SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

July 2026, Vol. 69, Iss. 7, 172101:1–172101:16
https://doi.org/10.1007/s11432-025-4685-0

An asynchronous Chinese speller integrating eye tracking and large language models into a reactive BCI

Hao LI[1†], Yang YU[1†], Yifan ZHANG[1], Xin CHEN[2], Liang ZHOU[2], Jinfang LIU[2], Ming ZHANG[1], Ling-Li ZENG[1*] & Dewen HU[1]

[1]*College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China*
[2]*Department of Neurosurgery, National Clinical Medical Research Center for Geriatric Diseases, Xiangya Hospital, Central South University, Changsha 410008, China*

Received 17 April 2025/Revised 9 September 2025/Accepted 12 November 2025/Published online 11 June 2026

Abstract Reactive brain-computer interface (BCI) systems provide an alternative communication and control pathway for patients with severe motor impairments. However, these systems often face challenges in balancing speed, robustness, and user autonomy. These challenges are particularly encountered in logographic languages, such as Chinese, where the vast set of characters complicates text entry. In this study, we develop a practical Chinese-specific BCI speller (eLLM-BCI speller). The proposed eLLM-BCI speller combines eye tracking (ET) technology and large language models in a reactive BCI system to improve communication efficiency. The ET technology enables implicit, low-load asynchronous control; concurrently, the spatial distribution of the ET data is analyzed via flattened Gaussian kernel density estimation, providing additional spatial information that strengthens the joint decoder. Furthermore, we integrate the fine-tuned BLOOM model into the Chinese BCI system to further improve the interaction efficiency. The average output speed of the proposed eLLM-BCI speller was 5.81 sinograms/min, and the associative word usage rate increased from 13.96% (traditional word frequency-based baseline method) to 41.98% (proposed method). The LLM-based optimization significantly improved ($p < 0.001$) the communication speed; this is attributed to its strong context comprehension abilities. This study provides a comprehensive approach for developing efficient and reliable BCI spellers, addressing the challenging tradeoffs of user comfort, stability, and high-speed communication in real-world applications.

Keywords electroencephalogram, brain-computer interface, asynchronous Chinese speller, eye tracking, large language model

Citation Li H, Yu Y, Zhang Y F, et al. An asynchronous Chinese speller integrating eye tracking and large language models into a reactive BCI. Sci China Inf Sci, 2026, 69(7): 172101, https://doi.org/10.1007/s11432-025-4685-0

1 Introduction

Brain-computer interface (BCI) technology is highly significant for individuals with severe neuromuscular disorders, such as amyotrophic lateral sclerosis (ALS) [1,2], stroke [3], and spinal cord injury [4], because it provides an alternative communication channel between the human brain and the external world, helping patients to regain communication and control capabilities to some extent. Among the various BCI applications, spellers are particularly important for their ability to effectively convey user intentions, enabling accurate communication. Most electroencephalogram (EEG)-based spellers are reactive BCI systems that rely on signals in response to external stimuli. Widely-known examples include the P300 spellers [5] and the steady-state visual evoked potential (SSVEP)-based spellers [6]. However, they exhibit a critical tradeoff between communication speed and user comfort. Although the SSVEP-based spelling systems achieve high information transfer rates (ITRs) by employing large stimulus matrices (often exceeding 120 targets) and optimized decoding methods [7,8], the required flickering stimuli often cause significant visual fatigue and even seizures in susceptible individuals [9]. In contrast, the P300 spellers are generally more comfortable than the SSVEP-based spellers; however, they often require multiple stimulus repetitions to ensure accuracy [10,11], thereby limiting communication speed. Additionally, these systems have been mainly developed for alphabetic languages such as English. Consequently, the EEG-based Chinese spellers (EEGCSs) have received comparatively less attention. The main challenge is the vast number of Chinese sinograms (over 11000 Chinese characters) compared with the 26-letter English alphabet [12]. Addressing this language-specific challenge is crucial for the development of universally applicable BCI spellers.

* Corresponding author (email: zengphd@nudt.edu.cn)
† These authors contributed equally to this work.

A significant functional limitation associated with the two dominant paradigms is their typically synchronous nature. They operate in rigid, system-paced cycles, forcing the user to respond within a predefined time window. To overcome this limitation, asynchronous systems that allow users to communicate at their own pace have been actively investigated. Self-paced BCI control was achieved by employing two main approaches. The first decodes user intent directly from EEG signals, such as attention states or event-related potentials [13–15]; however, the inherent variability of EEGs often leads to instability. The second is based on hybrid BCI (hBCI) designs that use explicit physiological signals, such as electrooculography (EOG) as a control switch [16–18]. Despite their stability, these switches demand deliberate actions that increase the user's cognitive load [19], posing significant challenges to users with severe disabilities. Therefore, an implicit, low-load asynchronous control mechanism is crucial.

Beyond asynchronous control, the practical performance of BCI spellers critically depends on the accuracy and robustness of their decoding process. Although significant efforts have been made to optimize decoding by developing advanced filtering methods [10, 20] and deep learning-based classifiers [21–23], the performance of BCI spellers is ultimately limited by the inherent low signal-to-noise ratio (SNR) of EEG signals. The eye tracking technology has emerged as a promising method to overcome this limitation. In visual-based reactive BCI systems, such as the P300 speller, users direct their visual attention toward specific regions to evoke the event-related potential. Therefore, the gaze information can be synchronously collected and analyzed without introducing extra cognitive loads [24], thus enabling efficient decoding. In reactive BCI systems, ET data are combined with EEG signals by employing two approaches: serial and parallel. The serial approaches initially employ an eye tracker to identify the general area of the target, followed by an accurate selection by employing a visually evoked paradigm [25, 26]. In contrast, the parallel approaches simultaneously combine eye movement with EEG data to achieve joint target identification [27]. Although these hybrid systems improve the decoding accuracy in BCI systems [24, 26, 28, 29], ensuring stability against systematic bias in ET data, which can arise from minor, unperceived shifts in the user, remains a critical challenge [30]. This can lead to a sharp decrease in the accuracy of the eye tracking and, consequently, a significant reduction in the overall performance of the hybrid system. Therefore, a hybrid decoding approach specifically designed to improve robustness against this particular failure mode is required.

Even by achieving asynchronous interaction and robust decoding accuracy, the overall communication performance is still limited because the selection time per character approaches a practical limit. This challenge is particularly encountered in EEGCSs, as methods based on Pinyin [31–33] or strokes [34–36] require multiple selections to input a single character; consequently, the efficiency of semantic expression further decreases. Text generation techniques [37–39] can reduce the number of selections required for semantic expression, thus providing an effective approach to further improve the CS of reactive BCI spellers. Based on input data (e.g., sequences or keywords), text generation, which is commonly known as natural language generation, produces coherent and readable text in human language [40]. In previous studies, methods that integrate traditional language models (e.g., n-gram) into BCI systems have been investigated to optimize performance [41, 42]. However, the limitations of such language models in handling and understanding context impede the improvement in communication performance. Recently, pre-trained large language models (LLMs), such as GPT series [43] and BERT [44], have demonstrated a remarkable performance in natural language processing, particularly in understanding and predicting user intentions. By integrating LLMs into BCI systems, the powerful language comprehension and generation capabilities of LLMs can be exploited to understand the context of users' historical inputs and predict subsequent text [45, 46]; as a result, the users' input load can be reduced, and the spelling efficiency can be improved.

Overall, the practical application of a BCI speller is limited by three interconnected challenges: enabling asynchronous control, high-performance decoding, and efficient semantic expression. Although the above-mentioned solutions have been largely investigated, they were developed in isolation. A hybrid ET system cannot overcome the limitation of multistep input inefficiency, and an LLM-powered speller struggles with low decoding accuracy and usability without employing a robust, asynchronous BCI front end. Furthermore, the integration of LLMs into BCI spellers remains in a preliminary stage [45–47], lacking practical system implementation and empirical evaluation, especially for inputting Chinese characters. These problems highlight the significant potential of an integrated framework and corresponding methodological improvements to overcome them.

In this study, we propose a novel hybrid BCI speller for inputting Chinese characters. The proposed speller is based on the P300 system for better user comfort. We effectively combined ET with LLMs to overcome the inherent limitations of the P300 system in practical applications. The proposed system improves traditional methods in two key aspects. (i) A novel dual-purpose integration of ET is introduced to improve both system usability and decoding robustness. ET data not only enable implicit, self-paced asynchronous control, but more importantly, they serve as a robust secondary modality in the proposed joint decoding framework without imposing additional cognitive load. Gaussian kernel density estimation (KDE) is employed to flatten the data distribution, ensuring system reliability, even under non-ideal conditions. This approach makes the proposed joint decoder robust, preventing significant

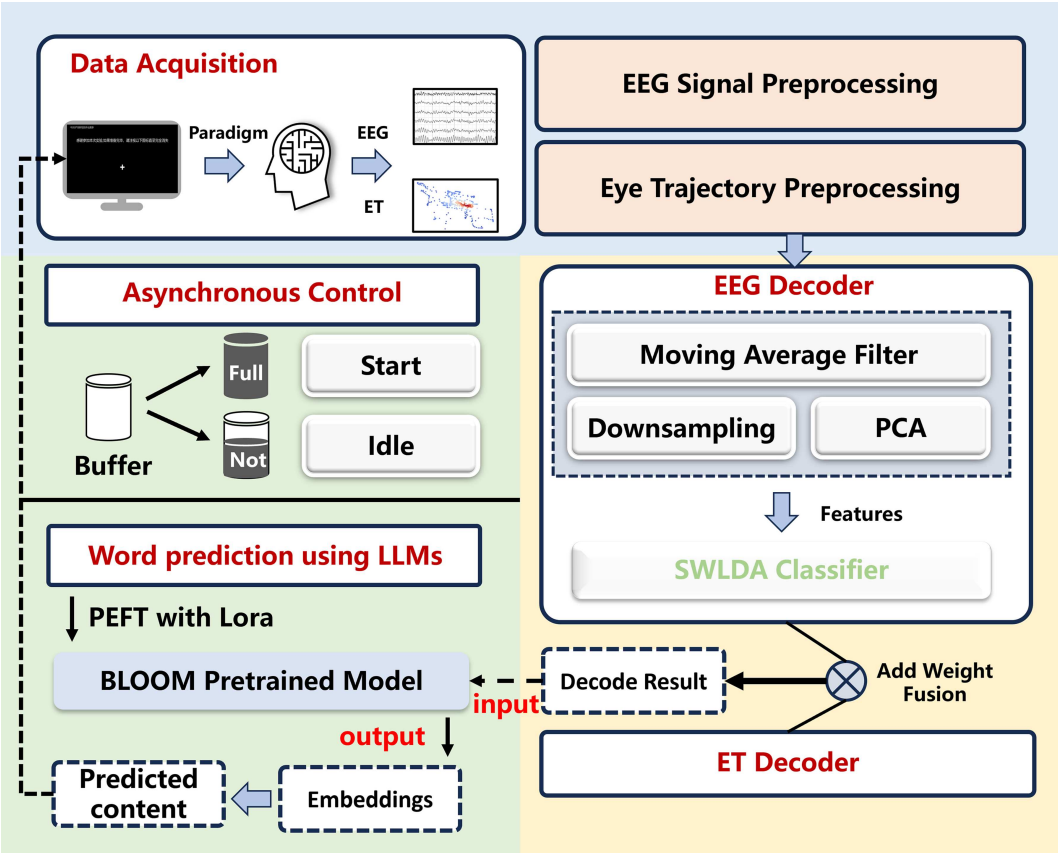

Figure 1 (Color online) Block diagram of the methodological framework developed in this study: data collection, asynchronous control, word prediction using LLMs, and a joint decoding strategy for combining EEG with ET data (ET: eye tracking, PEFT: parameter-efficient fine-tuning).

performance degradation caused by systemic errors, such as those arising from head posture. (ii) The BLOOM LLM is integrated into the proposed hybrid system to further improve the efficiency of inputting Chinese characters. By fine-tuning this LLM for the Chinese language and incorporating it into our speller, the input efficiency of the system was significantly improved ($p < 0.001$) by predicting associative words. Our study delivers the first implementation of such a holistic BCI framework, addressing the challenging tradeoffs of user comfort, stability, and high-speed communication in real-world applications.

2 Materials and methods

This section presents the methodology, detailing the implementation of the asynchronous Chinese spelling system optimized with eye tracking and LLM. Figure 1 illustrates the overview of the methodological framework developed in this study.

2.1 Data acquisition

The study involved ten healthy participants (eight males and two females), aged between 22 and 25, noted as S1–S10. Each participant possessed normal or corrected-to-normal vision and reported no history of neurological or psychiatric conditions. Prior to the experiment, all participants were informed of the experiment's content and agreed to participate. The tasks to be completed during the experiment were also clearly explained to participants. Upon the completion of the experiment, participants were remunerated for their participation. This study was approved by the Ethics Committee of the Xiangya Hospital of Central South University. The protocol number is 2021111249.

EEG data were acquired using the actiCHamp Plus EEG analyzer at a sampling rate of 500 Hz. During data collection, a notch filter at 50 Hz was applied to eliminate power frequency interference. Raw EEG signals were band-pass filtered between 0.05 and 40 Hz to remove low-frequency drifts and high-frequency noise. Following the international 10–20 system, eight wet active electrode channels (Fz, Cz, Pz, Oz, PO3, PO4, P3, and P4) were

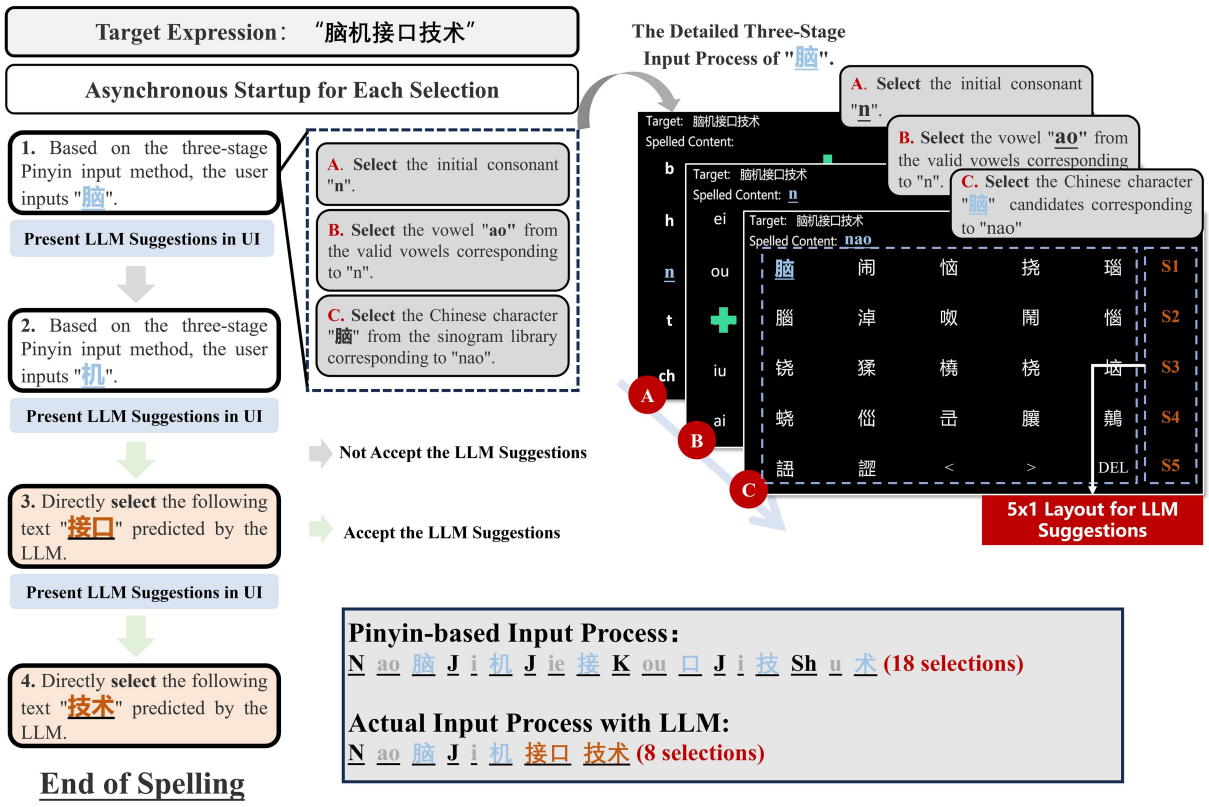

Figure 2 (Color online) Interactive workflow of our asynchronous system with LLM suggestions. Green cross icon: flash stimulus.

employed. The ground electrode was placed at the AFz position, while the reference electrode was positioned at the left mastoid. Finally, the experiment began after ensuring that the impedance was reduced below 15 kΩ.

In parallel, eye movement data were recorded using Tobii Nano eye tracker at 60 Hz. This device operates on the principle of video-based eye-tracking, which uses an infrared camera to capture images of the eye, and computer vision algorithms calculate the precise gaze position based on the pupil center and corneal reflection. This device can accurately capture the user's real time gaze position on the screen, providing an additional modality that complements the EEG signals.

2.2 Chinese BCI speller with asynchronous startup

The asynchronous Chinese BCI speller was implemented by combining eye tracking data with P300 signals within a reactive BCI framework. This approach aimed to enhance the efficiency and accuracy of Chinese character input. Figure 2 shows the workflow of our LLM-based BCI speller.

The speller's user interface (UI) utilized a 5 × 6 matrix layout, where each cell contained a Chinese character or phonetic symbol. To maintain consistency and minimize user adaptation costs throughout the spelling process, the first five columns (5 × 5) were dedicated to three-stage Pinyin-based input, and the sixth column (5 × 1) was reserved for presenting LLM suggestions. The input process within the 5 × 5 area was as follows. First, the initial consonant interface was displayed, containing all Pinyin initial consonants and the vowels a, o, e, along with a "DEL" key for deleting previous input. After an initial was selected, the system dynamically displayed only the valid vowels that could be paired with the selected initial consonant. Upon completion of a full Pinyin, the interface presented the corresponding Chinese characters. If the number of valid vowels or Chinese characters exceeded the available space, navigation keys ("<" and ">") allowed navigation between pages. Once a character was selected, the LLM provided predictive suggestions based on the context, which were displayed in the reserved sixth column. The system followed the row-column P300 stimulus paradigm [48]. In this paradigm, rows and columns of the matrix were flashed randomly. When the stimulus at the target position was flashed, it elicited a P300 response in the user's EEG signal.

Asynchronous control was achieved by applying a gaze position buffer. A thread was initiated to detect the user's gaze state. When the distance between the current eye tracking sample point and the previous one was less than

Table 1 Hyperparameter settings in supervised fine-tuning with LoRA.

Hyperparameter	Value
Learning rate	5×10^{-4}
Batchsize	128
Optimizer	AdamW
Epochs	10
LoRA: α	32
LoRA: rank	8
Dropout rate	0.1

the threshold, the current gaze position was added to the buffer; otherwise, the buffer was cleared. Consequently, the spelling process was initiated only when the buffer was full; otherwise, the system remained suspended. After a target was selected, the buffer was automatically cleared to prepare for the next selection, which allowed for consecutive input. This approach allowed users to interact with the speller at their own pace.

2.3 LLM-based word prediction

In this study, we utilized the BLOOM 1.7 B pre-trained model [49] for word prediction, which comprises 1.7 billion parameters. This model employs an autoregressive architecture, similar to GPT-3, and has been pre-trained on a large corpus to learn general language representations. The training dataset includes 46 different natural languages and 13 programming languages, with approximately 18.3% of the data being in Chinese. This makes the model particularly suitable for optimizing the Chinese BCI spelling system. Therefore, fine-tuning was performed on the pre-trained BLOOM model to enhance its performance in application scenarios (e.g., Chinese daily conversations). The fine-tuning dataset was derived from the LCCC Chinese dialogue corpus [50], comprising approximately 6.8 million rounds of Chinese dialogues.

During the fine-tuning process, we utilized the low-rank adaptation (LoRA) [51] for parameter-efficient fine-tuning, which freezes the pre-trained model weights and injects trainable low-rank matrices into each layer of the Transformer architecture, significantly reducing the number of trainable parameters. This PEFT method was chosen not only because full parameter fine-tuning was computationally prohibitive for a model of this scale, but also because it effectively adapts the model to our dialogue corpus with a minimal number of trainable parameters [52]. The fine-tuning process was conducted on 8×NVIDIA RTX4090 24 GB GPUs with mixed-precision training, requiring approximately 39 h to complete. The fine-tuning hyperparameters are detailed in Table 1.

Given the spelled Chinese text C, tokenization was performed using the BLOOM tokenizer's byte-level byte-pair encoding (BPE) algorithm [53], which employed a 250680-token vocabulary specifically optimized for multilingual processing. The vectorized token sequence $s = s_1 s_2 \cdots s_n$ was then mapped through the pre-trained embedding layer, preserving positional information through absolute positional embeddings.

$$v_i = \text{Embedding}(s_i) = v_i^t + v_i^p, \tag{1}$$

where v_i^t is the word vector of s_i and v_i^p is the position vector.

Finally, the output from the last transformer layer was passed to the linear and softmax layers to generate the probability distribution of each token.

$$P(s_i|s_1 \cdots s_{i-1}) = \text{Softmax}(W^{\text{out}} h_i^{(L)} + b^{\text{out}}), \tag{2}$$

where W^{out} is the weight matrix of the output layer, $h_i^{(L)}$ is the final output of the transformer layer, and b^{out} is the bias of the output layer.

2.4 Joint decoder integrating EEG and eye tracking data

2.4.1 *EEG decoder*

The raw EEG signals were first processed using a causal moving average filter with a window size of 5 samples to preserve temporal causality. The filtered signals were then downsampled by a factor of 10 to 50 Hz using a polyphase antialiasing filter. Feature extraction was performed through principal component analysis (PCA), retaining 150 components.

For P300 classification, three established methods available in scikit-learn were implemented. (i) Linear discriminant analysis (LDA) with singular value decomposition solver. (ii) Support vector machine (SVM) with linear

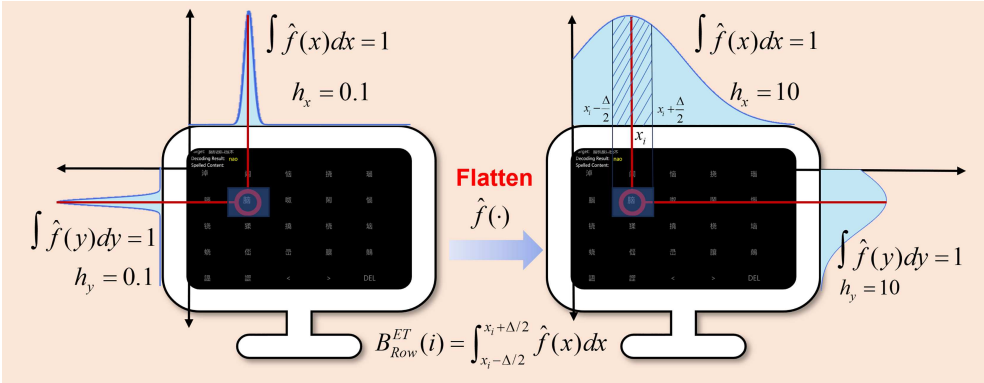

Figure 3 (Color online) ET decoding confidence calculation based on flatten Gaussian kernel probability density function.

kernel. (iii) Stepwise linear discriminant analysis (SWLDA) using a forward/backward selection threshold of $p = 0.1/0.15$.

The discriminant function $D(\cdot)$ of these classifiers can be uniformly defined as a linear combination of the n selected features (SVM/LDA: $n = 150$; SWLDA: $n <150$)

$$D(x) = w_0 + \sum_{i=1}^{n} w_i x_i, \tag{3}$$

where w_0 is the bias term, w_i are the weights, and x_i are the selected features. Based on the parameters derived from the trained discriminant function, the confidence scores for each row i and column j in each spelling trial can be calculated as $B^{\mathrm{EEG}}_{\mathrm{Row}}(i)$ and $B^{\mathrm{EEG}}_{\mathrm{Col}}(j)$ respectively.

$$B^{\mathrm{EEG}}_{\mathrm{Row}}(i) = \sum_{l} D\left(x^{(l)}\right) \text{ if } l = i, \tag{4}$$

where $x^{(l)}$ denotes the EEG epoch corresponding to the flashing of the l-th row.

2.4.2 *Eye tracking decoder*

For participants with normal visual function, when attention is focused on a fixed target, the gaze points data show low variance characteristics ($\sigma_x^2 \leqslant 0.05$, $\sigma_y^2 \leqslant 0.05$). The distribution ratio of the gaze points in each row and column area can be calculated as

$$P_{\mathrm{row}_i} = \frac{N_{\mathrm{row}_i}}{N_{\mathrm{total}}}, \quad P_{\mathrm{col}_j} = \frac{N_{\mathrm{col}_j}}{N_{\mathrm{total}}}, \tag{5}$$

where N_{row_i} and N_{col_j} represent the number of valid gaze points in the i-th row and j-th column areas, respectively, and N_{total} represents the total number of sampling points in a single selection period.

A baseline approach is to use this ratio directly as the confidence score. However, this method is highly susceptible to systemic bias. For example, if the user's posture changes, it could cause a relative positional change with the eye tracker without the user's awareness, leading to a shift in the collected gaze point coordinates. When this shift exceeds the character cell's region, the baseline method could result in error outputs, causing catastrophic reductions in accuracy. Even in hybrid systems, when the target remains unchanged, the spike distribution of the eye tracking data causes the joint decoder to overly trust the biased eye tracking data, which degrades the performance of the joint decoder.

To address this, we proposed a confidence calculation method based on Gaussian KDE. This approach was specifically chosen to enhance stability by making the confidence score less sensitive to the precise coordinates of gaze points and more robust against potential systemic bias. Figure 3 shows the confidence calculation approach of the eye tracking decoder. Probability density functions (PDFs) were constructed in both horizontal and vertical directions. For the x-direction

$$\hat{f}(x) = \frac{1}{nh_x} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_x}\right), \tag{6}$$

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \tag{7}$$

where $K(\cdot)$ represents the Gaussian kernel function, h_x is the bandwidth parameter in the x-direction, and n represents the number of valid gaze points. Similarly, for the y-direction, we have $\hat{f}(y)$. By adjusting the bandwidth parameter h, the data distribution was flattened, and the final confidence score was obtained through the integral of the probability density function

$$B_{\text{Row}}^{\text{ET}}(i) = \int_{x_i-\Delta/2}^{x_i+\Delta/2} \hat{f}(x)\,dx, \tag{8}$$

$$B_{\text{Col}}^{\text{ET}}(j) = \int_{y_j-\Delta/2}^{y_j+\Delta/2} \hat{f}(y)\,dy, \tag{9}$$

where Δ represents the width of the character cell, $B_{\text{Row}}^{\text{ET}}(i)$ and $B_{\text{Col}}^{\text{ET}}(j)$ represent the confidence scores for the i-th row and j-th column, respectively. $\hat{f}(x)$ and $\hat{f}(y)$ denote the probability density functions in the x and y directions.

2.4.3 *Joint decoder*

To enhance decoding accuracy, a joint strategy integrating EEG and eye tracking data was implemented. The confidence scores from the EEG-based and eye tracking-based were weighted and averaged to produce a fused confidence score for each row and column

$$B_{\text{Row}}^{\text{Fusion}}(i) = \alpha \cdot B_{\text{Row}}^{\text{EEG}}(i) + \beta \cdot B_{\text{Row}}^{\text{ET}}(i), \tag{10}$$

$$B_{\text{col}}^{\text{Fusion}}(j) = \alpha \cdot B_{\text{col}}^{\text{EEG}}(j) + \beta \cdot B_{\text{col}}^{\text{ET}}(j), \tag{11}$$

where α and β are weighting factors determined by the reliability of each modality. We referenced an improved method [54] to determine the weights of two decoders

$$\begin{cases} \dfrac{\alpha}{\text{acc}_1} = \dfrac{\beta}{\text{acc}_2}, \\ \alpha + \beta = 1, \end{cases} \tag{12}$$

where acc_1 and acc_2 denote the decoding accuracy using single EEG data and single eye tracking data collected from the offline session, respectively. The row and column with the highest fused confidence scores were selected as the target.

2.5 Experimental procedure

In this study, participants were instructed to engage in a series of experiments designed to evaluate the proposed asynchronous Chinese speller. The experimental procedure was divided into an offline training session, an online testing session, a copy spelling session and a free spelling session. Figure 4 illustrates the experimental procedure.

The offline training session comprised three runs, each consisting of 15 blocks with 8 trials per block. During the experiment, participants were instructed to focus their attention on the highlighted target and covertly count the number of times the target stimuli appeared. After the completion of the training session, the data acquired from the three runs will undergo offline analysis. Three different classifiers were used to calculate the information transfer rate corresponding to each trial. Ultimately, the trial number and classifier combination yielding the highest ITR was selected as the hyperparameter for online decoding. Short breaks of 5–10 min were provided between runs to avoid fatigue. After that, the online testing session was implemented to evaluate the online performance of the speller in real-time with the best hyperparameters obtained through offline analysis.

Following this, the copy and free spelling session were conducted to investigate the improvements brought by integrating BLOOM into the system, compared with that based on word frequency. To ensure a fair comparison, the dataset for word frequency-based method was derived from the same LCCC dataset used for fine-tuning the LLM. For this method, word-level segmentations were first performed on the corpus with the jieba library, after which we filtered the results to retain only meaningful words composed of two or more Chinese characters. Finally, the frequencies of these words were calculated to construct a predictive model based on Bigram probabilities. This model provided suggestions by identifying the most probable words to follow the user's previously entered content. Participants were not prompted with a target to focus on, and they needed to accomplish the tasks with their own cognitive resources. In the copy spelling session, participants were randomly assigned with a short sentence containing 12–20 Chinese characters (from our fine-tuning test dataset) in each run and were required to spell the sentence in its entirety. During the free spelling session, longer sentences containing 30–60 words, which provided

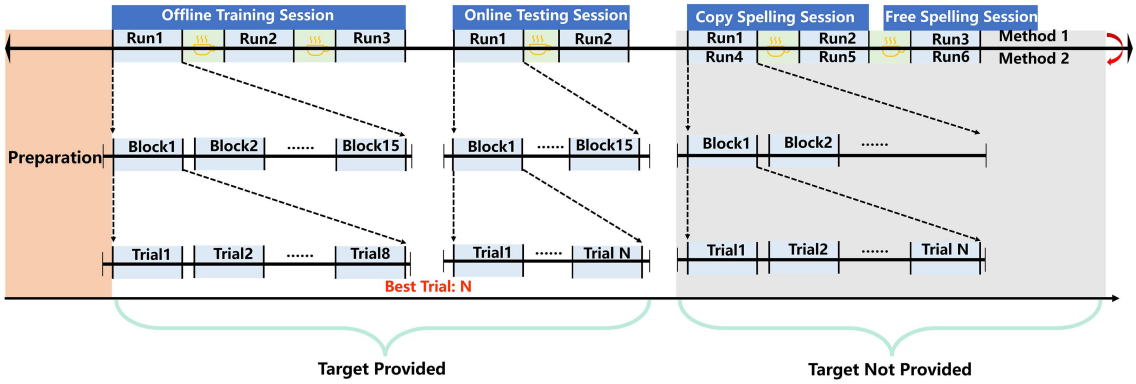

Figure 4 (Color online) Experimental procedure in this study. Method 1: optimization based on word frequency; Method 2: optimization based on LLM; Trial: consisting of 11 flashes (5 rows and 6 columns) to stimulate ERP responses; Block: consisting of n trials, and each block outputs a selection.

richer contextual information, were selected for the participants to spell. This allowed a more flexible way to convey the target semantics, since the result may not be as same as the target sentence. The participants could press the space key to end the spelling when they deemed the spelled content adequately represented the intended semantics.

2.6 Performance evaluation metrics

2.6.1 *Information transfer rate*

In our study, the information transfer rate [55] was introduced to assess the communication rate of our BCI system in the offline training and online testing sessions. It simultaneously considers the number of selectable targets N, the target recognition accuracy P, and the time T for single target selection. Raw bitrate for a single target selection is first calculated

$$B = \log_2 N + P\log_2 P + (1-P)\log_2\left(\frac{(1-P)}{(N-1)}\right), \tag{13}$$

where B is measured in bits per selection. Furthermore, the value of ITR represents the amount of information output by the system per minute.

$$\mathrm{ITR} = B \cdot \left(\frac{60}{T}\right), \quad T = t_s + t_b, \tag{14}$$

where t_s denotes the stimulation time per selection and t_b denotes the inter-selection pause time.

2.6.2 *Semantic textual similarity (STS)*

In the free spelling session, participants were not required to ensure that their spelled content exactly matched the target sentence. This necessitated the introduction of a method to evaluate the semantic similarity between the participants' spelled sentences and the target sentences. Therefore, this study employed the Sentence-BERT (SBERT) [56] model and cosine similarity to assess STS. This method aimed to quantify the accuracy of the BCI system in conveying semantics when optimized by different associative input methods, and further quantified the practical effectiveness of the BCI system.

Given two sentences S_1 and S_2, we firstly encoded them into high-dimensional vector representations using the SBERT model in semantic space. Let f denote the SBERT encoding function; then the vector representations of the sentences are

$$v_1 = f(S_1), \quad v_2 = f(S_2), \tag{15}$$

where v_1 and v_2 are the vector representations of sentences S_1 and S_2, respectively. Then, cosine similarity was used to measure the similarity between the two vectors.

$$\mathrm{cosine\_similarity}(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\|\|v_2\|}, \tag{16}$$

where $v_1 \cdot v_2$ denotes the dot product of the vectors, and $\|\cdot\|$ denotes the norms of the vector.

2.6.3 *Communication efficiency metrics*

During the free spelling session, this study also calculated each participant's communication speed (CS), total number of spellings (TS), and the associative words usage rate (AUR). These metrics were used to evaluate the performance of the system in practical application scenarios.

$$\mathrm{CS} = \frac{N_c \cdot 60}{T_{\mathrm{all}}}, \tag{17}$$

$$\mathrm{AUR} = \frac{N_{\mathrm{asso.}}}{\mathrm{TS}}, \tag{18}$$

where N_c represents the total number of Chinese characters spelled, $N_{\mathrm{asso.}}$ and TS represent the suggested word usage times and the total number of selections, respectively. T_{all} denotes the total time spent completing the task for this run.

3 Results

3.1 Event-related potential (ERP) analysis

As illustrated in Figure 5(a), the grand average ERP responses to target and non-target stimuli across all subjects were presented for each channel. Under target stimuli, the EEG signal exhibited a slight decline around 200 ms (N200), followed by a significant positive amplitude peak between 300 and 400 ms. In contrast, the non-target stimuli elicited relatively stable EEG activity with no significant amplitude fluctuations. Notably, the ERP amplitudes recorded at the Fz, PO3, and PO4 channels were lower compared to those in other channels.

Figure 5(b) shows the t-value heatmap obtained from t-tests comparing grand-averaged target and non-target EEG epochs, where blue represents negative and red represents positive differences. Significant differences in ERP responses were primarily observed in the 300–400 ms range, with a significance level of $p = 0.01$. This spatiotemporal dynamic was further illustrated through six-phase topographic maps shown in Figure 5(c). In the early stage (50–150 ms), ERP responses exhibited low amplitudes, indicating a resting-state pattern. Between 150 and 250 ms, a slight decrease in ERP activity appeared in the occipital region. During 250–350 ms, a positive, higher amplitude ERP response emerged in the occipital lobe and showed a propagation trend toward the parietal region. Finally, between 350 and 450 ms, a robust P300 response was elicited in the parietal lobe, which subsequently returned to baseline.

3.2 Target decoding results

Figure 6 shows the variations in accuracy and ITR of the BCI system under different numbers of trials per block during the offline training session, using three classifiers (SWLDA, LDA, and SVM) for EEG decoding. For all participants, the decoding accuracy showed an overall upward trend as the number of trials per block increased. Notably, participant S1, who had prior experience with similar experiments, achieved approximately 90% decoding accuracy under a single trial. The ITRs of most participants exhibited a trend of initially increasing and then decreasing, with the peak typically occurring at two or three trials. When eye tracking data were incorporated, the fusion decoding accuracy remained consistently high (>95%) even with only one trial. This result indicates that the joint decoding strategy further enhanced both the efficiency and stability of the hybrid speller. In subsequent sessions, the optimal trial number and classifier type for each participant were utilized as hyperparameters in the hybrid speller to maximize the ITR. During the online testing session, participants were instructed to spell 30 random targets using the hybrid system. On average, the joint decoding strategy achieved a 37.19% higher decoding accuracy compared with EEG-only decoding, and an ITR that was 49.43 bits/min greater.

3.3 Sentences spelling without target prompt

We subsequently conducted two sentences spelling sessions to evaluate the performance of our hybrid BCI speller in a realistic application scenario. Table 2 summarizes the TS, CS, and AUR for participants (S1–S10) in the copy spelling session. Table 3 presents the CS, AUR and STS results from the free spelling session.

In the copy spelling session, the results in Table 2 indicate that the LLM-based optimization of the Chinese BCI speller significantly reduced the TS required to spell short sentences by an average of 18.8 selections (paired t-test, $p < 0.005$) and increased the CS by an average of 1.054 sinograms/min (paired t-test, $p < 0.005$). Moreover, the

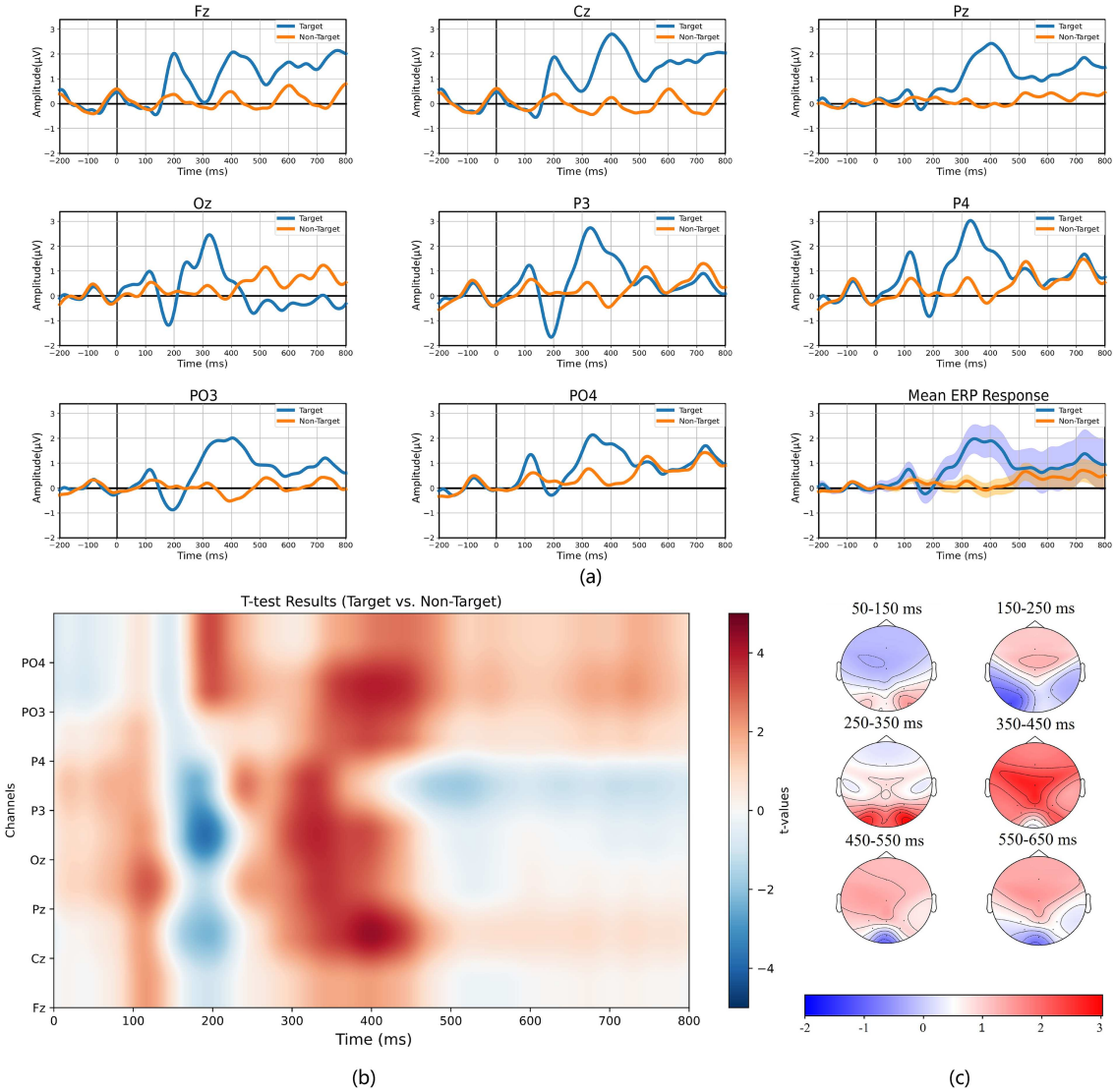

Figure 5 (Color online) ERP analysis in this study. (a) The grand average ERPs over 8 electrodes for each channel across all subjects S1–S10; (b) the t-test values of ERPs during 0–800 ms for 8 channels; (c) the topographical maps of EEG epochs under target stimuli across six different time periods.

Table 2 The TS, CS, and AUR for different subjects in the copy spelling session. Bold values indicate the highest average performance. Paired t-test, NS: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.005$; ****: $p < 0.001$.

Subject	Words-frequency-based			LLM-based		
	TS	CS	AUR (%)	TS$^{(***)}$	CS$^{(***)}$	AUR$^{(**)}$ (%)
Subject1	107	3.01	12.34	92	3.77	16.22
Subject2	110	2.91	10.97	106	3.09	13.17
Subject3	112	2.85	13.63	99	3.31	18.38
Subject4	43	2.03	7.32	27	3.26	22.22
Subject5	115	2.77	10.43	89	3.87	15.79
Subject6	119	2.69	10.00	93	3.71	18.04
Subject7	82	2.77	12.33	38	3.66	31.58
Subject8	80	2.83	6.59	59	3.64	20.94
Subject9	76	2.86	14.31	77	3.64	16.75
Subject10	79	2.85	10.00	59	3.66	18.84
Average	92.3	2.757	11.19	**73.5**	**3.811**	**18.49**
Std.	±22.97	±0.26	±2.54	±25.27	±0.80	±5.29

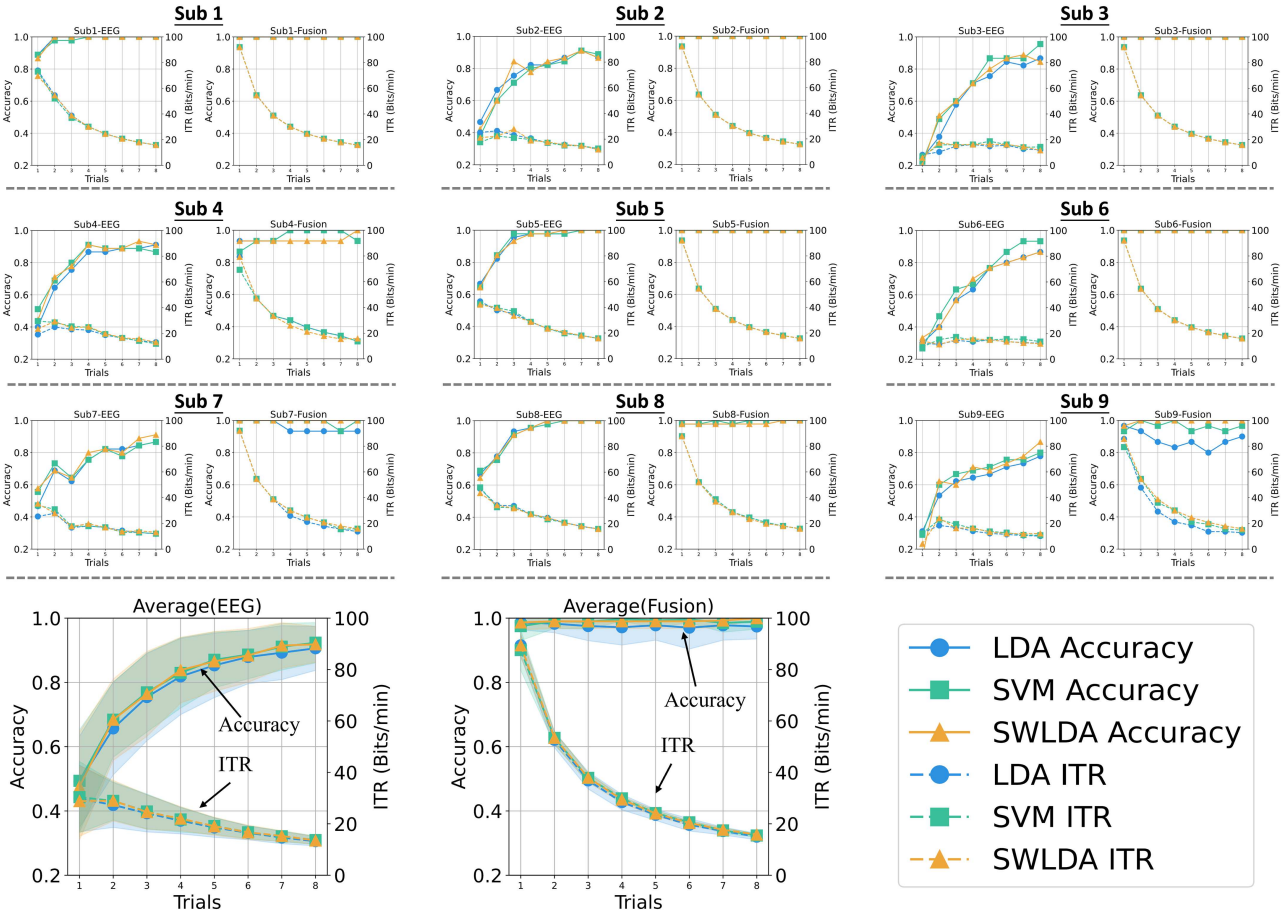

Figure 6 (Color online) Accuracy and ITR changes with trials per block (1–8) across 9 subjects (S10 ET data missing in training session), utilizing three different classifiers (SWLDA, LDA, and SVM). Solid line indicates accuracy, dashed line indicates ITR.

Table 3 The CS, AUR, and STS for different subjects in the free spelling session. Bold values indicate the highest average performance. Paired t-test, NS: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.005$; ****: $p < 0.001$.

Subject	Words-frequency-based			LLM-based		
	CS	AUR (%)	STS (%)	$CS^{(****)}$	$AUR^{(****)}$ (%)	$STS^{(NS)}$ (%)
Subject1	2.134	10.00	89.95	5.965	51.40	90.14
Subject2	3.302	12.50	96.43	4.920	30.61	95.01
Subject3	3.556	17.39	93.73	4.448	25.00	92.25
Subject4	2.898	12.96	92.71	6.733	48.65	95.74
Subject5	3.017	18.18	98.61	6.875	50.00	96.39
Subject6	2.886	7.25	97.24	5.105	43.48	99.00
Subject7	3.112	15.63	95.48	5.293	35.90	91.50
Subject8	3.309	15.84	98.32	5.911	42.25	95.73
Subject9	2.844	10.43	96.11	7.005	53.85	97.84
Subject10	3.337	19.39	96.05	5.880	38.67	96.58
Average	3.039	13.96	**95.46**	**5.813**	**41.98**	95.02
Std.	±0.375	±3.76	±2.53	±0.829	±8.98	±2.70

LLM-based word prediction improved the AUR by 7.3% compared with the frequency-based method (paired t-test, $p < 0.01$), suggesting that the integration of LLMs significantly accelerated Chinese sentence spelling.

As shown in Table 3, during the free spelling session, the LLM-based optimization significantly improved the CS by 2.774 sinograms/min (paired t-test, $p < 0.001$) and increased the AUR from 13.96% to 41.98% (paired t-test, $p < 0.001$). These findings demonstrated that the LLM-based optimization method achieved better words prediction during long sentence spelling, thereby enabling faster and more coherent semantic expression. In addition, the semantic cosine similarity between the spelled content and the target sentences was compared across both methods.

Both approaches achieved high STS exceeding 95%, with the LLM-based method attaining an average semantic similarity of 95.02% to the target sentences. No significant difference was observed between the two methods (paired t-test, $p > 0.05$), indicating that both produced semantically consistent sentence outputs.

4 Discussion

In Subsection 3.1, we analyzed the differences in the spatiotemporal dynamic changes of the EEG signals under target and non-target stimuli to investigate (to a certain extent) the interpretability of EEG decoding. We then analyzed the offline data to compare the performance achieved by the proposed joint decoding strategy with that obtained using EEG alone. In the sentence spelling experiments, we assessed the performance of the proposed LLM-based BCI speller in short and long sentence tasks, respectively. In this section, we further discuss the findings to demonstrate the advantages of the proposed speller.

4.1 Interaction between EEG and eye tracking

Firstly, asynchronous control was achieved by employing a gaze position buffer. Typically, low variance in the ET data activates the system, whereas a high variance deactivates it. This mechanism ensures that the system responds quickly when the user intends to interact; in contrast, the system remains inactive when there is no clear intent, thus preventing false activations.

A key challenge in ET-based interaction interfaces is the Midas Touch problem [57], where the system struggles to distinguish between simply looking at an interface element and an intentional action. This often leads to the accidental triggering of commands not intended by the user. In the proposed asynchronous hybrid BCI speller, which is based on the P300 paradigm, the dwell time required for target selection is determined automatically. The ET and EEG data are collected synchronously without setting a separate time window for the ET decoder, thereby mitigating the Midas Touch problem to some extent.

The integration of the ET data into the BCI system introduces a new data modality without increasing the cognitive load. As a result, the combined advantages of BCI and the ET data improve the decoding efficiency. In the offline training and online testing sessions, incorporating ET data significantly improved ($p < 0.001$) the decoding accuracy of the spelling system. When only EEG data were used for decoding, the ITR for the participants during the offline experiment initially increased and then decreased with the number of trials. In contrast, by employing ET data, the ITR for the participants was maximized at trial number 1, corresponding to an improvement of 49.43 bits/min compared with the EEG-only spelling system during the online testing session. This indicates that the incorporation of ET data improves the spelling speed; thus, the issue of low decoding accuracy in the traditional P300 system is partially addressed during single-trial decoding.

Despite the overall improvement, no significant differences were found between the results of the joint and the ET-only decoder ($p > 0.05$) under ideal calibration conditions; both decoders exhibited high accuracy and ITR. This finding can be attributed to the low variance of the ET coordinates collected from healthy participants after calibration. The spatial distribution of the ET data, modeled using the flattened KDE, produces stable and accurate decoding outcomes. Therefore, the joint decoder has limited potential for further accuracy enhancement under ideal conditions.

Next, we conducted a bias injection experiment to further investigate whether the joint decoding method could simultaneously exploit the spatiotemporal complementarity of EEG and eye movement data, rather than just the eye movement data playing a role. During the selection process, a fixed offset $(\text{bias}^x, \text{bias}^y)$ was applied to all ET sampling points. The bias term indicates the potential systematic bias risk of video-based eye movement data [30]. The offset in each selection was kept constant across three parameter gradients (0.25Δ, 0.5Δ, and 0.75Δ), representing a slight impact, a critical impact, and a complete ET decoder failure, respectively. The fusion weights were calculated using unbiased ET data.

Figure 7 shows that when $\text{bias}^{x/y} = 0.25\Delta$, the performance of the decoder on the ET-only data did not decrease significantly; additionally, the performance of the hybrid decoder was relatively high. However, when the bias increased to 0.5Δ, the ET-data decoding accuracy dropped significantly to 51.43%. In the joint decoding strategy, the accuracies improved by 8.62% and 18.01% (paired t-test, both $p < 0.001$) relative to single-modality data. Even in the extreme case of almost complete failure ($\text{bias}^{x/y} = 0.75\Delta$), the joint decoder improved the decoding accuracy to a certain extent (paired t-test, $p < 0.05$) compared with the EEG-only decoder. The above results demonstrate that the proposed joint decoding strategy effectively leverages the combined characteristics of the EEG and ET data to improve the decoding performance. Additionally, it exhibits a certain tolerance compared with the sharp performance decrease regarding the single-modal data, indicating the stability of the proposed system. Moreover,

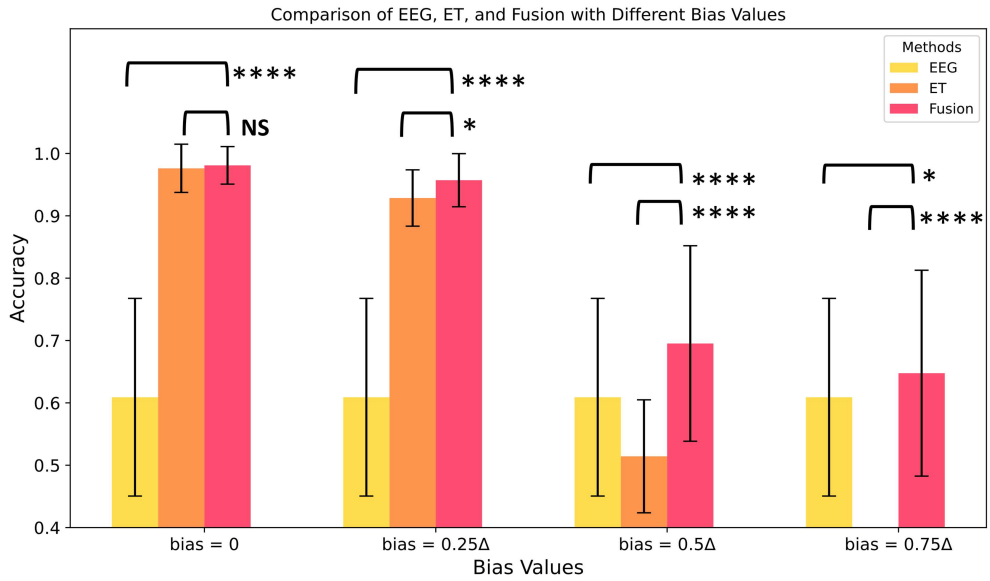

Figure 7 (Color online) Bias injection experimental results (paired t-test, NS: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.005$, ****: $p < 0.001$).

the fusion mechanism provides interpretability. For instance, when the EEG decoder fails to discriminate between the first and fifth rows, the spatial distribution of the gaze points supplies additional information to improve the decoding accuracy.

4.2 A deeper insight into the weighting mechanism of the KDE-based fusion

In our method, a data-driven adaptive weight mechanism was adopted to determine the weights based on the actual performance of different modal data in the calibration phase. The rationale of this approach appears intuitive: the modality with the higher reliability should justifiably be assigned greater confidence. However, we empirically investigated whether our adaptive weights functioned optimally or at least contributed significantly to the performance of our hybrid system. Contrary to expectations, our findings indicated that this was not the case.

Based on the bias injection experiment described in Subsection 4.1 ($\text{bias}^{x/y} = 0.5\Delta$), where the hybrid system demonstrated a significant (paired t-test, $p < 0.001$) improvement over single-modal systems, we further investigated its performance under different weighting strategies. In addition to the system's original adaptive weighting strategy, we introduced five distinct fixed-weighting schemes; these are: pure-ET ($\alpha = 0$), super heavy-ET ($\alpha = 0.1$), heavy-ET ($\alpha = 0.25$), balanced ($\alpha = 0.5$), and heavy-EEG ($\alpha = 0.75$). These schemes were compared against the adaptive scheme ($\alpha = 0.370 \pm 0.073$). We then conducted experiments to evaluate the system's performance under each weighting scheme. The results were presented in Figure 8.

We observed that the overall system performance initially improves rapidly; then, it deteriorated as the weight of the EEG data increases. Notably, all hybrid systems incorporating EEG data consistently outperformed the ET-only system. This observation reconfirmed the conclusion made in Subsection 4.1. As the EEG-related weight increased, the system performance was maximized at $73.33\% \pm 15.73\%$ under the heavy-ET weighting scheme. Beyond this point, the performance gradually decreased, approaching that of the pure-EEG system, as shown in Figure 7. When $\alpha = 0.75$, the performance of the hybrid system did not significantly differ from that of the pure-EEG system, indicating that the weighting mechanism assigned a value close to 1 to the EEG modality. In contrast, the performances of both the heavy-ET and super heavy-ET configurations surpassed that of the adaptive scheme; additionally, the mean error and variance decreased. These findings indicate that our data-driven weighting mechanism is suboptimal and fails to achieve the intended optimization.

The key finding of our weighting analysis is that the system optimality is achieved at a remarkably low EEG weight ($\alpha = 0.1/0.25$), indicating that our intuitive, data-driven adaptive weighting scheme is, in fact, suboptimal. This counterintuitive observation can be attributed to the flattened KDE-based method, which transforms the ET data into a low-amplitude, high-bandwidth distribution, thereby increasing their sensitivity to the integrated EEG data. Consequently, an EEG weight that appears small can be sufficient to push the system beyond its optimal point. This indicates that our adaptive method is not fundamentally flawed; instead, it requires targeted modification. For example, the EEG component can be artificially downweighed during the weight estimation to

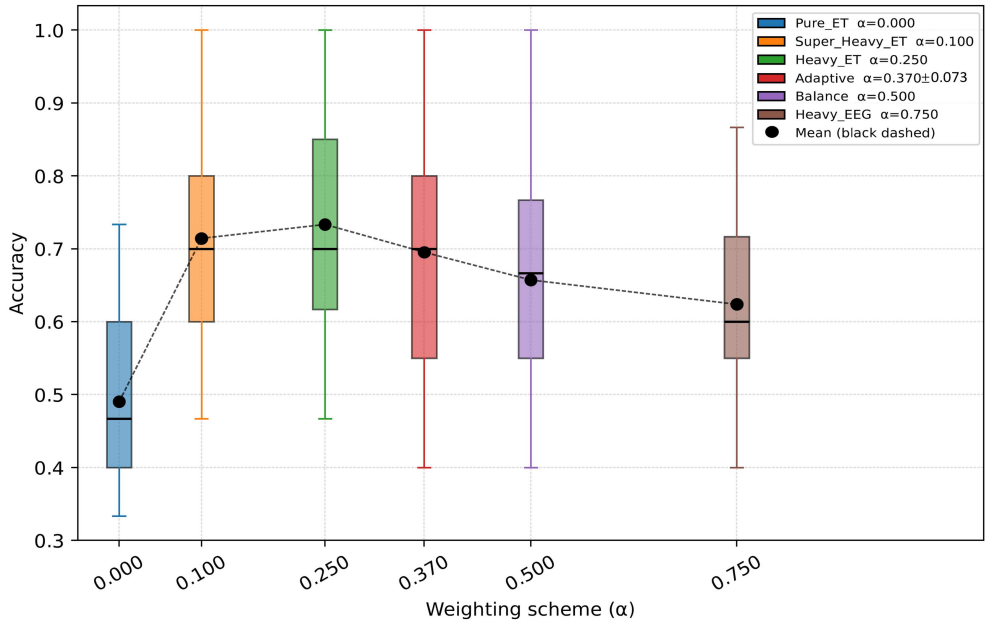

Figure 8 (Color online) Hybrid system performance under different weight schemes ($\text{bias}^{x/y} = 0.5\Delta$).

better suit the KDE-based fusion framework.

From a broader perspective, the asymmetrical and nonlinear performance of our system curve leads to a new research direction. Our findings indicate that the design of the optimal weighting mechanism for such hybrid systems should be based on the intrinsic properties of the processed data modalities, instead of being modeled as a simple linear relationship with performance metrics such as accuracy. Additionally, a deeper investigation of the nonlinear relationship in the weight-performance curve could provide valuable insights for the future optimization of the proposed hybrid system.

4.3 Optimization using LLMs

In this study, we also investigated the effect of the LLM-based prediction methods on the optimization of the Chinese hybrid BCI speller.

The results from the copy and free spelling sessions demonstrated that the LLM-based optimizations significantly improved the CS of the BCI speller (paired t-test, $p < 0.005$ and $p < 0.001$). In the copy spelling session, the average CS increased from 2.757 to 3.811 sinograms/min, indicating a 38.2% improvement in the CS. But compared with the free spelling session (from 3.039 to 5.813 sinograms/min in CS and from 13.96% to 41.98% in AUR), it demonstrated a slight increase. This indicates that using sufficient contextual information allows the LLM to better capture semantic relationships, thereby leading to improved word prediction performance and overall spelling efficiency. Our eLLM-BCI speller exhibited prominent performance among the current Chinese BCI spelling systems by achieving a communication speed of 5.813 sinograms/min. According to the Pinyin-based Chinese character encoding scheme, this was equivalent to 3.43 s to output a correct choice. Furthermore, during the free spelling session, the average STS exceeded 95% in both methods, showing no significant difference (paired t-test, $p > 0.05$). This indicates that the LLM-optimized BCI speller is not only faster than other spellers but also semantically accurate and relevant. This improvement is crucial for practical applications, as it ensures faster and more accurate input.

4.4 Limitations and future work

Our method exhibits some limitations and areas for improvement.

The primary objective of this study was to optimize the reactive BCI speller by combining ET data and LLMs. Therefore, the EEG decoder was not further optimized. The current EEG decoder is essentially a linear classifier based on the P300 detection, and its preprocessing steps are based on traditional baseline methods. In future work, more advanced classifiers such as the EEG-inception or improved preprocessing filters should be investigated to improve the accuracy of the EEG-only decoding, particularly for single-trial classification. Combining the above with our spatial distribution processing method for ET data could result in a more efficient BCI speller.

In this study, user comfort was prioritized by selecting a P300-based BCI system over other mainstream paradigms. Although this choice is critical for our target users (i.e., Chinese ALS patients), SSVEP-based systems could further improve ITR, especially in wider groups (such as healthy individuals with a greater tolerance for visual fatigue). Therefore, a promising future research direction is to integrate our framework into SSVEP-based spellers, potentially achieving high-speed performance with enhanced usability.

Additionally, constrained by the available computing resource, we simply conducted fine-tuning on the Big Science BLOOM model with 1.7B parameters, which is not competitive compared to the current large language models with tens of billions of parameters. Employing more advanced LLMs could further improve word prediction performance, thereby optimizing our system.

5 Conclusion

We presented a novel hybrid BCI speller for inputting Chinese characters that integrates ET and LLMs into the reactive BCI framework. Our approach improves traditional methods regarding the following aspects. (i) The ET device enables asynchronous operation, thereby improving user autonomy and decreasing dependence on external assistance. (ii) Combining ET data with EEG signals improves the decoding accuracy without increasing the cognitive load. Additionally, using our KDE-based method for the ET data, the proposed speller exhibits robustness against the sharp decreases in the performance of single-modal data. (iii) Our fine-tuned LLM significantly improves the efficiency of Chinese semantic expression by predicting associative words. Overall, this study demonstrates the effectiveness of combining eye tracking and LLMs to optimize Chinese BCI spelling systems, paving the way for more efficient Chinese communication systems for individuals with severe motor impairments.

Acknowledgements This work was supported by Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project (Grant No. 2022ZD0208903), National Natural Science Foundation of China (Grant Nos. U24A20339, T2521006), and Science and Technology Innovation Program of Hunan Province (Grant Nos. 2023RC1004, 2024QK2006).

References

1 Borgheai S B, McLinden J, Zisk A H, et al. Enhancing communication for people in late-stage ALS using an fNIRS-based BCI system. IEEE Trans Neural Syst Rehabil Eng, 2020, 28: 1198–1207
2 Luo S, Angrick M, Coogan C, et al. Stable decoding from a speech BCI enables control for an individual with ALS without recalibration for 3 months. Adv Sci, 2023, 10: 2304853
3 Khan M A, Das R, Iversen H K, et al. Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: from designing to application. Comput Biol Med, 2020, 123: 103843
4 Chen X, Yu Y, Tang J, et al. Clinical validation of BCI-controlled wheelchairs in subjects with severe spinal cord injury. IEEE Trans Neural Syst Rehabil Eng, 2022, 30: 579–589
5 Krusienski D J, Sellers E W, Cabestaing F, et al. A comparison of classification techniques for the P300 speller. J Neural Eng, 2006, 3: 299–305
6 Chen X, Chen Z, Gao S, et al. A high-ITR SSVEP-based BCI speller. Brain-Comput Interfaces, 2014, 1: 181–191
7 Chen Y, Yang C, Ye X, et al. Implementing a calibration-free SSVEP-based BCI system with 160 targets. J Neural Eng, 2021, 18: 046094
8 Chen X, Liu B, Wang Y, et al. A spectrally-dense encoding method for designing a high-speed SSVEP-BCI with 120 stimuli. IEEE Trans Neural Syst Rehabil Eng, 2022, 30: 2764–2772
9 Azadi Moghadam M, Maleki A. Fatigue factors and fatigue indices in SSVEP-based brain-computer interfaces: a systematic review and meta-analysis. Front Hum Neurosci, 2023, 17: 1248474
10 Blanco-Díaz C F, Guerrero-Méndez C D, Ruiz-Olaya A F. Enhancing P300 detection using a band-selective filter bank for a visual P300 speller. IRBM, 2023, 44: 100751
11 Gu Z, Chen Z, Zhang J, et al. An online interactive paradigm for P300 brain-computer interface speller. IEEE Trans Neural Syst Rehabil Eng, 2019, 27: 152–161
12 Shi M H, Zhou C L, Jiang M, et al. Advancement in the EEG-based Chinese spelling systems. In: Proceedings of the 9th International Conference on Intelligent Robotics and Applications, 2016. 110–117
13 Pan J, Li Y, Zhang R, et al. Discrimination between control and idle states in asynchronous SSVEP-based brain switches: a pseudo-key-based approach. IEEE Trans Neural Syst Rehabil Eng, 2013, 21: 435–443
14 Schettini F, Aloise F, Aricó P, et al. Self-calibration algorithm in an asynchronous P300-based brain-computer interface. J Neural Eng, 2014, 11: 035004
15 Yang C, Yan X, Wang Y, et al. Spatio-temporal equalization multi-window algorithm for asynchronous SSVEP-based BCI. J Neural Eng, 2021, 18: 0460b7
16 Hu L, Zhu J, Chen S, et al. A wearable asynchronous brain-computer interface based on EEG-EOG signals with fewer channels. IEEE Trans Biomed Eng, 2023, 71: 504–513
17 Zhou Y, He S, Huang Q, et al. A hybrid asynchronous brain-computer interface combining SSVEP and EOG signals. IEEE Trans Biomed Eng, 2020, 67: 2881–2892
18 He S, Zhou Y, Yu T, et al. EEG- and EOG-based asynchronous hybrid BCI: a system integrating a speller, a web browser, an e-mail client, and a file explorer. IEEE Trans Neural Syst Rehabil Eng, 2020, 28: 519–530
19 Mussi M G, Adams K D. EEG hybrid brain-computer interfaces: a scoping review applying an existing hybrid-BCI taxonomy and considerations for pediatric applications. Front Hum Neurosci, 2022, 16: 1007136
20 Zhang G, Garrett D R, Luck S J. Optimal filters for ERP research I: a general approach for selecting filter settings. Psychophysiology, 2024, 61: e14531
21 Wang H, Pei Z, Xu L, et al. Performance enhancement of P300 detection by multiscale-CNN. IEEE Trans Instrum Meas, 2021, 70: 1–12
22 Wang Z, Chen C, Li J, et al. ST-CapsNet: linking spatial and temporal attention with capsule network for P300 detection improvement. IEEE Trans Neural Syst Rehabil Eng, 2023, 31: 991–1000
23 Li S, Daly I, Guan C, et al. Inter-participant transfer learning with attention based domain adversarial training for P300 detection. Neural Netws, 2024, 180: 106655

24 Golenia J E, Wenzel M A, Bogojeski M, et al. Implicit relevance feedback from electroencephalography and eye tracking in image search. J Neural Eng, 2018, 15: 026002
25 Mannan M M N, Kamran M A, Kang S, et al. A hybrid speller design using eye tracking and SSVEP brain-computer interface. Sensors, 2020, 20: 891
26 Stawicki P, Gembler F, Rezeika A, et al. A novel hybrid mental spelling application based on eye tracking and SSVEP-based BCI. Brain Sci, 2017, 7: 35
27 Kalika D, Collins L, Caves K, et al. Fusion of P300 and eye-tracker data for spelling using BCI2000. J Neural Eng, 2017, 14: 056010
28 Wang J, Xu S, Dai Y, et al. An eye tracking and brain-computer interface-based human-environment interactive system for amyotrophic lateral sclerosis patients. IEEE Sens J, 2022, 23: 24095–24106
29 Yi Z, Pan J, Chen Z, et al. A hybrid BCI integrating EEG and eye-tracking for assisting clinical communication in patients with disorders of consciousness. IEEE Trans Neural Syst Rehabil Eng, 2024, 32: 2759–2771
30 Gunawardena N, Ginige J A, Javadi B. Eye-tracking technologies in mobile devices using edge computing: a systematic review. ACM Comput Surv, 2022, 55: 1–33
31 Xu X, Fang H. A P300-based BCI system for online Chinese input. J Huaqiao Univ (Nat Sci), 2015, 36: 269–274
32 Yu Y, Liu Y, Yin E, et al. An asynchronous hybrid spelling approach based on EEG-EOG signals for Chinese character input. IEEE Trans Neural Syst Rehabil Eng, 2019, 27: 1292–1302
33 Shi N, Wang L, Chen Y, et al. Steady-state visual evoked potential (SSVEP)-based brain-computer interface (BCI) of Chinese speller for a patient with amyotrophic lateral sclerosis: a case report. J Neurorestoratology, 2020, 8: 40–52
34 Wu B, Su Y, Zhang J, et al. A virtual Chinese keyboard BCI system based on P300 potentials. Acta Electronica Sinica, 2009, 37: 1733–1745
35 Jin J, Allison B Z, Brunner C, et al. P300 Chinese input system based on Bayesian LDA. Biomedizinische Technik BioMed Eng, 2010, 55: 5–18
36 Minett J W, Zheng H Y, Fong M C M, et al. A Chinese text input brain-computer interface based on the P300 speller. Int J Hum-Comput Interaction, 2012, 28: 472–483
37 Guo D, Chen B, Lu R, et al. Recurrent hierarchical topic-guided RNN for language generation. 2019. ArXiv:1912.10337
38 de Rosa G H, Papa J P. A survey on text generation using generative adversarial networks. Pattern Recognition, 2021, 119: 108098
39 Zhang H, Song H, Li S, et al. A survey of controllable text generation using transformer-based pre-trained language models. 2023. ArXiv:2201.05337
40 Li J, Tang T, Zhao W X, et al. Pre-trained language models for text generation: a survey. 2022. ArXiv:2201.05273
41 Speier W, Arnold C, Chandravadia N, et al. Improving P300 spelling rate using language models and predictive spelling. Brain-Comput Interfaces, 2018, 5: 13–22
42 Speier W, Arnold C W, Deshpande A, et al. Incorporating advanced language models into the P300 speller using particle filtering. J Neural Eng, 2015, 12: 046018
43 Brown T B, Mann B, Ryder N, et al. Language models are few-shot learners. 2020. ArXiv:2005.14165
44 Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 17th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019. 4171–4186
45 Caría A. Towards predictive communication: the fusion of large language models and brain-computer interface. Sensors, 2025, 25: 3987
46 Hong J, Wang W, Najafizadeh L. ChatBCI: a P300 speller BCI leveraging large language models for improved sentence composition in realistic scenarios. 2024. ArXiv:2411.15395
47 Parthasarathy N, Soetedjo J, Panchavati S, et al. High performance P300 spellers using GPT2 word prediction with cross-subject training. Brain-Comput Interfaces, 2024, 11: 210–224
48 Farwell L A, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography Clin NeuroPhysiol, 1988, 70: 510–523
49 Scao T L, Fan A, Akiki C, et al. BLOOM: a 176b-parameter open-access multilingual language model. 2022. ArXiv:2211.05100
50 Wang Y, Ke P, Zheng Y, et al. A large-scale Chinese short-text conversation dataset. 2020. ArXiv:2008.03946
51 Hu E J, Shen Y, Wallis P, et al. LoRA: low-rank adaptation of large language models. 2021. ArXiv:2106.09685
52 Han Z, Gao C, Liu J, et al. Parameter-efficient fine-tuning for large models: a comprehensive survey. 2024. ArXiv:2403.14608
53 Provilkov I, Emelianenko D, Voita E. BPE-Dropout: simple and effective subword regularization. 2019. ArXiv:1910.13267
54 Bai X, Li M, Qi S, et al. A hybrid P300-SSVEP brain-computer interface speller with a frequency enhanced row and column paradigm. Front Neurosci, 2023, 17: 1133933
55 Wolpaw J R, Ramoser H, McFarland D J, et al. EEG-based communication: improved accuracy by response verification. IEEE Trans Rehab Eng, 1998, 6: 326–333
56 Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using siamese BERT-networks. 2019. ArXiv:1908.10084
57 Niu Y F, He J X, Jin L. Magilock: a reliable control triggering method in multi-channel eye-control systems. Front Hum Neurosci, 2024, 18: 1365838