ORIGINAL ARTICLE

Towards Improving Motor Imagery Brain–Computer Interface Using Multimodal Speech Imagery

Jigang Tong[1] · Zhengxing Xing[1] · Xiaoying Wei[1] · Chao Yue[1] · Enzeng Dong[1] · Shengzhi Du[2] · Zhe Sun[3] ·
Jordi Solé-Casals[4] 🄳 · Cesar F. Caiafa[5]

Abstract

Purpose The brain–computer interface (BCI) based on motor imagery (MI) has attracted extensive interest due to its spontaneity and convenience. However, the traditional MI paradigm is limited by weak features in evoked EEG signal, which often leads to lower classification performance.

Methods In this paper, a novel paradigm is proposed to improve the BCI performance, by the speech imaginary combined with silent reading (SR) and writing imagery (WI), instead of imagining the body movements. In this multimodal (imaginary voices and movements) paradigm, the subjects silently read Chinese Pinyin (pronunciation) and imaginarily write the Chinese characters, according to a cue.

Results Eight subjects participated in binary classification tasks, by carrying out the traditional MI and the proposed paradigm in different experiments for comparison. 77.03% average classification accuracy was obtained by the new paradigm versus 68.96% by the traditional paradigm.

Conclusion The results of experiments show that the proposed paradigm evokes stronger features, which benefits the classification. This work opens a new view on evoking stronger EEG features by multimodal activities/stimuli using specific paradigms for BCI.

Keywords Brain–computer interface · Chinese pinyin · Event-related desynchronization (ERD) · Motor imagery (MI) · Speech imagery (SI)

✉ Jordi Solé-Casals
jordi.sole@uvic.cat

✉ Cesar F. Caiafa
ccaiafa@fi.uba.ar

1 Tianjin Key Laboratory of Control Theory and Applications in Complicated Systems, Tianjin University of Technology, Tianjin 300384, China

2 Department of Electrical Engineering, Tshwane University of Technology, Pretoria 0001, South Africa

3 Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, RIKEN, Saitama, Japan

4 Data and Signal Processing Research Group, University of Vic-Central University of Catalonia, Vic, Catalonia, Spain

5 Instituto Argentino de Radioastronomía, CONICET CCT La Plata/CIC-PBA/UNLP, Villa Elisa, Buenos Aires, Argentina

1 Introduction

Brain–computer interface systems (BCIs) enable a person to communicate directly with the outside world directly by the brain signal, differentiating to the traditional neural channels [1, 2]. It can help disabled patients achieve the need to control external equipment and assist in rehabilitation training [3]. With the rapid development of electroencephalogram (EEG) technology, more attention has been focussed on EEG-based BCIs. In just a few decades, classic paradigms such as P300, steady-state visual-evoked potentials (SSVEP), motor imagery (MI), and emotion BCIs have been developed [4].

Compared with other paradigms, the event-related desynchronization (ERD) and event-related synchronization (ERS) are commonly considered in MI-BCI. Motor imagery (MI) evokes a spontaneous EEG signal, which has great research values and potentials in rehabilitation. References [5–7] have shown that both the actual movement of

limbs or the imagery of such movements similarly result in the Mu and Beta rhythms of the motor cortex in the brain attenuated or increased, in the form of ERD and ERS [8]. The MI-BCI uses these phenomena to extract features from EEG signals of different imagery tasks for classification and recognition. Pfurtscheller designed a virtual keyboard input method based on MI-BCI [9], which laid a solid foundation for the relevant research. Hochberg et al. helped patients to complete the activity of using a mechanical arm to drink by recording electrical signals from the motor cortical area of patients with cerebral palsy [10]. Yasunri Hashimoto et al. [11] combined the MI-BCI with the virtual reality system, where the individuals realized the manipulation of the avatar in the virtual reality system by imagining the movement of the left hand, right hand, and feet.

The MI-BCI brings much convenience to the disabled, but there are still many unresolved problems: (1) the EEG signal of motor imagery is spontaneous, whose non-stationarity results in significant individual diversity; (2) the task is based on motor imagery patterns that are closely related to the individuality of the subjects, and are not feasible for all people [12]; (3) the traditional MI paradigm has shortcomings, such as non-significant features and low classification performance, which affect the use of MI-BCI.

Many researchers have developed advanced machine learning algorithms to improve the performance of BCI [13–16]. These classification algorithms produce high classification accuracy but often require large training datasets. In addition, many feature extraction methods are used to differentiate signals under different tasks for more efficient classification [17–19].

However, despite advanced feature extraction and classification algorithms, the BCI cannot perform well if the subject is not able to adjust the brain activity to the BCI requirements [20]. In recent years, hybrid Brain–Computer Interface has become a hot spot in relevant research. It combines two or more types of EEG and other bioelectric signals to further improve the performance of a BCI system. It provides an attractive solution to the improvement of the overall operating capacity of users. Allison et al. [21] proposed a hybrid brain–computer interface that combined motor imagery and steady-state visual-evoked potential. The system achieved a recognition rate of 81.0%. In contrast, SSVEP achieved 76.9%, while MI had an even lower success rate (74.8%). Long et al. [22] proposed a hybrid brain–computer interface using P300 visual-evoked features and imaginary movement ERD features, which also showed a very good application prospect in target object selection applications. Yao et al. [23] combined motor imagery with selective sensation to enhance the discrimination between left and right mental tasks. In the study by Kim et al. [24], features were extracted from three different patterns, including RPs, the ERD/ERS, and the ERP, and the results showed

that the method had high accuracy. Zhuang et al. [25] used a combination of wavelet and canonical correlation analysis to reveal the ERD/ERS pattern and combine multiple learning algorithms in a classifier to obtain the best learning results. Lu & Bi [26] designed a longitude-based algorithm to control the speed of simulated vehicles. In this algorithm, Common Spatial Pattern (CSP) was used to enhance the SNR of EEG signals, and then PSD features were extracted and SSVEP patterns were classified using traditional RBF kernel support vector machine (SVM) [27].

References [28–31] have shown that the generation of EEG signals from speech imagery (SI) is related to different brain areas, such as left inferior frontal gyrus, supplementary motor area, premotor area, superior temporal gyrus, and middle temporal gyrus. Bocquelet et al. [32] believed that the advanced processing of speech imagery was related to the left hemisphere. Ikeda et al. [33] studied the electrocorticography (ECoG) signals of vowel speech imagery and found that the alpha band (8–13 Hz) and beta band (13–30 Hz) may contain information for decoding single vowel speech imagery activities. Research [34] conducted by Sereshkeh et al. using EEG signals showed that during speech imagery (SI) activities, the EEG channel activation level increases in the 20–30 Hz frequency band in the Broca area and the band below 15 Hz in the Wernicke area. From these studies, one finds that SI activities generate EEG in a similar manner to right-hand MI. Tong et al. analysed the EEG under Chinese phoneme imagination and MI tasks and the result was better than traditional MI [35]. Therefore, the SI activities can be used to enhance the alpha and beta bands of the EEG signals generated on the left cortex during MI. This paper aims to conduct in-depth analysis on this hypothesis, as well as to improve the performance of the MI-BCI through SI. A novel paradigm using SI to enhance MI-BCI is proposed to solve the problems of non-significant feature differences and lower classification accuracy in the traditional MI paradigm.

The content of this paper is organized as follows. Section 2 introduces the proposed paradigm and methods used, and the conduction of offline data collection experiments; the third section gives the results of offline data analysis and classification; Sect. 4 discusses the advantages and disadvantages of the proposed paradigm, while Sect. 5 summarizes this paper.

2 Methods

2.1 Participants

This experiment was approved by the Ethical Committee of the Tianjin University of Technology. Eight healthy and right-handed subjects (2 females and 6 males, 22–26 years

old) participated in our experiments, labelled as S1–S8. All the subjects were native Chinese speakers, with normal or corrected-to-normal vision and had never participated in relevant experiments before. They had been informed and consented to participate in the experiments. The average duration of the experiment for each subject (including preparation) was 2.5 h.

2.2 Experimental Paradigm

The proposed paradigm aims to improve the traditional right-handed MI task by combining it with the SI task, that is, the subjects need to silently read Chinese Pinyin (pronunciation of Chinese characters) and imagine writing the corresponding Chinese characters with the right hand. These tasks of speech imagery are denoted as Speech and Write Motor Imagery (SW-MI) in the paper, to differentiate itself to the Left-Hand Motor Imagery (LH-MI) and Right-Hand Motor Imagery (RH-MI). In order to verify the effectiveness of the proposed paradigm while minimizing the impacts caused by time-varying features of EEG, the SW-MI, LH-MI, and RH-MI tasks are mixed randomly for comparison.

During the preparation process prior to an experiment, subjects sit quietly in a comfortable chair, with their hands resting on the armrests, looking at the screen (60 Hz refresh rate, 1920 × 1080 resolution) about 60 cm far, and waiting for the experiment to start. The detailed flowchart of the experimental paradigm is shown in Fig. 1.

Each trial is divided into three temporal parts. The first part is the setup. The "cue1" instruction is displayed as a cross, lasting 3 s, in the centre of the screen, informing the subject that the experiment is about to start. After the cross disappears, the second part starts with "cue2", showing one of the following images: "←", "→", or "yòu shǒu" (the pronunciation of "right hand" in Chinese), representing LH-MI, RH-MI, and SW-MI, respectively. All cues are displayed in the same colour and size in the centre of the screen, to minimize the effects of eye movements and vision stimuli to colours. In the SW-MI, the subject imagines the

Pinyin and imagines writing the Chinese character using the right hand at the same time. The subject performs imaginary activities according to the cue for 5 s. Then, the last part of "cue3" stops the imagination tasks by displaying the "●" in the centre of the screen, reminding the subject that the imagination task is finished and that they should enter a relaxed state in which they are allowed to blink, take deep breaths, or engage in other activities. One complete experiment consists of 30 trails, ten times for each type of tasks. Each subject conducts seven experiments on the same day, a total of 210 imagination tasks. Each subject is asked verbally at the end of the experiment whether he/she has conducted the experiment seriously and whether he/she has committed any infractions, such as sudden body movements or inattention during the experiment.

2.3 EEG Recording

A SynAmps2 amplifier (Neuroscan, version 4.5) was used to acquire the EEG signals. As shown in Fig. 2, a total of 64 electrodes were used. The electrodes were placed according to the international 10–20 systems. Throughout experiments, the impedances of all channels were kept below 5000 Ω. The sampling frequency of the system was set to 1000 Hz, and the bandpass filter was set to 0.1–100 Hz. After the experiments, the recorded data were saved as a '.cnt' format file for subsequent analysis and processing.

It was commonly noted that even a tiny action, such as scalp movement, frowning, biting, swallowing, and so on, can generate electromyography recorded by electrodes attached on the skin. It is difficult to separate the required EEG signals from the superimposed electrical signals, therefore, it is necessary to remove these electromyographic artefacts. In this work, independent component analysis (ICA) was applied to identify the artefactual components of the EEG signal, so that most of the artefacts were isolated and eliminated.

For a better signal data quality, it is necessary to avoid disturbances and noises as much as possible. Therefore,

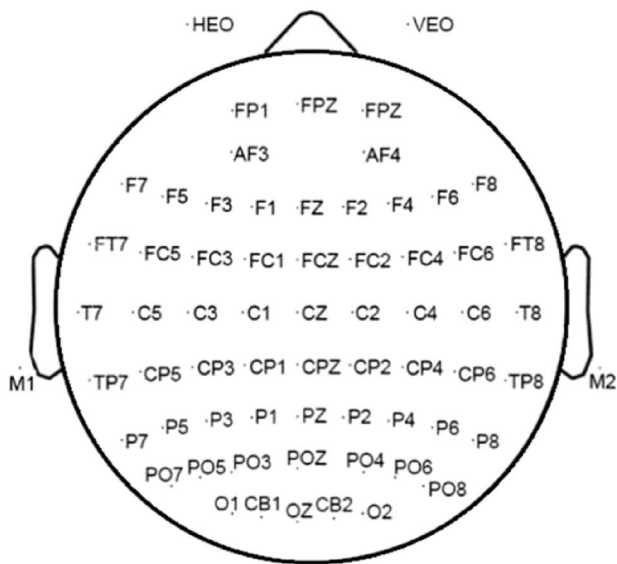**Fig. 1** Time course of the proposed experimental paradigm
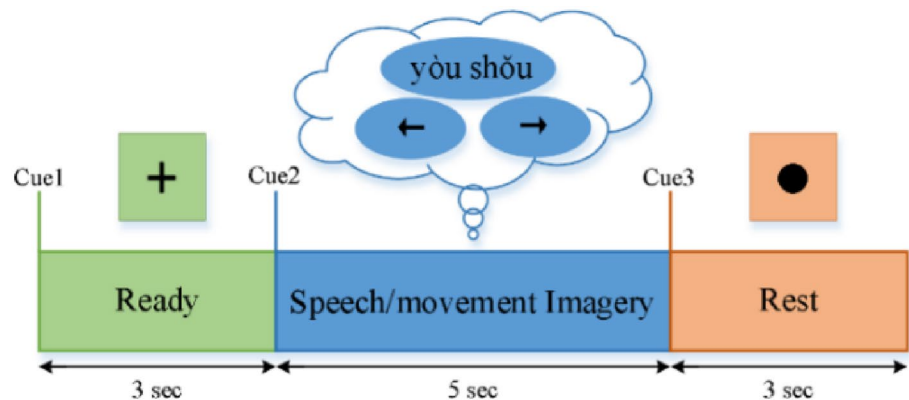
Fig. 2 Location of the 64 electrodes in the experiments

Fig. 3 EEG signal processing flowchart

three experts observed the subjects through visual inspection. If the subject exhibited excessive action or inattention during the experiment, making the experimental data difficult to process, then the experiment was conducted again to ensure the artefactual noise was eliminated.

In addition, a questionnaire was completed by the subject after the experiment to ensure that the subject performed the speech imagery according to the clues. The experimental data were processed using MATLAB software and NVIDIA GeForce GTX 1650 GPU (8 GB).

2.4 Implementation of the Proposed Method

For each subject, the recorded data were first pre-processed before being split into three categories according to the labels corresponding to the three different tasks. Afterwards, the time–frequency analysis was applied to processed data. Finally, in the classification step, two types of signals were used to build a set of spatial filters that were used to extract features for SVM classification. These steps are shown in Fig. 3 and detailed in the next section.

2.4.1 Preprocessing

The recorded EEG data were pre-processed by EEGLAB [36]. First, the multi-channel EEG signals were re-referenced using the bilateral mastoid electrodes as reference. Second, 6–30 Hz bandpass filtering was performed to remove baseline drift, power frequency interference, and EMG artefacts. Third, sampling was downsampled to 250 Hz and independent component analysis (ICA) [37] was applied to remove electrooculographic artefacts from the EEG signals. To
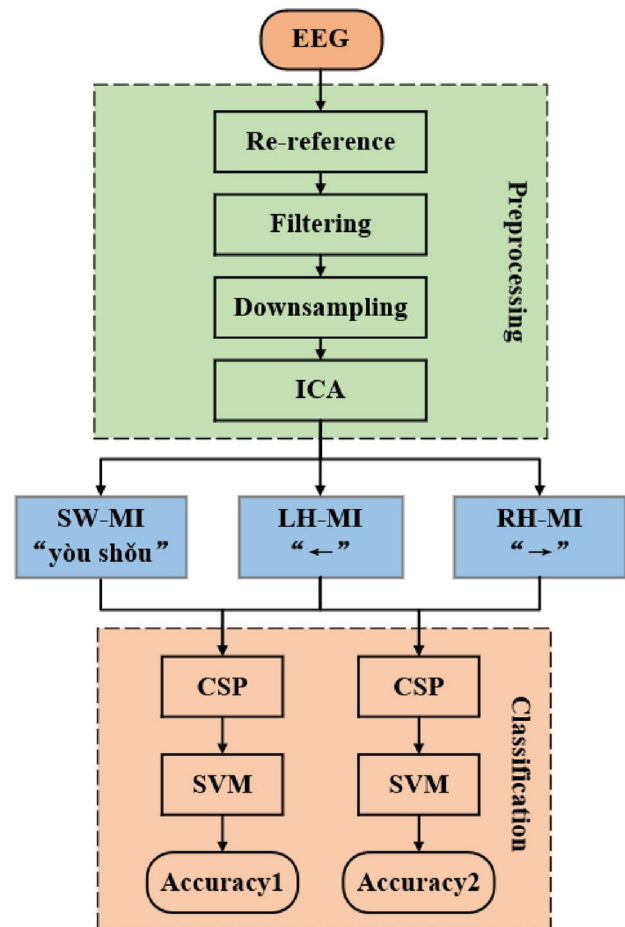
preserve data characteristics as much as possible, EEG data were standardized. Z-Score uses the formula of $(x- \mu)/\sigma$ to convert two or more groups of data into unitless Z-Core scores to unify data. Finally, EEG signals were obtained for the three types of tasks.

2.4.2 Signal Analysis

EEG signals are non-stationary and time-varying and cannot be fully analysed in the time or frequency domain alone. Therefore, the time–frequency analysis method is adopted in which the time–frequency plot is used to observe the evoked ERD in RH-MI and SW-MI tasks.

Time–frequency analysis, also known as joint Time–Frequency Analysis, is commonly used for processing and analysing EEG signals. The basic idea of this method is to design a joint function of time and frequency to simultaneously describe the energy density at different times and frequencies. Using the time–frequency distribution of the signals, one gets the instantaneous frequency components and their amplitudes

at a given time point, which can be used to analyse time-varying signals.

The short-time Fourier transform is a well-known time–frequency distribution function, where short data segments are selected by conjugating the signal with a sliding window, then the Fourier transform is applied on each short segment separately. The definition of short-time Fourier transform is shown in Eq. (1):

$$\text{STFT}(t,f) = \int_{-\infty}^{+\infty} x(\tau) W^*(\tau - t) e^{-j2\pi f\tau} d\tau, \tag{1}$$

where f is the frequency, $x(\tau)$ is the EEG signal of one trail, $W(t)$ is the observation window, $W(\tau - t)$ is obtained by shifting the time observation window to the right to any time point t, $W^*(\tau - t)$ is the complex conjugate function of $W(\tau - t)$, and $x(\tau)W(\tau - t)$ is the signal obtained after windowing the original signal.

In this paper, to improve the spectrum leakage of rectangular window, the hamming window with t=2 s was used. In order to explore the activation of the cerebral motor cortex under different tasks, event-related spectrum disturbances were calculated using short-time Fourier transform. Event-related spectrum disturbances are defined as changes in event-related spectrum power relative to the previous baseline or reference period. The event-related spectrum disturbance can be calculated as in Eq. (2):

$$\text{ERSP}(f, t) = \frac{1}{N} \sum_{k=1}^{N} \left(F_k(f, t)^2 \right), \tag{2}$$

where N represents the number of trials, and $F_k(f, t)$ represents the frequency spectrum estimation at the frequency f at the time t of the k-th experiment.

Considering the time difference between subjects seeing the motor imagery task clue and performing the motor imagery, data from 0.5 s to 5 s in the process of motor imagery are selected for classification.

In EEG signal processing, CSP is a well-known and effective feature extraction algorithm [38]. The basic idea of CSP is to find the optimal direction of EEG signal spatial projection by constructing a set of spatial filters [39]. CSP is suitable for binary classification. For the two classes of EEG signals, single trail EEG data can be expressed as $N \times T$ matrix X_i, where N is the number of channels, T is the number of sampling points for each epoch, i (1 or 2) represents the i-th class. The sample data of the two classes are categorized as X_1 and X_2, respectively.

For the segmented raw data, the covariance matrix R_i is obtained as shown in Eq. (3):

$$R_i = \frac{X_i X_i^T}{\text{trace}\left(X_i X_i^T\right)} (i = 1, 2), \tag{3}$$

where trace(X) is the trace of the matrix X, and i represents the i-th class. R_i is the expectation of the spatial covariance matrix of the sample data of the i-th class. The sum of the spatial covariance matrices of the two classes is calculated in Eq. (4):

$$R_c = R_1 + R_2, \tag{4}$$

where R_c is a positive definite matrix, which is obtained from the singular value decomposition theorem

$$R_c = U_C \wedge_C U_C^T, \tag{5}$$

where $\wedge_c$ is the corresponding diagonal matrix of eigenvalues of R_c sorted in the descending order and U_c is the corresponding eigenvector matrix.

By the whitening conversion on U_c, one obtains the whiten eigenvector matrix P in Eq. (6):

$$P = \frac{1}{\sqrt{\wedge_c}} \cdot U_c^T. \tag{6}$$

Applying matrix P to R_1 and R_2, one gets the matrices S_1 and S_2 as shown in Eq. (7):

$$S_1 = PR_1P^T, S_2 = PR_2P^T. \tag{7}$$

S_1 and S_2 have common eigenvectors, so there exist two diagonal matrices $\wedge_1, \wedge_2$, and eigenvector matrix B, which meet the following conditions:

$$S_1 = B \wedge_1 B^T, S_2 = B \wedge_2 B^T, \wedge_1 + \wedge_2 = I, \tag{8}$$

where I is the identity matrix, which means that the sum of the characteristic values of S_1 and S_2 is equal to unity.

In the eigenvector matrix B, if one class gets greater eigenvalue than the other, then the data correspond to the class with the larger eigenvalue. Therefore, matrix B can be used to solve the two-class classification problem, by means of the projection matrix W in Eq. (9):

$$W = \left(B^T P\right)^T. \tag{9}$$

The original EEG signal X can be projected through the matrix W to obtain the characteristic matrix in Eq. (10), where the dimensions of each matrix are detailed for clarification.

$$Z_{N\times T} = W_{N\times N} X_{N\times T}. \tag{10}$$

The first m rows and the last m rows ($2m < M$) of Z are selected as the characteristics of the original input data. Then, the final feature is obtained using Eq. (11):

$$y_i = \log\left(\frac{var(Z_i)}{\sum_{n=1}^{2m} var(Z_n)}\right), \tag{11}$$

where y_i is the normalized characteristic matrix of the i-th sample.

2.4.3 Classification

To properly estimate the classification accuracy, an SVM model was used to classify the feature vectors of the EEG signals. SVM is a machine learning method based on statistical learning theory [40]. It seeks to minimize structured risk to improve the generalization ability of the model, and to minimize the experience risk and confidence range. In the case of low-dimensional samples, the purpose of good statistical law can also be obtained by using kernel functions to map the input vector to a high-dimensional space and construct an optimal classification function that approximates the classification hyperplane. Here, we do not optimize the hyper-parameters specifically, but use the default parameters for classification. The effectiveness of the SVM model depends on the kernel function and the penalty factor. The kernel function determines the spatial distribution of the training samples, and the penalty factor controls the proportion of the empirical risk and the confidence range. The kernel function used in this paper is the Gaussian Kernel. The grid search method of 10×10 cross-validation is used to find the optimal parameters g and c, by searching in the range 2^{-10}–2^{10}.

3 Results

In this paper, the proposed method is used to demonstrate the potential of the multimodal paradigm in improving the separability of EEG features for BCI. The EEG data of 8 subjects (S1–S8) are used to validate the hypothesis that the speech imagery improves the EEG BCI performance based on traditional MI paradigm.

Because the imagination process designed in our experiments is time-consuming, and the performance of each subject in each experiment is different to some extent, the generation of ERD/ERS phenomenon may be different over the time. Therefore, in this paper, in order to better observe the ERD/ERS phenomenon in each experiment, a visual inspection of the 5 s of imagination process of a complete experiment is performed from the energy perspective. Figure 4 shows the corresponding Event-Related Spectral Perturbation (ERSP) plots of the right-hand tasks in two paradigms, for subjects S2 and S7. As the speech imagery task is added to the new paradigm, the C3 channel of the sensorimotor area and the FC3 channel of Broca's area are selected for analysis. As shown in Fig. 4, before $t = 0$, the subject is in resting state. The time $t = 0$ corresponds to the moment when the "cue2" appears in one trail, indicating that the subject started the imagery task. When the subject performed SW-MI and RH-MI tasks, the energy of the electrode channels C3 and FC3 in the alpha and beta bands decreased from the baseline level (as shown in blue in Fig. 4), due to the ERD phenomenon. Compared with RH-MI (the second row in Fig. 4), the ERD produced by SW-MI (the first row) is significantly stronger over a wider frequency band and lasts longer, especially in the alpha band, which means the SW-MI is more suitable for classification.

According to the ERSP diagrams, the 6–30 Hz frequency band is selected as the characteristic segment for filtering. The spatial pattern is extracted using the CSP method. Figure 5 shows the topographic maps of the two most important spatial patterns obtained, which clearly show the spatial
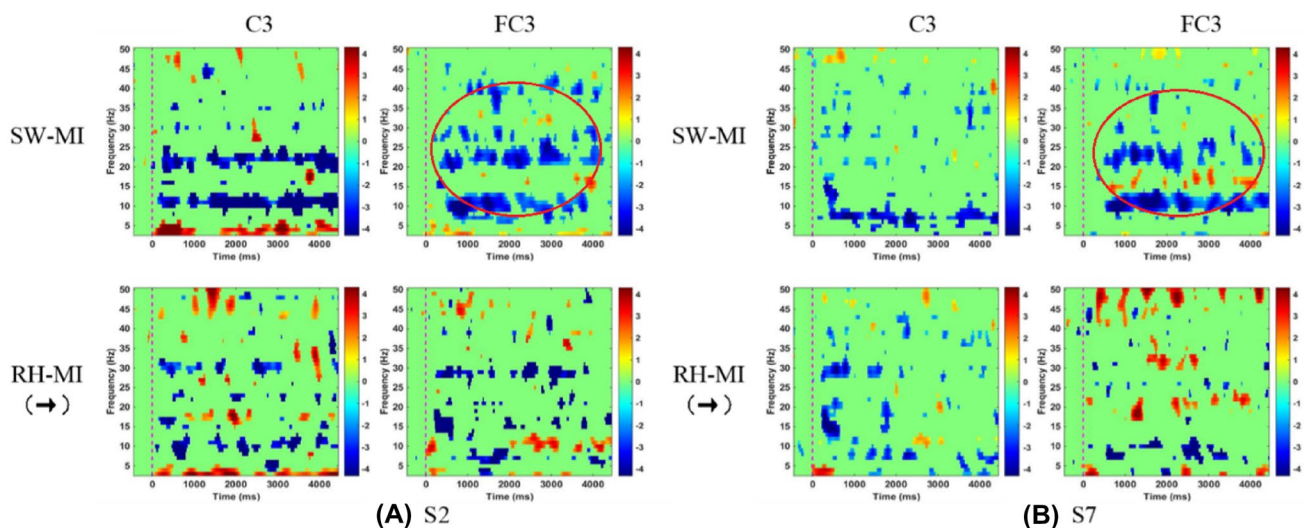

Fig. 4 ERSPs of SW-MI and RH-MI in channels C3 and FC3 for subject S2 and S7
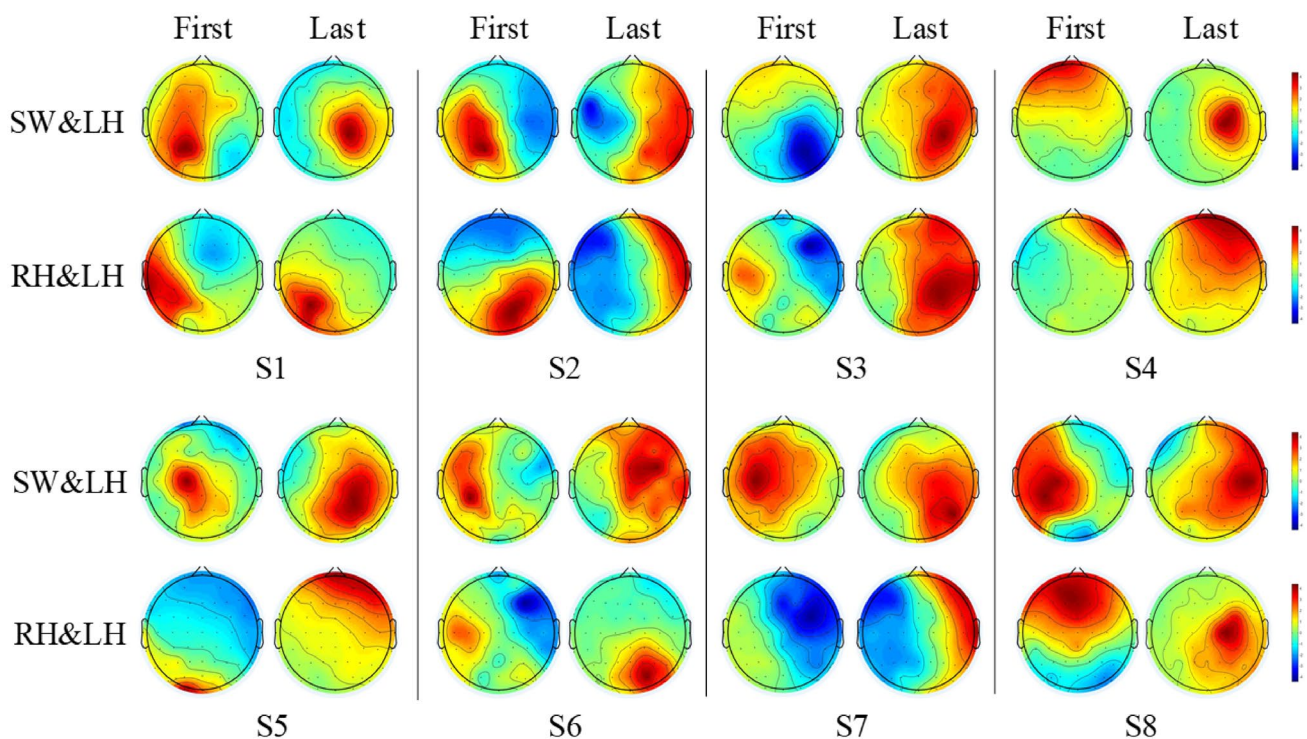
Fig. 5 Topographic map of eight subjects after CSP

pattern of the brain area when each subject performs motor imagery. The two pictures in the first row of each subject are from the first column and the last column of the feature matrix in the common spatial pattern extracted by the spatial filter constructed on the SW-MI and LH-MI-EEG signals, respectively. The second row comes from RH-MI and LH-MI. From the topographic maps of SW-MI and LH-MI, the spatial difference between the two tasks is observed.

The difference is not limited to the sensory-motor areas on both sides of the brain, but there is a larger activated area for SW-MI. This shows that, during 70 speech imagery tasks, subjects have a stronger ERD for SW-MI, thus differentiating it from hand MI. On the other side, the spatial patterns of left-hand and right-hand motor imagery of most subjects under the traditional paradigm are similar, with small differences observed in the feature space. This proves that, with the proposed paradigm, the spatial discrimination between classes is higher, which is helpful to improve the classification accuracy.

Figure 6 shows the classification accuracy of each subject calculated using the 10×10 cross-validation method, where the 'mean' represents the average classification accuracy of 8 subjects, which is 68.96% under the traditional paradigm versus 77.03%, under the new paradigm with an increase of 8%. The highest classification accuracy comes from the subject S6 between SW and LH, which is 91.29%. At the same time, the classification accuracies of the S2 and S7 have

increased by 12.0% and 13.5% under the proposed paradigm compared with the traditional paradigm, respectively. From Fig. 6 it can be clearly seen that our proposed paradigm (SW&LH) obtains higher classification accuracies than the traditional MI paradigm (RH&LH).

To further analyse the differences in classification accuracy between the two paradigms, the data were classified in the alpha and beta frequency bands according to the frequency band energy changes in the ERSP chart and the spectrum chart. A tenfold cross-validation method was adopted, and the results are shown in Table 1.

For the alpha band, the classification accuracy of 5 subjects in the traditional paradigm is less than 70%. In MI-EEG signal classification, 70% is a standard result [41], where the classification accuracy rate below 70% is not conducive to perform BCI. In contrast, using the proposed paradigm, the classification accuracy of all subjects' EEG signals in the alpha band exceeds 70%, which meets the criteria for BCI. This is an encouraging result, suggesting that the proposed paradigm is helpful.

The beta band features are also considered for the classification. As shown in Table 1, the classification accuracy of most of the subjects using the alpha band is significantly better than that of the subjects using the beta band, except for subject S3, demonstrating that the use of alpha band EEG data for classification is a good choice for the paradigm proposed in this paper. From the same table, the classification

Fig. 6 The average classification accuracy of subjects in the two paradigms ($p < 0.01$)
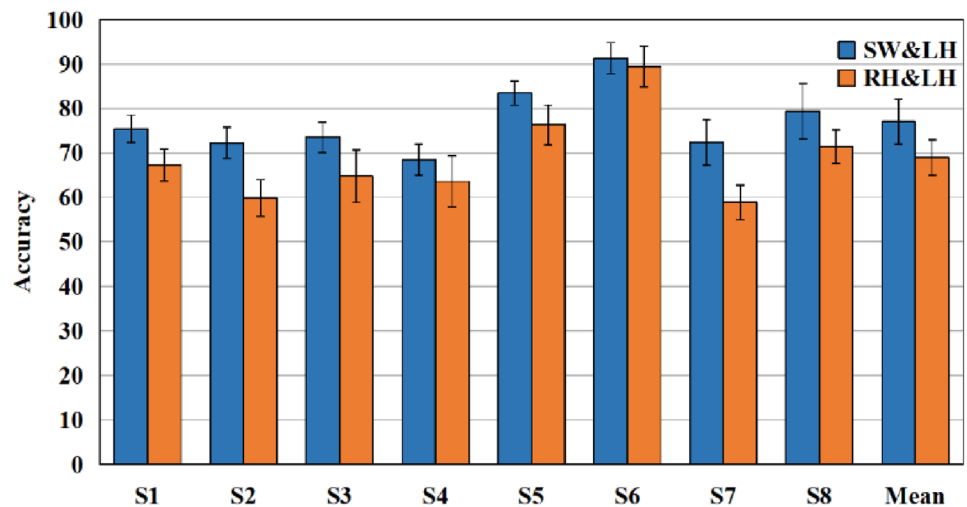

Table 1 Classification accuracy of two paradigms in two bands

Subject	Alpha		Beta	
	SW&LH	RH&LH	SW&LH	RH&LH
S1	76.32	68.34	71.21	65.56
S2	73.45	58.36	68.43	53.95
S3	70.46	62.45	71.57	66.57
S4	70.53	66.58	65.49	60.54
S5	83.51	75.77	78.64	73.78
S6	91.34	89.43	89.91	85.08
S7	73.87	58.37	70.52	55.95
S8	79.65	73.85	75.36	69.39
Mean	77.39	69.14	73.89	66.35

accuracies in beta band are higher than the traditional paradigm, which further proves the effectiveness of the proposed paradigm.

4 Discussion

This paper proposed a novel paradigm using Chinese speech imagery to enhance the EEG features in MI-BCI. Compared to the traditional motor imagery paradigm, the proposed paradigm generated clearer spatial features among classes, therefore, leading to higher classification accuracy than the traditional paradigm.

4.1 Speech Imagery Enhances the ERD in BCI

This research demonstrated that it is feasible to merge speech imagery and motor imagery in BCI by silently reading words relevant to specific classes while imagining the writing activities of the words. Imagery on writing activities awakens the subject's hand sensation from semantic

understanding, as suggested in the literature [42, 43], thus activating the sensory-motor area in the brain resulting in a significant ERD.

Since the MI-BCI is based on the ERD/ERS phenomenon of the brain [44], when the subject imagines left-hand or right-hand movement, the opposite side of the brain will produce energy changes in the Mu rhythm. In the proposed paradigm, the activated brain cortex areas of speech imagination are mainly concentrated in the Broca area and Wernicke area [45], which are close to the motor area. Due to the connectivity of the brain, when the subject imagines speech, the motor area of the brain will also show energy changes, and this change is similar to the ERD phenomenon in motor imagery. Therefore, when the subjects simultaneously perform motor and speech imagery, the brain will be affected by these two tasks at the same time, which is supported by the energy differences in Figs. 4 and 5. It can be seen from the ERSP diagram in Fig. 4 that, under the proposed paradigm, the motor area of the subject induced a stable and long-lasting ERD phenomenon.

4.2 Languages Used in Speech Imagery may Make Differences

It should be taken into account that each person has different ability on speech understanding. In the literature, most of the BCIs based on speech imagery have been developed for English users. Due to the significant differences in language structure and pronunciation, it is worthy to study the BCI based on Chinese speech imagery by analysing the influence of Chinese Pinyin and semantic information on motor imagery. Through the proposed paradigm, the Chinese silent reading and writing imagination were added to the tasks conducted in BCI as a form of multimodal imagery activities. Other multimodal activities can also be considered, for example, integrating different sensory imagining tasks into

the traditional MI-BCI, towards higher separability among classes.

4.3 Complex Movement Imagination Helps the Concentration with Less Training

An interesting point deserving attention is that, after each trial, we asked the subject's experiences. Although the proposed SW-MI task is more complicated, most subjects stated that when performing SW-MI it is easier to maintain their concentration. However, when performing the other two easier tasks (RH-MI and LH-MI), attention is easily distracted and susceptible to interference.

It is commonly recognized in literature that training is necessary for subjects to achieve effective motor imagination. Many subjects do not know how to imagine the hand movements, and it will take a long time to train subjects to achieve good results. However, when performing the SW-MI, the subjects will involuntarily concentrate on the tasks of silent reading and writing imagination, which improves the quality of the signal to a certain extent. This point is supported by the literature [46], which reported that imagining complex movements can activate the motor area of the brain more than simple movements, which explains why the proposed paradigm is more effective.

4.4 Future Work

The combination of speech and motor imagery, if fully developed, will further improve BCI performance, which opens up interesting topics for future research. In addition, it will also be worth studying how to use multiple types of imagery activities in this new paradigm, and how to accurately convert the corresponding EEG classification into multiple control commands for controlled devices.

In this paper, 64 electrode channels were used, which may cause the actual use of the BCI to be cumbersome. In the future, the choice of electrode channels will be further optimized.

At the same time, the subjects in this work are all healthy people, but the main targeting population of the existing MI-BCI system are people with disabilities [47]. The decoding method of healthy subjects' motor imaging tasks may not necessarily be the same as such patients. The SW-MI tasks need to involve speech imaging, so the proposed paradigm may not be suitable for subjects with diseases resulting in language disorders and atresia syndrome. Therefore, in future research, it is necessary to recruit subjects from different population categories, not only from the healthy group.

This work only uses EEG as an example to demonstrate the effectiveness of the proposed paradigm. A wide range of techniques based on the proposed paradigm, e.g. functional magnetic resonance imaging (fMRI) or others, can also be used to conduct in-depth discussions on brain activities in speech imagination and other perspectives.

The CSP method is used in the paper to demonstrate the potential of the proposed paradigm on improving the BCI performances. However, the single use of the CSP method still has some limitations and further research could test more feature extraction and classification algorithms.

5 Conclusion

In this work, we proposed a novel BCI paradigm combining motor imagery and speech imagery, in which the SW-MI tasks are proven to be superior to traditional hand motor imagery. The experimental results of eight subjects provided evidence for the interaction between speech imagery and motor imagery and showed that speech imagery is helpful for MI-BCI classification.

According to the subjects' oral comments, it is easier to maintain concentration in this new paradigm, and it is less boring. Therefore, this new paradigm not only improves the classification accuracy, but also helps new users to adapt to the BCI system faster and more efficiently. In line with the traditional paradigm, this experimental paradigm does not require any execution of the action, making it suitable for cases where physical movements are not feasible.

However, the work has limitations and can be improved in the future. The most important limitation is the number of subjects available in the experiments ($n = 8$). For a priori assumption on a large effect size (Cohen's d of 0.8), the error probability (alpha) set at 0.05 and a false-negative rate (beta) set at 0.2 (i.e. the power of 0.8), a minimum sample size of 15 subjects is needed for the paired t test. Another limitation is the language used in the experiments, so the results may not be general for other languages with different characteristics. In addition, we used the relatively simple CSP algorithm to conduct experiments, and other methods should be explored. These limitations point out the future research.

Data Availability Data are available from the authors upon reasonable request.

Declarations

Conflicts of interest No conflict of interest exists.

Ethical Approval All procedures performed in the study involving human participants were in accordance with the ethical standards of the Institutional Review Board at Tianjin University of Technology and with the 1964 Helsinki declaration and its later amendments.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Chaudhary, U., Birbaumer, N., & Ramos-Murguialday, A. (2016). Brain-computer interfaces for communication and rehabilitation. *Nature Reviews Neurology, 12*, 513–525. https://doi.org/10.1038/nrneurol.2016.113

2. Yuan, H., & He, B. (2014). Brain-computer interfaces using sensorimotor rhythms: Current state and future perspectives. *IEEE Transactions on Biomedical Engi-neering, 61*, 1425–1435. https://doi.org/10.1109/TBME.2014.2312397

3. Horki, P., Solis-Escalante, T., Neuper, C., & Gernot, M. (2011). Combined motor imagery and SSVEP based BCI control of a 2 DoF artificial upper limb. *Medical and Biological Engineering and Computing, 49*, 567–577. https://doi.org/10.1007/s11517-011-0750-2

4. Abiri, R., Borhani, S., Sellers, E. W., Jiang, Y., & Zhao, X. (2019). A comprehensive review of EEG-based brain-computer interface paradigms. *Journal of Neural Engineering, 16*, 1–21. https://doi.org/10.1088/1741-2552/aaf12e

5. Martin, L., & Ulrike, H. (2006). Motor imagery. *Journal of Physiology-Paris, 99*, 386–395. https://doi.org/10.1016/j.jphysparis.2006.03.012

6. Gu, L., Yu, Z., Ma, T., Wang, H., & Fan, H. (2020). EEG-based classification of lower limb motor imagery with brain network analysis. *Neuroscience, 436*, 93–109. https://doi.org/10.1016/j.neuroscience.2020.04.006

7. Hamedi, M., Salleh, S. H., & Noor, A. M. (2016). Electro-encephalographic motor imagery brain connectivity analysis for BCI: A review. *Neural Computation, 28*, 999–1041. https://doi.org/10.1162/NECO_a_00838

8. Pfurtscheller, G. (2003). Induced oscillations in the alphaband: functional meaning. *Epilepsia, 44*, 2–8. https://doi.org/10.1111/j.0013-9580.2003.12001.x

9. Neuper, C., & Pfurtscheller, G. (2001). Event-related dy-namics of cortical rhythms: Frequency-specific features and functional correlates. *International Journal of Psychophysiology, 43*, 41–58. https://doi.org/10.1016/S0167-8760(01)00178-7

10. Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., & Donoghue, J. P. (2012). Reach and graspby people with tetraplegia using a neurally controlled robotic arm. *Nature, 485*, 372–375. https://doi.org/10.1038/nature11076

11. Kakui, T., Hashimoto, Y., Ushiba, J., Liu, M., & Ota, T. (2018). Development of rehabilitation system with brain computer interface for subacute stroke patients. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. https://doi.org/10.1109/SMC.2018.00018

12. Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., & Pfurtscheller, G. (2003). How many people are able to operate an EEG-based brain-computer interface (BCI)? *Neural Systems and Rehabilitation Engineering IEEE Transactions on, 11*(2), 145–147. https://doi.org/10.1109/TNSRE.2003.814481

13. Usama, N., Leerskov, K. K., Niazi, I. K., Dremstrup, K., & Jochumsen, M. (2020). Classification of error-related potentials from single-trial EEG in association with executed and imagined movements: A feature and classifier investigation. *Medical and biological engineering and computing, 58*, 2699–2710. https://doi.org/10.1007/s11517-020-02253-2

14. Gao, L., Cheng, W., Zhang, J., & Wang, J. (2016). EEG classification for motor imagery and resting state in BCI applications using multi-class Adaboost extreme learning machine. *Review of Scientific Instruments, 87*(8), 216–219. https://doi.org/10.1063/1.4959983

15. Dong, E., Zhou, K., Tong, J., & Du, S. (2020). A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification. *Biomed and Signal Process and Cntrol, 60*, 101991. https://doi.org/10.1016/j.bspc.2020.101991

16. Enzeng, D., Guangxu, Z., Chao, C., Jigang, T., Yingjie, J., & Shengzhi, D. (2018). Introducing chaos behavior to kernel relevance vector machine (RVM) for four-class EEG classification. *PLoS One, 13*(6), e0198786. https://doi.org/10.1371/journal.pone.0198786

17. Li, M. A., Han, J. F., & Yang, J. F. (2021). Automatic feature extraction and fusion recognition of motor imagery EEG using multilevel multiscale CNN. *Medical and Biological Engineering and Computing, 59*(10), 2037–2050. https://doi.org/10.1007/s11517-021-02396-w

18. Park, S. H., Lee, D., & Lee, S. G. (2018). Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26*, 498–505. https://doi.org/10.1109/TNSRE.2017.2757519

19. Li, M. A., Luo, X. Y., & Yang, J. F. (2016). Extracting the Nonlinear features of motor imagery EEG using para-metric t-SNE. *Neurocomputing, 218*, 371–381. https://doi.org/10.1016/j.neucom.2016.08.083

20. Qiu, Z., Allison, B. Z., Jing, J., Yu, Z., Wang, X., Wei, L., & Andrzej, C. (2017). Optimized motor imagery paradigm based on imagining Chinese characters writing movement. *IEEE Transactions on Neural Systems and Rehabilitation Engi-neering, 25*, 1009–1017. https://doi.org/10.1109/TNSRE.2017.2655542

21. Allison, B. Z., Brunner, C., Kaiser, V., Müller-Putz, G. R., Neuper, C., & Pfurtscheller, G. (2010). Toward a hybrid brain-computer interface based on imagined movement and visual attention. *Journal of Neural Engineering, 7*(2), 26007. https://doi.org/10.1088/1741-2560/7/2/026007

22. Long, J., Li, Y., Yu, T., & Gu, Z. (2011). Target selection with hybrid feature for BCI-based 2-D cursor control. *IEEE*

Transactions on Biomedical Engineering, 59, 132–140. https://doi.org/10.1109/TBME.2011.2167718

23. Yao, L., Meng, J., Zhang, D., Sheng, X., & Zhu, X. (2014). Combining motor imagery with selective sensation toward a hybrid-modality BCI. *IEEE Transactions on Biomedical Engineering, 61*, 2304–2312. https://doi.org/10.1109/TBME.2013.2287245

24. Kim, I. H., Kim, J. W., Haufe, S., & Lee, S. W. (2014). Detection of braking intention in diverse situations during simulated driving based on EEG feature combination. *Journal of Neural Engineering, 12*(1), 016001. https://doi.org/10.1088/1741-2560/12/1/016001

25. Zhuang, J., Geng, K., & Yin, G. (2019). Ensemble learning based brain-computer interface system for ground vehicle control. *IEEE Transactions on Systems, Man, and Cybernetics: Systems,*. https://doi.org/10.1109/TSMC.2019.2955478

26. Lu, Y., & Bi, L. (2019). EEG signals-based longitudinal control system for a brain-controlled vehicle. *IEEE Transactions on Neural Systems and Rehabilitation Engi-neering, 27*(2), 323–332. https://doi.org/10.1109/TNSRE.2018.2889483

27. Hekmatmanesh, A., Nardelli, P. H., & Handroos, H. (2021). Review of the state-of-the-art of brain-controlled vehicles. *IEEE Access, 9*, 110173–110193. https://doi.org/10.1109/ACCESS.2021.3100700

28. Jahangiri, A., & Sepulveda, F. (2019). The relative con-tribution of high-gamma linguistic processing stages of word production, and motor imagery of articulation in class separability of covert speech tasks in eeg data. *Journal of Medical Systems*. https://doi.org/10.1007/s10916-018-1137-9

29. André, A., Elia, F., Heidi, K., Peter, H., De, H. E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex, 15*, 221–228. https://doi.org/10.1093/cercor/bhh124

30. Mcguire, P. K., Silberswerg, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine, 26*, 29–38. https://doi.org/10.1017/s0033291700033699

31. Shuster, L. I., & Lemieux, S. K. (2004). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language, 93*, 20–31. https://doi.org/10.1016/j.bandl.2004.07.007

32. Bocquelet, F., Hueber, T., Girin, L., Stéphan, C., & Yvert, B. (2017). Key considerations in designing a speech brain-computer interface. *Journal of Physiology Paris, 110*, 392–401. https://doi.org/10.1016/j.jphysparis.2017.07.002

33. Shigeyuki, I., Tomohiro, S., Naoki, N., Rieko, O., Naohiro, T., Kazushi, I., & Amami, K. (2014). Neural decoding of single vowels during covert articulation using electro-corticography. *Frontiers in Human Neuroscience, 8*, 125. https://doi.org/10.3389/fnhum.2014.00125

34. Sereshkeh, A. (2017). Online EEG classification of covert speech for brain-computer interfacing. *International Journal of Neural Systems, 27*, 1750033. https://doi.org/10.1142/S0129065717500332

35. Tong, J., Wei, X., Dong, E., Sun, Z., Du, S., & Duan, F. (2022). Hybrid mental tasks based human computer inter-face via

integration of pronunciation and motor imagery. *Journal of Neural Engineering*. https://doi.org/10.1088/1741-2552/ac9a01

36. Delorme, A., Fernsler, T., Serby, H., Makeig S. (2006). EEGLAB Tutorial. *Journal of Hered, 101*.

37. Hung, C. I., Lee, P. L., Wu, Y. T., Chen, H. Y., & Hsieh, J. C. (2005). Recognition of motor imagery electro-encephalography using independent component analysis and machine classifiers. *Annals of Biomedical Engineering, 33*, 1053–1070. https://doi.org/10.1007/s10439-005-5772-1

38. Koles, Z. J., Lazar, M. S., & Zhou, S. Z. (1990). Spatial patterns underlying population differences in the back-ground EEG. *Brain Topography, 2*, 275–284. https://doi.org/10.1007/BF01129656

39. Koles, Z. J., Lind, J. C., & Soong, A. C. K. (1995). Spatiotemporal decomposition of the EEG: A general approach to theisolation and localization of sources. *Electroencephalography and Clinical Neurophysiology, 95*, 219–230. https://doi.org/10.1016/0013-4694(95)00083-B

40. Tang, F., Adam, L., & Si, B. (2018). Group feature selection with multiclass support vector machine. *Neuro-computing, 317*, 42–49. https://doi.org/10.1016/j.neucom.2018.07.012

41. Wang, T., Dong, E., Du, S., & Jia, C. (2019). A shallow convolutional neural network for classifying MI-EEG. *2019 Chinese Automation Congress (CAC)*. https://doi.org/10.1109/CAC48633.2019.8996981

42. Zhang, H., Sun, Y., Li, J., Wang, F., & Wang, Z. (2018). Covert verb reading contributes to signal classification of motor imagery in BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering A Publication of the IEEE Engineering in Medicine and Biology Society, 26*, 45–50. https://doi.org/10.1109/TNSRE.2017.2759241

43. Dasalla, C. S., Kambara, H., Sato, M., & Koike, Y. (2009). Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks, 22*, 1334–1339. https://doi.org/10.1016/j.neunet.2009.05.008

44. Neuper, C., Wörtz, M., & Pfurtscheller, G. (2006). ERD/ ERS patterns reflecting sensorimotor activation and deactivation. *Progress in Brain Research, 159*, 211–222. https://doi.org/10.1016/S0079-6123(06)59014-4

45. Wallentin, M., Michaelsen, J., Rynne, I., & Nielsen, R. H. (2014). Lateralized task shift effects in Broca's and Wernicke's regions and in visual word form area are selective for conceptual content and reflect trial history. *NeuroImage, 101*, 276–288. https://doi.org/10.1016/j.neuroimage.2014.07.012

46. Bian, Y., Qi, H., Zhao, L., Ming, D., Guo, T., & Fu, X. (2018). Improvements in event-related desynchronization and classification performance of motor imagery using instructive dynamic guidance and complex tasks. *Com-puters in biology and medicine, 96*, 266–273. https://doi.org/10.1016/j.compbiomed.2018.03.018

47. Bockbrader, M. A., Francisco, G., Lee, R., Olson, J., Solinsky, R., & Boninger, M. L. (2018). Brain computer interfaces in rehabilitation medicine. *PM and R, 10*, S233–S243. https://doi.org/10.1016/j.pmrj.2018.05.028