Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2026.050108
ISSN 2790-2692 Vol. 5, Issue 1: 71-78

A Hybrid Brain Computer Interface Character Input System Combining Eye Tracking and Motor Imagery

Jun Ye

Electrical and computer engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213, United States

Keywords: Eye tracking; motor imagery; hybrid brain-computer interface; character input system; multimodal fusion

Abstract: Brain-computer interface (BCI) character input system combining eye-tracking and motor imagery. Addressing the problems of low speed, high false triggering, significant fatigue, and large individual variability inherent in traditional single-paradigm BCIs for character input tasks, a three-level collaborative architecture of "eye-tracking coarse screening—motor imagery active confirmation—language model error correction enhancement" is proposed. The paper systematically reviews the development of non-invasive BCI spelling systems, the research progress of eye-tracking and hybrid interaction, and the latest trends in combining eye-tracking with character input tasks. Based on this, the key bottlenecks faced by the hybrid system are analyzed, including unstable eye-tracking localization, fluctuating motor imagery signals, multimodal temporal asynchrony, and insufficient efficiency of the complete input link. A unified modeling and fusion strategy is presented. Research shows that eye-tracking is suitable for rapid spatial localization, motor imagery is suitable for active confirmation, and language models can effectively improve error correction and associative input capabilities. Combined with publicly available research data, this type of system demonstrates high application value in terms of accuracy, throughput, and communication assistance potential. This article argues that future efforts should focus on improving synchronization accuracy, adaptive decoding, and closed-loop interaction design to advance hybrid brain-computer interface character input systems from experimental prototypes to practical applications.

1. Introduction

Brain-computer interface (BCI) character input systems are a research area in neuroengineering and intelligent interaction, aiming to provide stable, low-burden, and sustainable alternative communication methods for patients with severe movement disorders. Traditional EEG character input systems mostly employ single paradigms such as P300, SSVEP, or motor imagery. However, in real-world applications, they generally suffer from slow input speeds, high false trigger rates, fatigue after prolonged use, and significant individual variability. In recent years, eye tracking, with its natural spatial positioning, low interaction burden, and clear target orientation, has gradually entered BCI systems as a vehicle for visual screening or controlling priors. Furthermore, the motor imagery paradigm, due to its lack of continuous flashing stimulation and minimal interference with the external visual environment, still holds unique value, demonstrating its effectiveness in active confirmation and intent transmission. Combining eye tracking and motor imagery can create a complementary relationship in target localization and intent confirmation, thereby overcoming the structural contradiction between speed and robustness in single-modal character input systems. Based on the above ideas, research has begun to explore multimodal fusion, VR/desktop environment deployment, and assisted communication [1-3]. In this context, the design of a hybrid brain-computer interface system that combines eye tracking and motor imagery for character input tasks has both significant academic value and important engineering implications.

2. Current Status Analysis of the Research Topic

2.1 Development Context of Brain-Computer Interface Spelling Systems

The development of non-invasive BCI spelling system technology shows that the three paradigms of P300, SSVEP, and MI represent three different pathways: "passive induction-rapid steady-state response", "rapid steady-state-continuous response", and "active cognitive regulation". Maslova et al.'s review of BCI spellers suggests that hybrid design is the main way to improve the shortcomings of a single paradigm, compressing training time and improving user experience while ensuring accuracy. Wang et al.'s systematic review of the study on deep learning classification of motor imagery EEG shows that the MI paradigm has a long-term development prospect for spontaneous control and asynchronous interaction, but its stability, sample efficiency and cross-subject generalization level are still relatively low [5]. Cheng et al.'s network meta-analysis found that feature extraction and classification links are still important technical links that determine whether MI-BCI can be put into practical use, especially in small sample and online situations. These studies together show that as long as MI is used for high-dimensional character selection, it is difficult for the system to achieve both speed and accuracy at the same time.

2.2 Research Progress in Eye Tracking and Hybrid Interaction

Eye-tracking information can quickly provide the user's focus, making it very suitable for reducing candidate options and spatial positioning in character matrices. Larsen et al. used a VR environment to complete the synchronous acquisition of EEG and eye-tracking. The average data stream offset obtained in the experiment was 36ms, and the average jitter was 5.76ms, indicating that commercial devices can already support prototype-level mixed spelling applications in terms of timing. Reddy et al. used an XR scene to combine gaze and EEG to confirm the target, proving that neural signals can be used as a confirmation channel other than overt gaze to overcome the Midastouch problem [8]. The NeuroGaze system proposed by Coutray et al. shows that the interaction method of EEG and eye-tracking fusion can effectively reduce the damage caused by incorrect operation, but it has a time trade-off between speed and accuracy [9]. According to current research, eye-tracking is more suitable for solving the question of "where to look", while EEG is more suitable for answering the question of "whether to choose".

2.3 The trend of combining eye tracking with character input tasks

After eye tracking and BCI are combined, character input has evolved from the initial 2D screen-assisted selection to asynchronous input, semantic prediction and intelligent error correction. The asynchronous Chinese spelling system proposed by Hao et al. integrates eye tracking and large language model into reactive BCI, showing that the architecture of "coarse eye tracking localization + fine correction by neural signals or language model" is feasible[10]. In addition, the study by Zhang et al. uses a hybrid cursor control of eye tracking and MI, and the average recognition rate of the three categories is 87.92%, and the text input rate is 53.2 characters/min. It proves that eye tracking can perform continuous movement and MI can perform click and drag interaction logic on desktop input tasks[3]. Yi et al.'s hybrid EEG eye tracking interaction system also proves that when using multiple modes for interaction in clinical auxiliary communication, the effect is better than a single mode[2].

3. Raise questions

While the hybrid approach of eye tracking and motor imagery has shown promising results, four key issues remain when directly applied to character input systems. First, eye-tracking signals, while possessing high spatial resolution, are susceptible to subtle saccades, gaze jitter, and calibration drift, leading to unstable candidate character localization in dense character matrices. Second, motor imagery signals fluctuate significantly across time periods and subjects; using them as the sole confirmation signal can easily cause confirmation delays or false triggers. Third, inconsistencies in sampling rates, processing delays, and decision windows between the two modalities, without unified temporal control, make it difficult to achieve the fusion results required for online input. Fourth, most current systems prioritize single-shot recognition accuracy, rarely analyzing the efficiency of the closed loop of "gaze, confirmation, error correction, and associative recommendation" within the complete character input chain, resulting in usable systems with low character throughput. Therefore, the key factor determining the performance of a hybrid BCI character input system is not a single recognizer, but rather the lack of a unified design for multi-modal temporal collaboration, hierarchical decision-making mechanisms, and human-machine collaborative error correction.

4. Problem Solving/Strategies

4.1 System Overall Architecture Design

To address the issues raised above, a three-tiered character input system can be created: eye-tracking coarse screening, MI (Multi-modal input) active confirmation, and language model error correction enhancement. The first layer uses an eye-tracking module to coarsely locate character regions, compressing the full keyboard search space into a local candidate set. The second layer uses an MI module for selection confirmation, mapping left-hand and right-hand motor imagery and relaxation states to cancel, confirm, and hold states, respectively. The third layer uses a word-level or character-level language model for context prediction and error correction, reducing the burden on neural decision-making. This architecture follows the principle of "spatial prioritization, neural confirmation subsequent, and semantic optimization as a fallback," allowing each modality to perform its most suitable task, thereby improving the entire input process and reducing the likelihood of errors. Table 1 shows the corresponding relevant representative studies and the focus of this paper over the past three years.

4.2 Key Mathematical Models and Fusion Mechanisms

To ensure the online stability of the character input system, a unified model should be used to describe the motion image recognition, gaze confidence judgment, multimodal fusion decision-making, and system throughput. First, the output of the MI classifier is represented by posterior probabilities:

$$P(c_k \mid x) = \frac{\exp(z_k)}{\sum_{j=1}^{K} \exp(z_j)} \tag{1}$$

The classification of k-class motion imagery is represented by EEG feature vectors, and the k-th class of motion imagery is represented as p(c_k|x), which is the level of confidence in a certain state within the current time window. This formula is used to determine the boundary judgment of the three states: confirmation, cancellation, and idle. Second, eye-tracking localization uses the distribution ratio of fixation points within the candidate character region to determine the gaze confidence.

$$G_i = \frac{n_i}{N} \tag{2}$$

In the formula, the number of gaze points in the i-th candidate region is the number of effective gaze points in the i-th candidate region, and the total number of gaze points in the i-th candidate region is N. The larger G_i is, the more attention the user pays to the characters in that region. To combine the two sets of information, a weighted fusion decision function can be obtained.

$$F_i = \alpha G_i + \beta M_i + \gamma s_i \tag{3}$$

The fusion score F_i is the fusion score of the i-th character, α and β are the weights of eye tracking and MI, respectively, and s_i is the semantic prior score obtained from the language model. The system selects the character with the largest F_i as the output candidate within one decision cycle. To measure the effective communication capability of the character input system, the classic transmission rate can be used.

$$ITR = \frac{60}{T} I(P, N) \tag{4}$$

N can be the number of commands to choose from, P is the recognition accuracy, and T is the time spent making a character decision. This metric represents the relationship between speed and accuracy. Finally, for practical text input tasks, character throughput can be used.

$$CPM = \frac{C}{t/60} \tag{5}$$

Here, C represents the number of correctly entered characters, and t represents the total input time. CPM more closely reflects the real document entry process than ITR and is more suitable as an indicator of the final effect of a character input system.

4.3 Key Module Design and Implementation Strategies

From an implementation perspective, the eye-tracking module should use a combination of fixed-point duration threshold and dispersion threshold to reduce misselection caused by brief saccades, while the MI module should use sliding window and adaptive update to improve its online classification robustness [3][6]. Fusion decision should first perform time alignment and then make hierarchical decisions, and cannot simply splice the two original features together. The synchronization layer completes the time-stamp synchronization of the eye-tracking stream and the EEG stream, the perception layer outputs G_i and P(c_k|x), the decision layer uses the weighted fusion of equation (3) to make selection, and the interaction layer relies on the language model to produce associative candidate words. Users can significantly enhance the actual character processing speed by confirming the input of complete characters or phrases [10]. From an engineering perspective, the hierarchical system design is conducive to system parameter tuning and also conducive to personalized customization for different user groups.

4.4 Performance Discussion Based on Publicly Available Data

Table 1. Representative related works about the hybrid BCI character Input & Interaction.

Study	Year	Modalities	Task	Key Metric	Implication
Maslova et al.	2023	EEG speller review	Review	Hybrid spellers reduce training burden	Paradigm comparison
Zhang et al.	2024	ET + MI	Cursor typing	87.92% acc.; 53.2 CPM	Feasible desktop input
Yi et al.	2024	EEG + ET	Clinical communication	100% healthy; 76.1% patient acc.	Robust multimodal assistance
Larsen et al.	2024	EEG + VR ET	Hybrid speller in VR	36 ms offset; 5.76 ms jitter	Synchronization is critical
Reddy et al.	2024	Gaze + EEG	XR target selection	Intent confirmation improves selection	Mitigates Midas touch
Hao et al.	2025	ET + reactive BCI + LLM	Chinese spelling	Asynchronous semantic assistance	Supports predictive input

Table 1 summarizes representative studies from the past three years in terms of task morphology, modality combination, and key indicators. As shown in the table, the key to enabling the system to be applied to actual character input is not relying on a single classifier to improve accuracy, but rather on a closed loop involving accuracy, gaze stability, active confirmation channels, and semantic compensation. The desktop input results of Zhang et al. and the clinical communication results of Yin et al. both demonstrate the engineering usability and auxiliary value of this approach.

Table 2. Proposed hybrid character input system configuration

Layer	Module	Input	Output	Design Goal
Perception	Eye tracker	Gaze stream	Candidate region	Fast spatial localization
Decoding	MI classifier	EEG window	Confirm/Cancel/Idle	Active intention detection
Fusion	Decision engine	Gaze + MI + context	Character score	Stable online decision
Language	Predictive model	History text	Word suggestion	Error correction and acceleration
Interaction	Output manager	Accepted command	Typed character/word	Low-burden text entry

Table 2 presents the hierarchical configuration for character input tasks. The design focuses on using eye tracking to narrow down the selection, motor imagery to provide active confirmation, and a language model to reduce the burden of single neural decision-making. This division of labor avoids placing all the interaction complexity on a single EEG decoder, better aligning with the stability and maintainability of long-term online input systems.

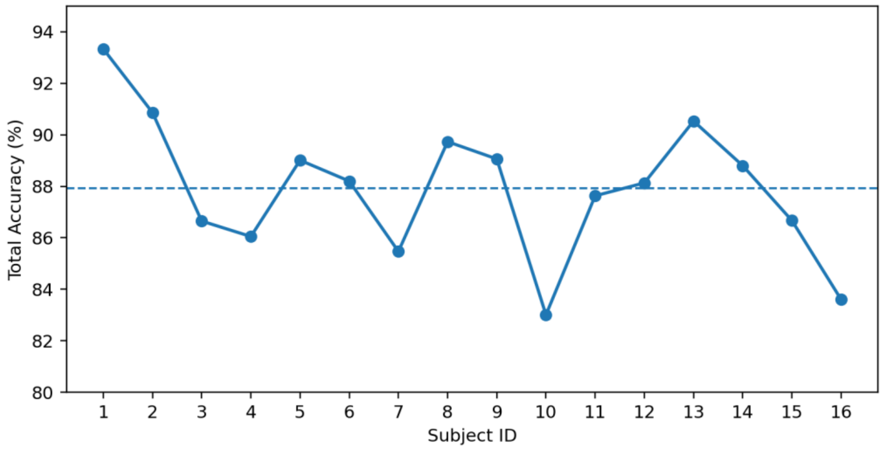

Figure 1. Zhang et al. (2024)'s total online accuracy on the 16th subject

Figure 1 shows the overall online recognition rate of 16 subjects, redrawn based on publicly available experimental data. Overall, the subjects' accuracy rates ranged around a mean of 87.92%, with a maximum of 93.33% and a minimum of 83.01%. Therefore, in online character input, the combination of eye-tracking and MI (Mimicry and Interference) can achieve both high average performance and a relatively acceptable range of fluctuation across subjects, laying a reliable foundation for the migration of character input interfaces to online systems.

Table 3. Reconstruct Key Performance Statistic for Recently opened report

Indicator	Value	Source	Interpretation
Average MI total accuracy	87.92 ± 2.67 %	Zhang et al., 2024	Stable online control baseline
Typing throughput	53.2 characters/min	Zhang et al., 2024	Shows practical text-entry potential
Patient communication accuracy	76.1 ± 7.9 %	Yi et al., 2024	Supports assistive communication
Synchronization offset	36 ms	Larsen et al., 2024	Acceptable multimodal alignment
Synchronization jitter	5.76 ms	Larsen et al., 2024	Low variance in device timing

Table 3 further summarizes and reorganizes the relevant indicators most closely related to the system design in this paper. It can be seen that to support realistic character input, it is not enough to rely solely on classification accuracy; throughput, clinical applicability, and multimodal synchronization quality also need to be considered. In particular, the 36ms offset combined with the 5.76ms jitter leads researchers to conclude that timing management itself becomes a crucial performance factor determining the final interactive experience.

5. Conclusion

The hybrid brain-computer interface (BCI) character input system, which combines eye-tracking and motor imagery, essentially creates a collaborative working mechanism that effectively coordinates spatial selection and cognitive confirmation. Existing research indicates that eye-tracking can effectively reduce the cost of target localization, motor imagery can provide a more proactive and safer confirmation channel, and language models can reduce input errors when neural signals are unstable [2-3][8-10]. This system addresses three aspects: improving the synchronization accuracy and low-latency processing capabilities of eye-tracking and EEG, designing a small-sample online adaptive MI decoding model, and forming a closed-loop interactive system with context prediction, error correction feedback, and user state awareness. Only by unifying hardware synchronization, signal decoding, interaction design, and language intelligence within a single structure can the hybrid BCI character input system evolve from a laboratory prototype into a long-term usable auxiliary communication tool.

References

[1] Liu, X., & Yang, D. (2025, March). LLM Data Strategy: Improving Data Availability and Efficiency. In Doctoral Symposium on Computational Intelligence (pp. 425-437). Singapore: Springer Nature Singapore.

[2] Yin, J. (2026, March). An Efficient Iterative Algorithm for Calibrating Agency MBS Estimation Parameters Based on Bayesian Optimization, Implemented in Python. In 2026 IEEE Madhya Pradesh Section Conference (MPCON) (pp. 1462-1467). IEEE.

[3] Zheng, H. (2026, April). Research on Cloud-Edge Collaborative Elastic Computing and Cost Optimization for High-Concurrency Scenarios. In 2026 IEEE 15th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 1489-1496). IEEE.

[4] Liu, H. (2025, December). Mlops Model Deployment System for Multi-Cloud Environments and Improvement of Commercial AI Service Availability. In 2025 IEEE International Conference on Communication Networks and Computing (CNC) (pp. 1047-1052). IEEE.

[5] Hong, Y. (2025, December). Chain Demand Mutation Identification and Emergency Response Decision-Making Integrated With Social Media Sentiment Analysis. In 2025 IEEE International Conference on Communication Networks and Computing (CNC) (pp. 425-432). IEEE.

[6] Chen, M. (2026, March). Design of a Privacy-Oriented AI Compliance Hook System Based on Static Code Analysis. In 2026 IEEE Madhya Pradesh Section Conference (MPCON) (pp. 648-654). IEEE.

[7] Lyu, N. (2026, April). Toward Robust AI Agents: A Closed-Loop Task Planning–Execution–Feedback Framework for Open Scenarios. In 2026 IEEE 15th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 1182-1188). IEEE.

[8] Zhang, K. (2025, December). Research on Cross-Selling Precision Marketing Strategy Based On Xgboost Model and SHAP Explanatory Framework. In 2025 IEEE 1st International Conference on Recent Trends in Computing and Smart Mobility (RCSM) (pp. 1-7). IEEE.

[9] Zhu, P. (2025, December). Construction of Multi-Scale Biostatistical Analysis Framework and its Application in Biomedical Signal Feature Recognition and Classification. In 2025 5th International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-7). IEEE.

[10] *Zhang, C., Han, J., Zou, Y., Dong, K., Li, Y., Ding, J., & Han, X. (2024, April). Detecting the anomalies in LiDAR pointcloud. In WCX SAE World Congress Experience. SAE Technical Paper.*

[11] *Wu, W. (2025, June). Construction and optimization of intelligent gateway software management platform based on jenkins cluster management under cloud edge integration architecture in industrial internet of things. In International Conference on 6G Communications Networking and Signal Processing (pp. 633-645). Singapore: Springer Nature Singapore.*

[12] *Pan, H. (2025, March). Research on Efficient Computing Model of Hartree Fock and Density Functional Theory Based on GPU Acceleration. In Doctoral Symposium on Computational Intelligence (pp. 485-496). Singapore: Springer Nature Singapore.*

[13] *Yiting Hong. Differentially Private High-Dimensional Business Data Publishing and Analysis Algorithm. International Journal of Business Management and Economics and Trade (2026), Vol. 7, Issue 1: 28-35.*

[14] *Wu, Y. (2025, October). Multi-Level Belief Rule Base Modeling Architecture and Intelligent Optimization Technology for Decision Support Systems. In 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-8). IEEE.*

[15] *Zhang, Q. (2025, October). Application of Reinforcement Learning in Dynamic Advertising Content Generation. In 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-5). IEEE.*

[16] *Zhang, Q. (2025, July). Research on Improving the Effectiveness of Programmatic Advertising Using Multi-armed Bandit Algorithms. In International Conference on Frontier Computing (pp. 543-552). Singapore: Springer Nature Singapore.*

[17] *Wang, Z. (2026). Exploration of the Application of Resource Circular Economy in the Agricultural Industry Chain.*

[18] *Wang, Y. (2025, December). Construction of an Intelligent Recommendation System for Personalized Training Plans for Sports Rehabilitation Patients Based on Graph Neural Networks. In 2025 IEEE 1st International Conference on Recent Trends in Computing and Smart Mobility (RCSM) (pp. 1-6). IEEE.*