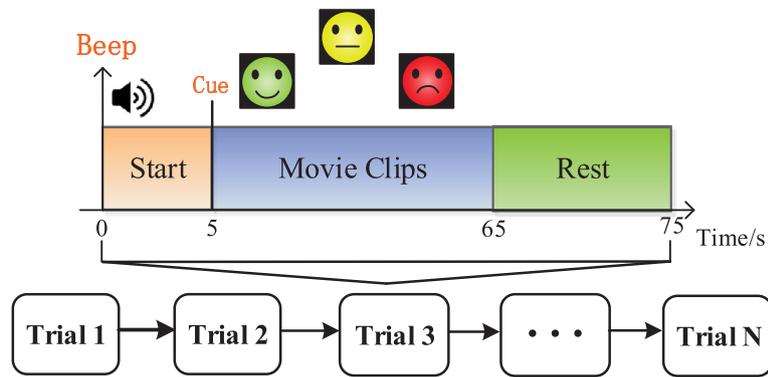


Fig. 2. Timing scheme of the experiment paradigm.

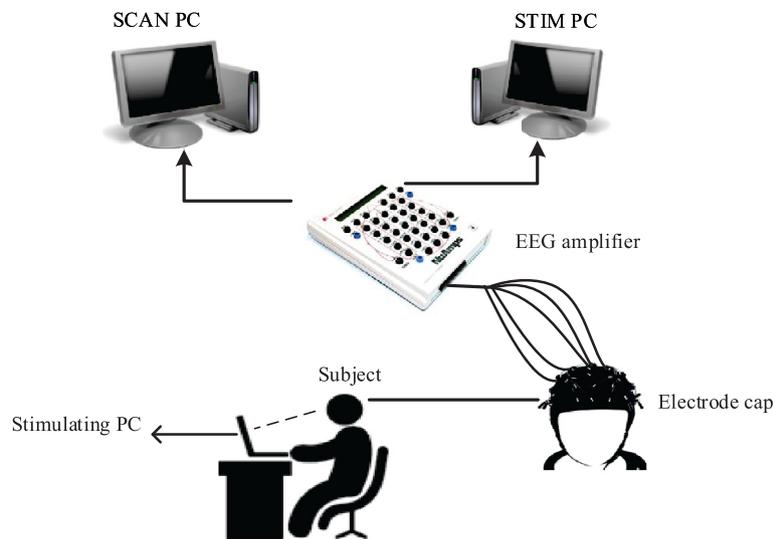


Fig. 3. Data acquisition system. The EEG amplifier is connected to the electrode cap worn by the subject, and the SCAN PC is used to collect and display the EEG signals synchronously. The STIM PC is employed to control both the stimulation time and the emotional labels, and the stimulating PC is used to display the stimuli-invoking movies.

the method proposed in [26]. Specifically, we initially selected a set of emotional film clips from Chinese movies obtained through a self-assessment of the subjects. We then divided each emotional intensity into five levels ranging from levels 1 to 5. Further, we invited volunteers who were not involved in subsequent experiments to assess the emotional intensities of all the clips, and score them as levels 1–5. The higher the score, the stronger the emotional intensity. Finally, we selected 36 highest scoring movie clips as the stimuli for the present study.

### 2.3.2. Experimental paradigm design

To induce emotional EEG signals effectively, we designed an experimental paradigm based on the time scheme in [27], as presented in Fig. 2. In this paradigm, each trial starts with a short warning tone (beep) indicating a 5-s preparation time. Subsequently, three different of 60-s durations are displayed at random. The subject is then allowed to rest for 10-s to complete the next trial better. During the experiment, each subject was asked to sit in a comfortable chair facing the computer screen and to focus on the movie clips as they were displayed. Notably, the subjects needed to keep their bodies maximally still, to avoid the additional interference caused by the micro-movements of the electrodes.

A total of 8 healthy volunteers (3 male and 5 female) aged between 23 and 26 (mean age of 24.63, with a standard deviation of 1.06) were involved in the present study, after receiving written informed consents from them. All the subjects were students from

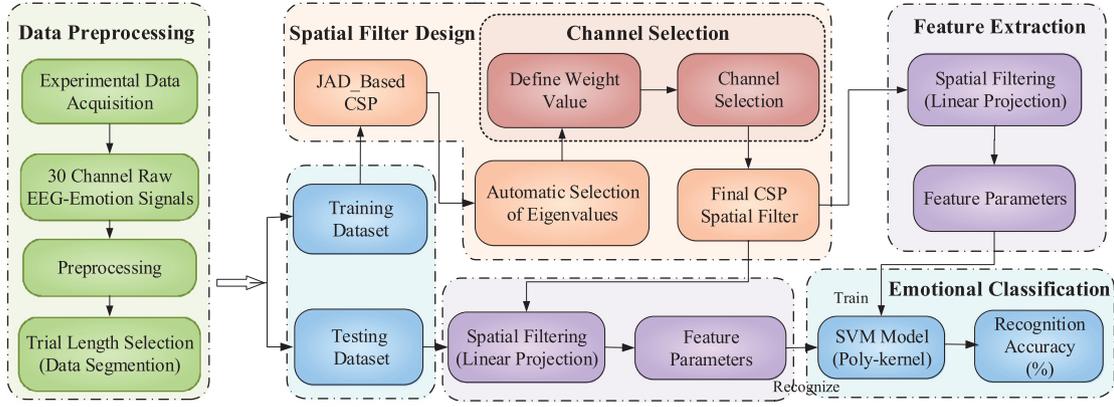
our laboratory who have normal or corrected to normal vision and a normal cognitive ability. Before the experiment, all the subjects were informed of the purpose and procedure of the following experiments. In addition, 32-channel emotional EEG signals were acquired using a SynAmps amplifier (Neuroscan Inc., Herndon, VA) with a 16-bit resolution at a sampling rate of 250 Hz. The locations of the acquisition electrodes were in line with the International 10–20 system [28]. The data acquisition system is displayed in Fig. 3.

## 3. Methods

The objective of the proposed ICSPCS method was to explore the feasibility of applying spatial features to EEG-based emotion recognition. The procedure includes preprocessing of the raw EEG signals, a spatial filter design, channel selection, feature extraction and emotional classification. A flow chart of the procedure is illustrated in Fig. 4.

### 3.1. Data preprocessing

To suppress the effects of artifacts effectively in EOG, electromyography (EMG), power line interference, and electromagnetic interference, we first apply a preprocessing step to the raw EEG signals. Specifically, a 50-Hz notch filter is initially set to suppress the power line interference. Next, considering the distributions of



**Fig. 4.** Flow chart of the proposed ICSPS method. The “data preprocessing” unit is used to acquire and filter the raw 30-channel emotional EEG signals. The “spatial filter design” unit is used to implement the spatial filter design based on the joint approximate diagonalization (JAD) algorithm. The “channel selection” unit is employed to select the appropriate channel using the proposed channel selection method. The “feature extraction” unit and “emotional classification” units are used to extract the EEG spatial features and recognize the emotions, respectively.

the effective emotion response frequency bands (the conducted verification experiment is described in Section 4.1), we employ a linear-phase finite impulse response filter with a cut-off frequency of 8–60 Hz, to filter the EEG signals. Finally, the mean removal operation is applied to the filtered signals to eliminate the influence of the DC component.

### 3.2. Feature extraction using the improved CSP

#### 3.2.1. Solution for multiclass CSP

The traditional CSP algorithm [29] focuses on solving a two-class problem; it distinguishes between different classes by determining the projection space that maximizes the energy difference between two types of samples. In the present study, we focus primarily the recognition of three emotional states (i.e., positive, neutral, and negative). Thus, it is necessary to extend the traditional CSP algorithm to a multiclass algorithm.

Currently, there are three different solutions for a multi-class CSP: one-to-one method, one versus rest method (OVR), and joint approximation diagonalization (JAD) method. Among them, the OVR and one-to-one CSPs are realized by conducting a two-class CSP on different combinations of classes. Indeed, this two-class CSP algorithm needs to be repeated several times, resulting in a more complex algorithm and a longer calculation time. In contrast, Gouy-Pailler et al. [30] and Nguyen et al. [31] demonstrated that the JAD method is more effective in multi-task EEG feature extraction and achieves powerful scalability.

Similar to a traditional two-class CSP algorithm, the JAD method designs spatial filters through the diagonalization of the covariance matrix for each class [32]. The main difference between them is the eigenvalue selection strategy. Specifically, when the eigenvalue of one class is maximized, the JAD method cannot ensure that it differs from the maximized eigenvalues of other classes. To solve this problem, all eigenvalues  $\lambda$  are mapped to the follow functional score [32], and the top  $M$  highest eigenvalues of each class are selected for the design of the CSP spatial filters as follows:

$$\text{score}(\lambda) = \max \left[ \lambda, \frac{1 - \lambda}{1 - \lambda + (N - 1)^2 \lambda} \right] \quad (1)$$

where  $N = 3$  represents the number of emotional classes, and  $\lambda$  is the diagonal element in the eigenvalue matrix  $\Sigma_i$ , ( $i = 1, 2, 3$ ). Motivated by the principle of the two-class CSP algorithm [25], the relationship between the eigenvector and the eigenvalue corresponding to the considered three different classes can be described as

follows:

$$\mathbf{U}_1 = \mathbf{U}_2 = \mathbf{U}_3, \quad \Sigma_1 + \Sigma_2 + \Sigma_3 = \mathbf{I} \quad (2)$$

Furthermore, the spatial filters  $\mathbf{W}_i$ , ( $i = 1, 2, 3$ ) can be designed by selecting the eigenvector matrices  $\mathbf{U}_a$ ,  $\mathbf{U}_b$ , and  $\mathbf{U}_c$ , which are the eigenvectors corresponding to the highest eigenvalue using Eq. (1) for each class, and  $\mathbf{P}$  is the pre-whitening transformation matrix, i.e.,

$$\mathbf{W}_1 = \mathbf{U}_a^T \cdot \mathbf{P}, \quad \mathbf{W}_2 = \mathbf{U}_b^T \cdot \mathbf{P}, \quad \mathbf{W}_3 = \mathbf{U}_c^T \cdot \mathbf{P} \quad (3)$$

Finally, the spatial features  $\mathbf{SF}_i$ , ( $i = 1, 2, 3$ ) of the three-classes of emotional states can be computed using the following improved equation, where  $\mathbf{X}$  is the pre-processed multi-channel EEG signal:

$$\mathbf{SF}_i = \mathbf{W}_i \cdot \frac{\mathbf{X} \cdot \mathbf{X}^T}{\text{trace}(\mathbf{X} \cdot \mathbf{X}^T)}, \quad (i = 1, 2, 3) \quad (4)$$

#### 3.2.2. Improved multiclass CSP

During the emotion recognition experiments, we note that the traditional JAD-based multiclass CSP method using the “highest score eigenvalue criterion” presents a phenomenon in which the eigenvalues with the highest scores for the different emotional states correspond to the same eigenvector. Specifically, the eigenvalues of the different emotional states are located at the same position in the diagonal matrix. Consequently, the eigenvectors are combined with the same transform matrix  $\mathbf{P}$  to obtain the spatial filters, which might represent different emotional states. It is remarkable that the spatial domain filter designed in such a case is an invalid filter, resulting in a poor recognition performance. To resolve this problem, some studies [33,34] have proposed using a method for selecting the next-highest transformed eigenvalue instead of a traditional selection method. Specifically, if one eigenvector is selected more than once, it will be replaced by the eigenvector with the next-highest transformed eigenvalue. However, this method may not be able to determine the spatial projection direction that maximizes the energy difference of three types of emotional EEG signals. To further enhance the performance of the algorithm, we propose an improved JAD method.

Based on all the possible position cases in which the eigenvalues with the highest scores of the three emotional states are located in the diagonal matrix, three different eigenvalue selection methods, JAD1, JAD2, and JAD3, are designed according to the following rules:

- The JAD1 method uses the highest scoring eigenvalues of each class (in which all the eigenvalues are arranged in descending

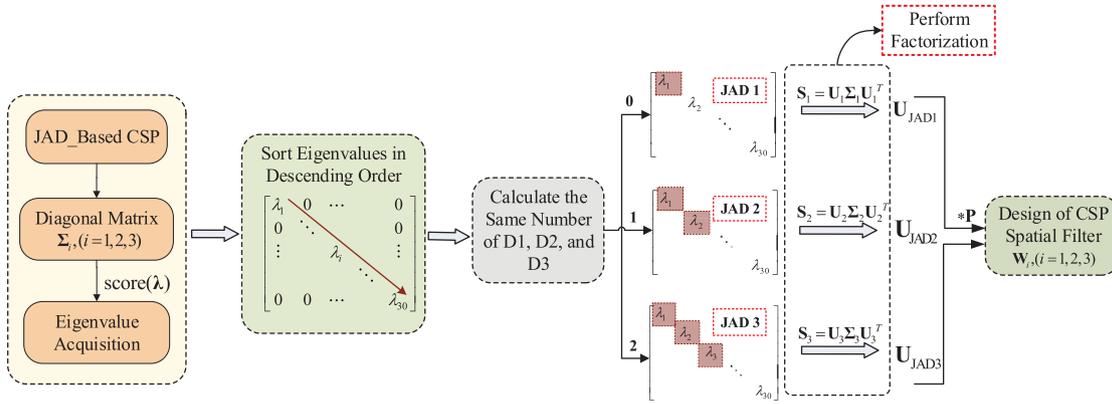


Fig. 5. Automatic selection algorithm of the eigenvalues based on JAD.

order), i.e., the first column of the eigenvector matrix  $\mathbf{U}$  is selected and denoted as  $\mathbf{U}_{\text{JAD1}}$ .

- The JAD2 method uses the top two eigenvalues with the highest scores in each class, i.e., the top two columns of the eigenvector matrix  $\mathbf{U}$  are selected and denoted as  $\mathbf{U}_{\text{JAD2}}$ .
- The JAD3 method employs the top three eigenvalues with the highest scores in each class, i.e., the top three columns of the eigenvector matrix  $\mathbf{U}$  are selected and denoted as  $\mathbf{U}_{\text{JAD3}}$ .

A block diagram of the improved CSP algorithm based on the above three predefined eigenvalue selection methods is presented in Fig. 5. This algorithm can be elucidated as follows. Initially, the traditional JAD method is used to obtain the diagonal matrix  $\Sigma_i$ , ( $i = 1, 2, 3$ ) of the three types of emotional tasks for a training set. The eigenvalues are subsequently obtained using the  $\text{score}(\lambda)$  function. The scores are arranged in descending order, and the positions of the highest scored eigenvalues in the diagonal matrix are saved as  $\mathbf{D}_i$ , ( $i = 1, 2, 3$ ). Finally, we compare the positions to determine which features of the selection method are suitable for the current condition. More specifically, if the positions of the eigenvalues with the highest scores are all different (i.e., the number of the same positions is equal to 0), we select the JAD1 algorithm for the design of the spatial filter. Similarly, if any two positions are the same (i.e., the number of the same positions is equal to 1), we select the JAD2 algorithm. Otherwise, JAD3 is chosen if all the positions are the same (i.e., the number of the same positions is equal to 2).

Accordingly, the improved spatial filters  $\mathbf{W}_i = \mathbf{U}_i^T \cdot \mathbf{P}$ , ( $i = 1, 2, 3$ ) of three emotional states are designed. On this basis, the spatial features for emotional EEG signals can be obtained using Eq. (4).

### 3.3. Channel selection strategy

It is necessary to perform a channel selection by considering the following factors: (1) Installation of the electrodes in the whole-channel mode will be extremely inconvenient during the experimental preparation period. (2) Because some EEG channels are irrelevant to emotion expression [14], some extra noises may be introduced, leading to the performance of the classifiers being degraded. (3) An excessive number of channels of features may result in a high computational complexity and low stability in identifying the signals. Therefore, determining a method to reduce the number of channels effectively while maintaining a high recognition accuracy has received increasing attention in the field of brain-computer interactions. To achieve this objective, some channel selection methods have been proposed. For example, Lan and Erdogmus et al. [35] used the mutual information

size between channels as the channel selection criterion, and Wang et al. [36] employed the maximum value of the spatial pattern vector in the scalp map to select important channels. In comparison, Lal et al. [37] applied a classifier to evaluate the features of different channels. The aforementioned methods mainly focus on the classification of two-class motor imagery tasks (i.e., the left and right hands). In addition, the distribution and number of optimum emotion-related channels may vary among different individuals, because their responses to different stimuli are varied. Motivated by the previous findings, we developed a new channel selection strategy based on the contribution of each channel to emotion recognition.

Initially, we define a weight value  $\mathbf{Q}$  for each channel by introducing a 2-norm [38], i.e.,

$$\mathbf{Q}(j) = \frac{\|\omega_j\|_2}{\|\mathbf{W}_{\text{CSP}}\|_2}, \quad (j = 1, 2, \dots, 30) \quad (5)$$

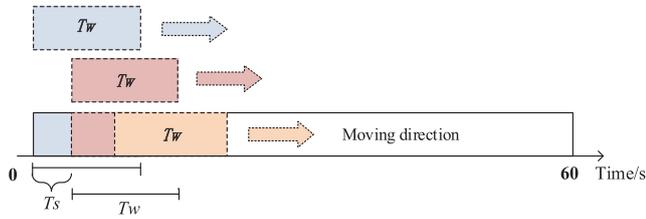
In Eq. (5), function  $\|\cdot\|_2$  indicates the operation of the 2-norm computation. Accordingly, we describe the relationship between the  $\mathbf{W}_{\text{CSP}}$  matrix and  $\omega_j$ , ( $j = 1, 2, \dots, 30$ ) vector as follows:

$$\mathbf{W}_{\text{CSP}} = [\mathbf{W}_1; \mathbf{W}_2; \mathbf{W}_3] = [\omega_1, \omega_2, \dots, \omega_{30}] \quad (6)$$

where the vector  $\omega_j$ , ( $j = 1, 2, \dots, 30$ ) represents the projection of a channel signal in the projection space and reflects its influence on the projected signal. The larger the  $\mathbf{Q}$  value of the observation channel is, the greater the impact on the emotion representation, and vice versa. Finally, we sort all the  $\mathbf{Q}$  values in descending order and select the top- $M$  channels to design the spatial filter.

## 4. Analysis of the experiment results

In the experiments, we collected EEG signals from eight subjects according to three types of emotional states (positive, neutral, and negative). To ensure the short-term stability of the observation signals, we further framed each trial (with a duration of 60-s) with a 6-s window and one-third sliding step length. The details of the sliding window for each trial are displayed in Fig. 6. Accordingly, a total of 972 samples can be obtained for each subject. We selected support vector machine (SVM) with polynomial kernel, to classify the emotional states. The penalty factor of the SVM model was empirically set as 1.0. Moreover, to ensure the reliability of the experimental results, we adopted the five-fold cross-validation scheme [39]. Specifically, the preprocessed EEG data were equally divided into five parts; four parts were used as the training sets, and one part was used as the test set, to acquire the recognition accuracy. Furthermore, we repeated this procedure ten times and averaged all the accuracies as the final output.



**Fig. 6.** Sliding window movement for each trial ( $T_w$  and  $T_s$  represent the duration of each trial and the length of the sliding step, respectively).

#### 4.1. Determination of the optimum parameters for CSP

The length of the sliding window and the observation frequency band are closely related to the recognition performance of the proposed CSP-based emotion recognition algorithm. Therefore, the selection of both the sliding window length and the critical frequency is a crucial step in achieving an effective emotional HCL. To determine the optimal length of the sliding window, we conducted comparison experiments by computing the recognition accuracy at different durations for each trial from 3-s to 10-s with a step size of 1-s (see Fig. 7).

Fig. 7 shows that the recognition accuracy presents a variation among all the subjects corresponding to the various durations of the emotional segments. Among all the durations, the method has the lowest accuracy (81.43%) when the length of the sliding window is 3-s. With an increase in the duration, the recognition accuracy gradually improves. When the duration is 6-s, the average accuracy over all the subjects is the highest (85.85%). Subsequently, as the duration increases, the accuracy attains a relatively stable state. Considering the balance between timeliness and accuracy, we finally determine 6 s as the optimum length of the sliding window to frame the ongoing emotional EEG signals and apply it to the following experiments.

Regarding frequency band selection, previous studies have provided some relevant references. For instance, Danny et al. [40] concluded that alpha and beta bands are more suitable for EEG emotion recognition. Zheng et al. [14] confirmed that the beta and gamma oscillations of a brain activity are more related to emotional processing than other frequency bands. Wang et al. [41] found that the emotions shown in an EEG are produced in three frequency bands (alpha, beta and gamma), whereas the delta and theta bands have no significant relation with emotions. To determine the optimum band frequencies, we applied a sliding window of 6 s (the optimal length) to extract the effective spatial features corresponding to each frequency band. Fig. 8 displays the average recognition accuracies for different frequency bands (delta, theta, alpha, beta, and gamma).

The results presented in Fig. 8 indicate that the classification accuracies of the alpha (8–12 Hz), beta (13–30 Hz), and gamma (31–60 Hz) frequency bands are remarkably higher than those of the delta (1–3 Hz) and theta (4–7 Hz) frequency bands. To decrease the computation load, we discarded the delta and theta bands while retaining the alpha, beta, and gamma bands. A closer evaluation of the accuracy under 8–60 Hz (a combination of the alpha, beta, and gamma frequency bands, which we refer as “ABG”) presents that its recognition performance is superior to the case when using only a single frequency band. This suggests that the high-frequency bands are more relevant to emotional cognitive activities than the low-frequency bands, which is consistent with the existing research conclusions to a large extent [40,41]. Therefore, we selected the frequency band to filter the raw 30-channel EEG signals to achieve a good recognition performance.

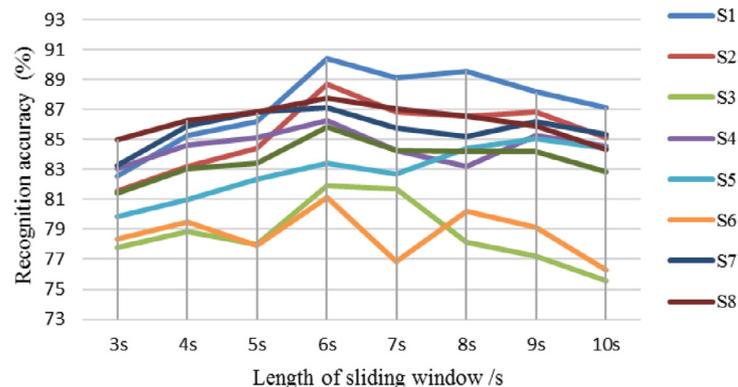
#### 4.2. Performance evaluation of emotion recognition

##### 4.2.1. Verification of channel selection strategy

To validate the effectiveness of the proposed channel-selection strategy, we conducted comparison experiments using the weight values and recognition accuracies of all the channels. The procedure used for the computation of these two criteria was as follows:

- The weight values are calculated  
We first employ the improved CSP algorithm (described in Section 3.2.2) to design a spatial-domain filter based on all the channels. We then compute the weight value  $Q(j)$ , ( $j = 1, 2, \dots, 30$ ) for each channel using Eq. (5).
- The recognition accuracies are calculated  
Assuming that vector  $IC_1, \dots, IC_j, \dots, IC_{30}$ , ( $j = 1, \dots, 30$ ) (where IC denotes independent component) represents the emotion-related spatial features corresponding to the 30 observation channels, we select  $IC_1, \dots, IC_{j-1}, IC_{j+1}, \dots, IC_{30}$ , ( $j = 1, \dots, 30$ ) instead of  $IC_j$  as the feature parameters of the emotional EEG signals. The “leave one IC out” method [42] is employed to select different ICs from each channel for emotion recognition. Specifically, 29 ICs are selected from the 30 channels, with  $IC_j$  being excluded. Accordingly, we can acquire the recognition accuracy ratios based on the current 29 channels. Removing the observation channels in turn, we thus obtain the recognition accuracies through 30 round tests. These accuracies represent the importance of the emotion recognition models. The higher the accuracy, the more important the IC.

Based on the degree of correlation between the channel and the emotion represented by the above two methods, the channels are arranged from high to low degree, and one channel is added to each test until reaching all the 30 channels. The experiment results are presented in Fig. 9.



**Fig. 7.** Emotion classification results under different sliding window lengths. S1–S8 denote the subjects.

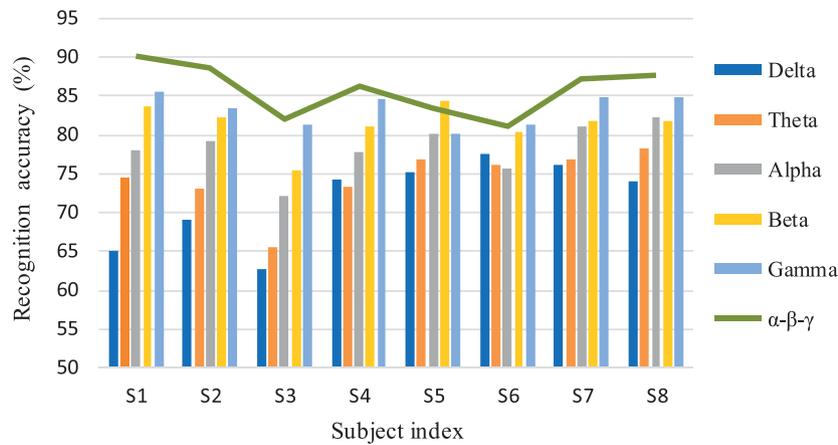


Fig. 8. Accuracies of the different frequency bands over all the subjects. The vertical axis is the average accuracy using a frequency band, and the horizontal axis denotes the index of the subject.

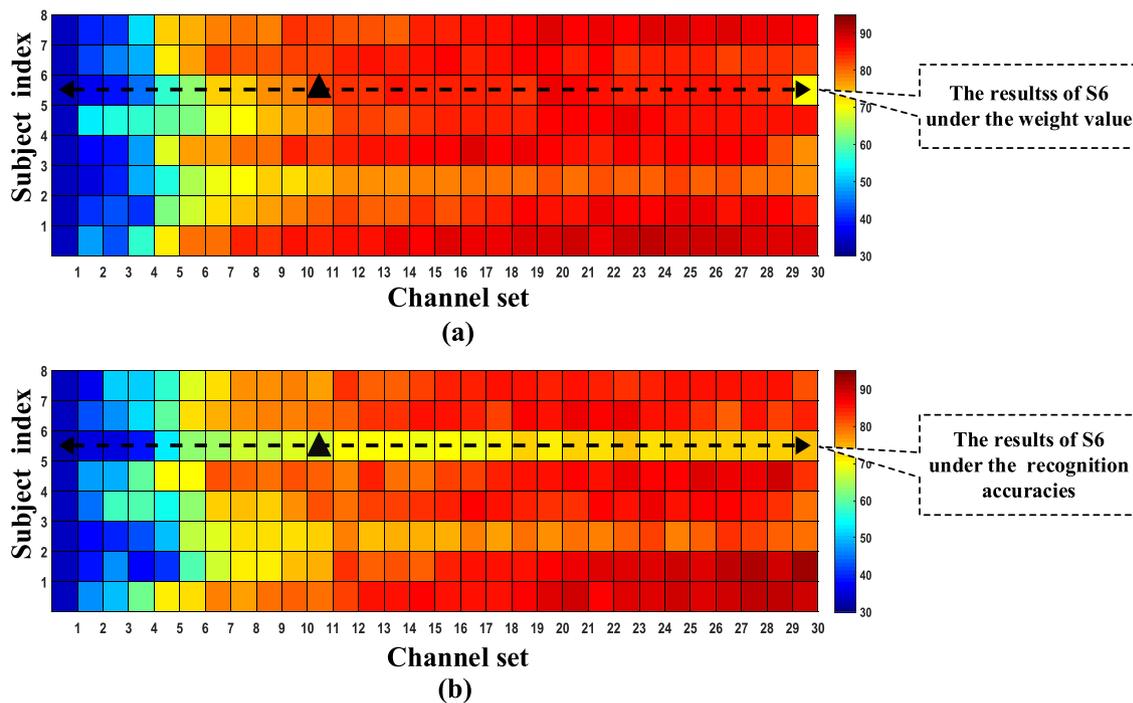


Fig. 9. Accuracy matrix for different numbers of channels for each subject: results using (a) the weight values and (b) the recognition accuracies. The rows indicate the number of selected channels, and the columns represent the indexes of the subjects. Each small block presents the classification accuracy obtained from the  $i$ th subject with  $j$  channels. The redder the color, the higher the accuracy, and the bluer the color, the lower the accuracy.

Comparing Fig. 9(a) and (b), we can note that the recognition accuracies using the proposed weight value-based method are remarkably superior to those of the recognition accuracy-based method. Particularly for subject S6 (the accuracies marked by black dashed lines in Fig. 9), the recognition performance of the proposed method, as displayed in Fig. 9(a), is effectively improved when the channel number is set as 11 (whose accuracies are marked by black triangles). By contrast, the performance when using the recognition accuracy-based method, as presented in Fig. 9(b), is relatively poor, even when more channels are selected for the analysis. The comparison results indicate that using the weight values to optimize the channels is effective for a CSP-based emotion recognition.

Furthermore, we explored the relationship between the number of channels and the recognition accuracy to determine the optimum channel set-up for emotion recognition. During the exper-

iment, we selected the top- $M$  channels ( $M = 1, 2, \dots, 30$ ) by descending weight value  $\mathbf{Q}$  to evaluate the influence on the recognition performance. The detailed accuracies for each subject are depicted in Fig. 10.

From Fig. 10, we can observe that the number of channels is related to the recognition performance. Specifically, it is difficult to obtain a high recognition accuracy by selecting numerous or very few channels. An overly small number of channels provides insufficient information regarding the neural activity of the brain under different emotional states, which makes ensuring the recognition performance problematic. Similarly, using excessive channels may introduce artifacts and contain unnecessary redundant information. In such a case, the CSP coefficients cannot exactly describe the spatial emotional information. In addition, we can also observe that the recognition accuracy presents a tendency to increase when the number of channels is less than 19; by contrast,

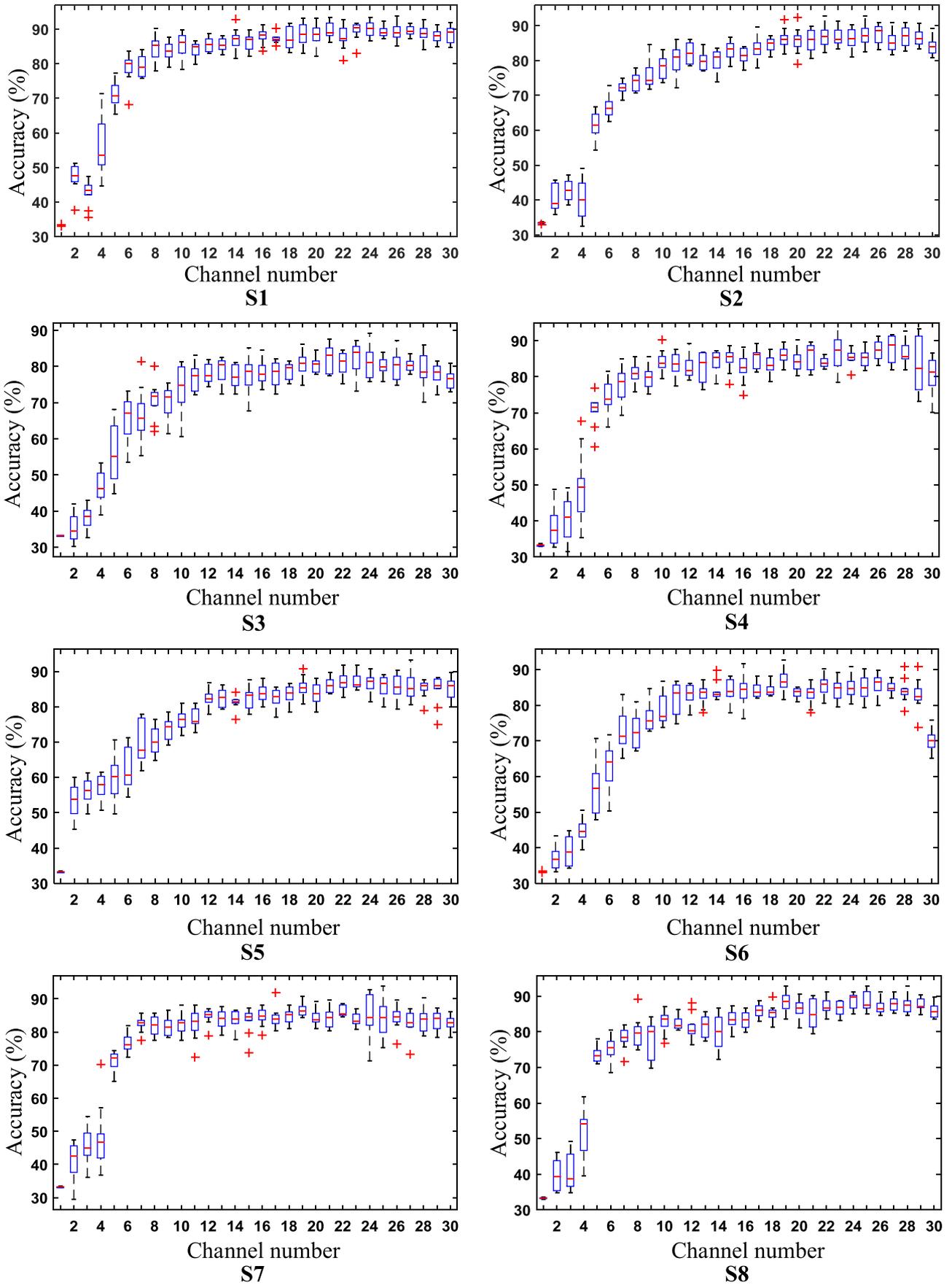
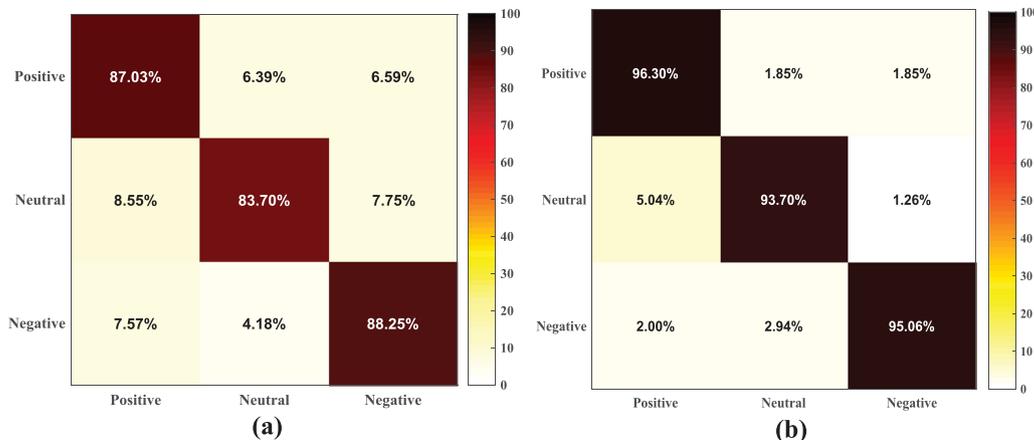


Fig. 10. Boxplots of the accuracies based on different channels for different subjects.

**Table 1**  
Average accuracies (%) and average  $F_1$ -scores (%) of the three different methods on two datasets.

Dataset	Methods					
	T_JAD		I_JAD		ICSPCS	
	Accuracy	$F_1$ -score	Accuracy	$F_1$ -score	Accuracy	$F_1$ -score
Self-collected dataset	74.92	74.84	83.04	82.75	85.85	85.78
MAHNOB-HCI dataset	81.29	80.34	92.70	91.31	94.13	93.26



**Fig. 11.** Summed confusion matrix from all the subjects: (a) dataset collected by our laboratory and, (b) MAHNOB-HCI dataset. The rows and columns in the matrix represent the predicted and real classes, respectively. The correct classifications are presented on the diagonal, and the substitution errors shown on the off-diagonal.

it gradually stabilizes as the number of channels is more than 19. Computing the distribution of the channel weights shows that, the sum of the weights from the top 19 channels is up to 85% of all channels; this suggests that the main components of the emotional cognitive activity are acquired. Furthermore, the recognition accuracy of subject S5 presents a slight decrease after applying the channel-selection strategy. If we increase the number of optimal channels to 22, the recognition accuracy reaches 86.23% (which is higher than that with 30 channels). Therefore, we speculate that the top 19 channels might not contain sufficient emotional information for subject S5 when using the ICSPCS method. Considering the ubiquity of the algorithm, recognition performance, and computational load, we finally selected the top 19 channels with the highest weight value as the observation channels.

#### 4.2.2. Experiment results on different datasets

The emotion recognition experiments were conducted under three conditions: the traditional JAD method (T\_JAD), the improved JAD method (I\_JAD), and a combination of the improved JAD and channel selection method (ICSPCS). Note that, for the T\_JAD and I\_JAD methods using the 30-channel EEG signals, and for the ICSPCS method, we selected the top 19 channels with the highest weight values  $Q$  as the observation signals. The average recognition results of the three emotional states according to the self-collected dataset of our laboratory and the MAHNOB-HCI dataset are summarized in Table 1.

As listed in Table 1, the average recognition accuracy and  $F_1$ -score vary for the two datasets. The probable reasons for this result are 1) the familiarity levels of the different subjects regarding the experiment environment as well as the differences in the response strengths to the stimulus-invoking movies, and in the self-control abilities of the induced emotions, 2) the differences in the educational backgrounds of the subjects as well as their cognitive preferences and life experiences, and 3) the differences in the data acquisition equipment applied. Closer examination of the experiment results reveals that for the self-collected dataset, the average recognition accuracies of the I\_JAD and ICSPCS methods reach 83.04%

and 85.85%, respectively, achieving remarkable corresponding improvements of 8.12% (7.91%) and 10.93% (10.94%) compared to the T\_JAD method. On the MAHNOB-HCI dataset, the average recognition results of the I\_JAD and ICSPCS methods for all the subjects reach 92.70% and 94.13%, respectively. The average recognition accuracies are improved by 11.41% and 12.84%, and the average  $F_1$ -scores are increased by 10.97% and 12.92%, respectively.

To further evaluate the performance of the proposed emotion recognition algorithm, we computed the summed confusion matrix (see Fig. 11) in the cases of positive, neutral, and negative emotional states across all subjects.

As presented in Fig. 11, the largest between-class substitution errors are 8.55% and 5.04% on the self-collected and MAHNOB-HCI datasets, respectively. Specifically, the probability of a “positive” emotional state falsely returning a “neutral” state is the highest. By analyzing the original stimuli videos and the emotional responses of the subjects after viewing the stimuli, the probable reasons for this result are: 1) the insufficient difference between the “positive” and “neutral” stimulus-invoking movies, which cannot effectively induce the corresponding emotional state, and 2) the diversity in the personalities and preferences of a subject, leading to differences in the cognition even for a particular stimulus-invoking movie. By contrast, (on the self-collected dataset) the smallest between-class substitution error (4.18%) occurs between the “neutral” and “negative” states. When inquired about their feelings, most of the subjects expressed experiencing remarkably different emotions when viewing different types of stimuli-invoking movies. Closer examination of the confusion matrix presents that the average accuracies of the “positive” (87.03% and 96.30%) and “negative” (88.25% and 95.06%) states are higher than those of the “neutral” state (83.7% and 93.7%). In fact, the strength of the brain activities will decrease because of the relaxation of the physical and psychological conditions under a “neutral” state. In such a case, the absolute difference in the values of all CSP coefficients will also reduce, resulting in a decrease in the abilities of the related sources to depict an emotion. Thus, the recognition accuracy is lower than that of the other two emotional states.

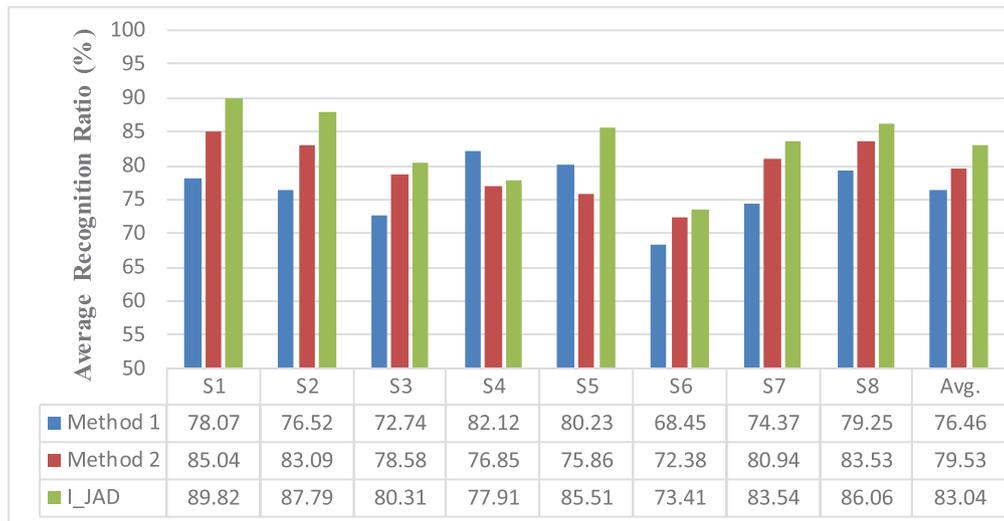


Fig. 12. Comparison of the emotional recognition classification results.

#### 4.3. Comparison experiments

To verify the effectiveness of the proposed spatial feature extracted algorithm further (regarding EEG emotion recognition), we compared the following two methods to the proposed I\_JAD method.

- Method 1: The PSD features and spectral asymmetries of the left and right sides of the brain were extracted from the delta, theta, alpha, beta, and gamma-frequency bands using the whole-channel EEG signals [13].
- Method 2: For the extension of the multi-class problem based on the JAD algorithm, if one eigenvector with the highest score was selected more than once, the next highest transformed eigenvalue was selected [33,34].

It is worth noting that the training and test datasets as well as the SVM parameters used in the current experiment were the same as those in the previous experiments. The comparison results are shown in Fig. 12.

Figure 12 shows that the average recognition accuracy of method 1 is 76.46%, which is 3.07% and 6.58% lower than the recognition accuracies of method 2 and the proposed method, respectively. This is because method 1 recognizes the emotional states by extracting the time/frequency features of the EEG signals, which are highly susceptible to noise interference. In this case, the recognition performance cannot be sufficiently ensured for noisy observation signals, because the spectra of the noises and the EEG signals may overlap. However, external noise signals, such as slight movements of the electrodes, unconscious electromyography, and impulse noises, are inevitable during the data acquisition procedure. Method 2 uses a spatial filtering algorithm to suppress the additional noises to a certain extent; however, it does not determine a spatial projection direction that maximizes the differences in the three types of emotional EEG energies. Thus, the recognition accuracy of method 2 is higher than that of method 1. Compared with methods 1 and 2, the I\_JAD method can extract the emotion-related components from the EEG signals and remove the irrelevant components and noise effectively while acquiring the largest differences between the different emotional states. Therefore, its emotion recognition accuracy is higher than those of both method 1 and method 2.

Fig. 13 depicts the time-domain waveforms and the corresponding coefficients of the CSP spatial filters for a random normal “posi-

Table 2

Relationships between the indexes of the electrodes and their positions.

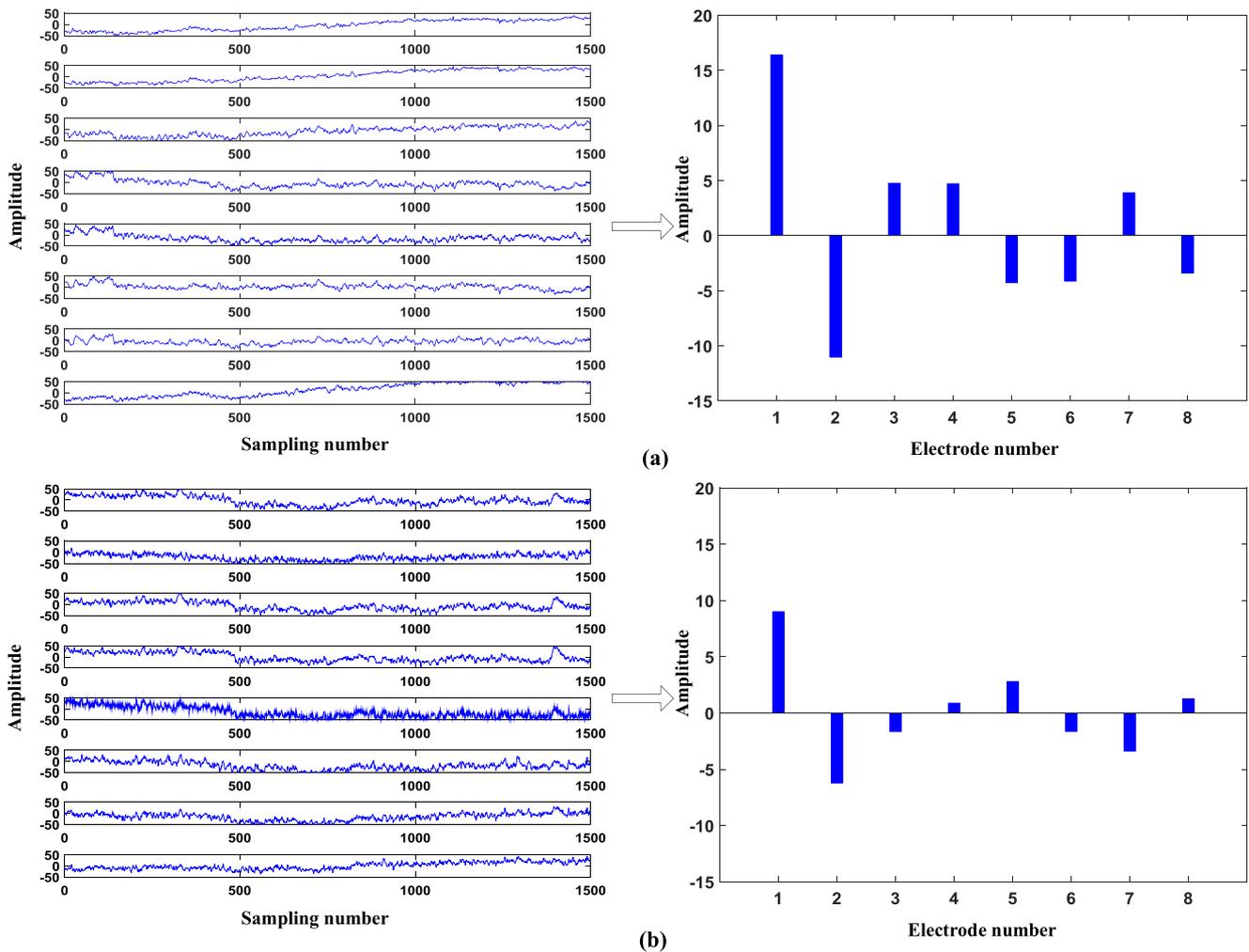
Number	Electrode position	Number	Electrode position
1	Oz	5	Fc3
2	O2	6	Fc4
3	P4	7	F3
4	Cz	8	Cp3

tive” trial and a noisy trial, respectively. For clarity, we only present the top eight channels based on descending values of the spatial coefficients. The relationships between the indexes of the electrodes and their positions are summarized in Table 2.

As can be observed in Fig. 13, the coefficients of the CSP spatial filters for the noisy trial deviate from those for the normal trial. Specifically, the absolute difference between the different electrodes corresponding to the noisy trial reduces in comparison to those of the normal trial. However, the top-two absolute coefficient values are still located on electrodes nos.1 and 2 (i.e., the Oz electrode and O2 electrode), respectively. The comparison results reveal that the CSP filters provide a certain robustness in describing the relative spatial position information of different emotion-related sources, which ensures the emotion recognition performance.

#### 5. Conclusions

The motivation of the present work is to explore the feasibility of using a CSP-based spatial filtering method to recognize emotions. To obtain a better recognition accuracy, we not only improved the traditional CSP algorithm but also developed a new channel selection strategy in terms of the characteristics of the emotional EEG signals. Recognition experiments focusing on three emotional states (i.e., positive, neutral, and negative) were conducted on the self-collected dataset of our lab and the MAHNOB-HCI dataset. The average accuracies of the proposed IC-SPCS method reached 83.01% and 94.13%, which were 8.09% and 12.84% higher than that of the traditional CSP method, respectively. The experiment results proved the effectiveness of the proposed spatial-feature based emotion recognition algorithm. As a promising method, the proposed algorithm is superior than other methods for user-dependent emotion recognition, offering a possibility for developing a high-performance emotion monitoring system. For



**Fig. 13.** Time-domain waveforms and coefficients of the CSP spatial filters of a “positive” emotional state for (a) a normal and (b) a noisy trial.

example, it can be employed in the emotion evaluation of a specific user (e.g., pilots, high-speed drivers, and soldiers) engaged in a high-risk work, which will provide a new approach to design a safety alert. In addition, it can also be used in the diagnosis and rehabilitation training of children with autism by detecting their emotional states, thereby allowing for their therapy to be adjusted in a timely manner.

Because the CSP filter has stringent requirements regarding the quality of the training data, manually detecting low-quality data generated by the fatigue of the subjects and other factors is difficult. Therefore, a future study should prioritize optimizing the training data and reduce the negative impact of “bad” trials on the recognition performance. In addition, we will make a distinct effort to address the problem of individual differences in the proposed channel selection strategy and enhance the practicability of the proposed algorithm.

### Conflicts of Interest

The authors declare no conflict of interest.

### Acknowledgements

The authors would like to thank the volunteers who participated in this study and the anonymous reviewers for their comments on the draft of this article. The authors also would like to thank Shasha Kang for providing the concept of the improved JAD

algorithm. This research work was supported by the [National Natural Science Fund of China](#) under Grant 61972437; Anhui Provincial Natural Science Research Project of Colleges and Universities Fund under Grant KJ2018A0008; Open fund for Discipline Construction, Institute of Physical Science and Information Technology, Anhui University; and Open Fund for Shanxi Key Laboratory of Network and System Security under grant NSSOF1900101. The study received approval from the Ethics Committee of Anhui University, and all the subjects signed informed consent. The authors declare no conflicts of interest.

### References

- [1] Diez PF, Müller SMT, Mut VA, Laciari E, Avila E, Bastos-Filho TF, et al. Commanding a robotic wheelchair with a high-frequency steady-state visual evoked potential based brain-computer interface. *Med Eng Phys* 2013;35(8):1155–64.
- [2] Negin F, Rodriguez P, Koperski M, Kerboua A, González J, Bourgeois J, et al. Praxis: towards automatic cognitive assessment using gesture recognition. *Expert Syst Appl* 2018;106:21–35.
- [3] Rautaray SS, Agrawal A. Vision based hand gesture recognition for human computer interaction: a survey. *Artif Intell Rev* 2015;43(1):1–54.
- [4] Cooper RA, Dicianno BE, Brewer B, LoPresti E, Ding D, Simpson R, et al. A perspective on intelligent devices and environments in medical rehabilitation. *Med Eng Phys* 2008;30(10):1387–98.
- [5] Kim M-K, Kim M, Oh E, Kim S-P. A review on the computational methods for emotional state estimation from the human eeg. *Comput Math Methods Med* 2013;2013.
- [6] Alonso JB, Cabrera J, Medina M, Travieso CM. New approach in quantification of emotional intensity from the speech signal: emotional temperature. *Expert Syst Appl* 2015;42(24):9554–64.

- [7] Wang K, An N, Li BN, Zhang Y, Li L. Speech emotion recognition using fourier parameters. *IEEE Trans Affect Comput* 2015;6(1):69–75.
- [8] Busso C, Deng Z, Yildirim S, Bulut M, Lee CM, Kazemzadeh A, et al. Analysis of emotion recognition using facial expressions, speech and multimodal information. In: *Proceedings of the 6th international conference on Multimodal interfaces*. ACM; 2004. p. 205–11.
- [9] Ali H, Hariharan M, Yaacob S, Adom AH. Facial emotion recognition using empirical mode decomposition. *Expert Syst Appl* 2015;42(3):1261–77.
- [10] Quintana DS, Guastella AJ, Outhred T, Hickie IB, Kemp AH. Heart rate variability is associated with emotion recognition: direct evidence for a relationship between the autonomic nervous system and social cognition. *Int J Psychophysiol* 2012;86(2):168–72.
- [11] Su Y, Chen P, Liu X, Li W, Lv Z. A spatial filtering approach to environmental emotion perception based on electroencephalography. *Med Eng Phys* 2018;60:77–85.
- [12] Chang C-Y, Chang C-W, Zheng J-Y, Chung P-C. Physiological emotion analysis using support vector regression. *Neurocomputing* 2013;122:79–87.
- [13] Bhardwaj A, Gupta A, Jain P, Rani A, Yadav J. Classification of human emotions from eeg signals using svm and lda classifiers. In: *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE; 2015. p. 180–5.
- [14] Zheng W-L, Lu B-L. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Trans Auton Ment Dev* 2015;7(3):162–75.
- [15] Hadjilidimitriou SK, Hadjileontiadis LJ. Toward an eeg-based recognition of music liking using time-frequency analysis. *IEEE Trans Biomed Eng* 2012;59(12):3498–510.
- [16] Wang Y, Hong B, Gao X, Gao S. Implementation of a brain-computer interface based on three states of motor imagery. In: *2007 29th annual international conference of the IEEE engineering in medicine and biology society*. IEEE; 2007. p. 5059–62.
- [17] Allison BZ, Brunner C, Altstätter C, Wagner IC, Grissmann S, Neuper C. A hybrid erd/ssvep bci for continuous simultaneous two dimensional cursor control. *J Neurosci Methods* 2012;209(2):299–307.
- [18] Lv Z, Zhang C, Zhou B, Gao X, Wu X. Design and implementation of an eye gesture perception system based on electrooculography. *Expert Syst Appl* 2018;91:310–21.
- [19] Nakamura M, Chen Q, Sugi T, Ikeda A, Shibasaki H. Technical quality evaluation of eeg recording based on electroencephalographers knowledge. *Med Eng Phys* 2005;27(1):93–100.
- [20] Saarimäki H, Gotsopoulos A, Jääskeläinen IP, Lampinen J, Vuilleumier P, Hari R, et al. Discrete neural signatures of basic emotions. *Cerebral cortex* 2015;26(6):2563–73.
- [21] Grekow J. Music emotion maps in the arousal-valence space. In: *From content-based music emotion recognition to emotion maps of musical pieces*. Springer; 2018. p. 95–106.
- [22] Soleymani M, Pantic M. Multimedia implicit tagging using eeg signals. In: *2013 IEEE International conference on multimedia and expo (ICME)*. IEEE; 2013. p. 1–6.
- [23] Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans Affect Comput* 2011;3(1):42–55.
- [24] Uhrig MK, Trautmann N, Baumgärtner U, Treede R-D, Henrich F, Hiller W, et al. Emotion elicitation: a comparison of pictures and films. *Front Psychol* 2016;7:180.
- [25] Zheng W-L, Dong B-N, Lu B-L. Multimodal emotion recognition using eeg and eye tracking data. In: *2014 36th annual international conference of the IEEE engineering in medicine and biology society*. IEEE; 2014. p. 5040–3.
- [26] Wang Y, Lv Z, Zheng Y. Automatic emotion perception using eye movement information for e-healthcare systems. *Sensors* 2018;18(9):2826.
- [27] Zheng WL, Zhu J-Y, Lu B-L. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Trans Affect Comput* 2017;10(3):417–29.
- [28] Dias NS, Carmo JP, Mendes PM, Correia JH. Wireless instrumentation system based on dry electrodes for acquiring eeg signals. *Med Eng Phys* 2012;34(7):972–81.
- [29] Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE Trans Rehabil Eng* 2000;8(4):441–6.
- [30] Gouy-Pailler C, Congedo M, Brunner C, Jutten C, Pfurtscheller G. Nonstationary brain source separation for multiclass motor imagery. *IEEE Trans Biomed Eng* 2009;57(2):469–78.
- [31] Nguyen T-H, Park S-M, Ko K-E, Sim K-B. Multi-class stationary csp for optimal feature separation of brain source in bci system. In: *2012 12th international conference on control, automation and systems*. IEEE; 2012. p. 1035–9.
- [32] Dornhege G, Blankertz B, Curio G, Müller K-R. Increase information transfer rates in bci by csp extension to multi-class. In: *Advances in neural information processing systems*; 2004. p. 733–40.
- [33] Dornhege G, Blankertz B, Curio G, Müller K-R. Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans Biomed Eng* 2004;51(6):993–1002.
- [34] Grosse-Wentrup M, Buss M. Multiclass common spatial patterns and information theoretic feature extraction. *IEEE Trans Biomed Eng* 2008;55(8):1991–2000.
- [35] Lan T, Erdogmus D, Adami A, Pavel M, Mathan S. Salient eeg channel selection in brain computer interfaces by mutual information maximization. In: *2005 IEEE engineering in medicine and biology 27th annual conference*. IEEE; 2006. p. 7064–7.
- [36] Wang Y, Gao S, Gao X. Common spatial pattern method for channel selection in motor imagery based brain-computer interface. In: *2005 IEEE engineering in medicine and biology 27th annual conference*. IEEE; 2006. p. 5392–5.
- [37] Lal TN, Schroder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N, et al. Support vector channel selection in bci. *IEEE Trans Biomed Eng* 2004;51(6):1003–10.
- [38] Zhang Y, Zhou G, Jin J, Wang X, Cichocki A. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface. *J Neurosci Methods* 2015;255:85–91.
- [39] Abibullaev B, An J, Jin S-H, Lee SH, Moon JI. Minimizing inter-subject variability in fnirs-based brain-computer interfaces via multiple-kernel support vector learning. *Med Eng Phys* 2013;35(12):1811–18.
- [40] Bos DO, et al. Eeg-based emotion recognition. *Influence Visual Auditory Stimuli* 2006;56(3):1–17.
- [41] Wang X-W, Nie D, Lu B-L. Emotional state classification from eeg data using machine learning approach. *Neurocomputing* 2014;129:94–106.
- [42] Stock M, Pahikkala T, Airola A, Waegeman W, De Baets B. Algebraic shortcuts for leave-one-out cross-validation in supervised network inference. *BioRxiv* 2018:242321.