



Research article

A novel brain-controlled prosthetic hand method integrating AR-SSVEP augmentation, asynchronous control, and machine vision assistance

Xiaodong Zhang^{a,c}, Teng Zhang^{b,c,*}, Yongyu Jiang^a, Weiming Zhang^a, Zhufeng Lu^a, Yu Wang^a, Qing Tao^d

^a School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shanxi, 710049, China

^b Zhejiang Normal University, Jinhua, Zhejiang, 321004, China

^c Shaanxi Key Laboratory of Intelligent Robot, Xi'an, Shanxi, 710049, China

^d School of Mechanical Engineering, Xinjiang University, Wulumuqi, Xinjiang, 830000, China

ARTICLE INFO

Keywords:

Brain-computer interface
Electroencephalogram
Steady-state visual evoked potential
Augmented reality
Pattern recognition
Asynchronous

ABSTRACT

Background and objective: The brain-computer interface (BCI) system based on steady-state visual evoked potentials (SSVEP) is expected to help disabled patients achieve alternative prosthetic hand assistance. However, the existing study still has some shortcomings in interaction aspects such as stimulus paradigm and control logic. The purpose of this study is to innovate the visual stimulus paradigm and asynchronous decoding/control strategy by integrating augmented reality technology, and propose an asynchronous pattern recognition algorithm, thereby improving the interaction logic and practical application capabilities of the prosthetic hand with the BCI system. **Methods:** An asynchronous visual stimulus paradigm based on an augmented reality (AR) interface was proposed in this paper, in which there were 8 control modes, including Grasp, Put down, Pinch, Point, Fist, Palm push, Hold pen, and Initial. According to the attentional orienting characteristics of the paradigm, a novel asynchronous pattern recognition algorithm that combines center extended canonical correlation analysis and support vector machine (Center-ECCA-SVM) was proposed. Then, this study proposed an intelligent BCI system switch based on a deep learning object detection algorithm (YOLOv4) to improve the level of user interaction. Finally, two experiments were designed to test the performance of the brain-controlled prosthetic hand system and its practical performance in real scenarios.

Results: Under the AR paradigm of this study, compared with the liquid crystal display (LCD) paradigm, the average SSVEP spectrum amplitude of multiple subjects increased by 17.41%, and the signal-noise ratio (SNR) increased by 3.52%. The average stimulus pattern recognition accuracy was $96.71 \pm 3.91\%$, which was 2.62% higher than the LCD paradigm. Under the data analysis time of 2s, the Center-ECCA-SVM classifier obtained $94.66 \pm 3.87\%$ and $97.40 \pm 2.78\%$ asynchronous pattern recognition accuracy under the Normal metric and the Tolerant metric, respectively. And the YOLOv4-tiny model achieves a speed of 25.29fps and a 96.4% confidence in the prosthetic hand in real-time detection. Finally, the brain-controlled prosthetic hand helped the subjects to complete 4 kinds of daily life tasks in the real scene, and the time-consuming were all within an acceptable range, which verified the effectiveness and practicability of the system.

* Corresponding author. Zhejiang Normal University, Jinhua, Zhejiang, 321004, China.

E-mail address: zhangteng@zjnu.edu.cn (T. Zhang).

<https://doi.org/10.1016/j.heliyon.2024.e26521>

Received 29 November 2022; Received in revised form 27 November 2023; Accepted 14 February 2024

Available online 19 February 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Conclusion: This research is based on improving the user interaction level of the prosthetic hand with the BCI system, and has made improvements in the SSVEP paradigm, asynchronous pattern recognition, interaction, and control logic. Furthermore, it also provides support for BCI areas for alternative prosthetic control, and movement disorder rehabilitation programs.

1. Introduction

Brain-computer interface (BCI) is based on the neural electrical activity generated by the brain, providing a communication and control pathway that does not depend on the conduction of peripheral nerves and muscles [1]. For some patients with normal brain function but physical disabilities, BCI technology can provide a way for people to interact with equipment or the environment [2]. As one of the most flexible parts of the human body, the hand covers a large part of the physical interaction between humans and the environment. Therefore, for patients with upper limb disabilities, it is a significant task to use BCI technology to restore their hand functions as much as possible. Many scholars have researched BCI-based prosthetic hand systems and achieved outstanding progress. In the first aspect, by studying the mechanism of brain cognition and neural activity, developing a user intention decoding algorithm, to use an electroencephalogram (EEG) to control an external intelligent prosthetic hand. For example, some scholars have made gratifying progress in the research on algorithm decoding accuracy [3,4], control logic naturalness and flexibility [5], and perception fineness [6]. The second aspect is the construction of BCI-driven neuro-rehabilitation prosthetic systems designed to modulate patients' brain activity and facilitate their neuro-rehabilitation, thereby restoring hand function. For example, some scholars have developed a BCI-controlled exoskeleton hand to help stroke patients with rehabilitation training, and they have also achieved exciting results [7,8].

The research in this paper belongs to the first aspect, and the representative paradigm in BCI-based prosthetic hand systems is steady-state visual evoked potential (SSVEP). It is an EEG component produced by the brain in response to steady-state visual stimulus, mainly in the occipital lobe (V1) [9]. SSVEP is an exogenous EEG signal that is largely unrelated to the spontaneous cognitive activity of the brain. Due to the obvious rhythmic assimilation effect of SSVEP, its characteristic information is easier to quantify, and it has a higher information transmission rate (ITR) and generalization ability [10]. At present, the research on SSVEP-BCI, especially in the control of external equipment based on the feedforward pathway, mainly includes visual stimulators, evoked paradigms, and SSVEP signal decoding [11–14].

Currently, the mainstream visual stimulators are liquid crystal displays (LCDs) and light-emitting diodes (LEDs). As the graphic display device of the modern computer system, LCDs cover most of the SSVEP research in recent years. With LCDs, the visual stimulus can be generated by computer programs with human-controlled precision and degrees of freedom (DOF) [15]. LCDs perform well in BCI systems for character input. Users can input characters to the computer by staring at the virtual keyboard on the LCD, and the ITR is higher than 300bit/min [16,17]. However, for the BCI used for external device control, especially the prosthetic hand system, the user needs to stare at the LCD on the side to receive visually evoked stimulus while manipulating the prosthetic hand, so the user cannot consider the LCD, the current environment, and the manipulated object at the same time. Unlike the SSVEP-BCI system used for character input, LCDs reduce the level of human-computer interaction during external device control and also increase the danger of certain scenarios [18].

AR-SSVEP paradigm enhancement. With the development of display devices, Augmented reality (AR) technology has injected fresh blood into the field of portable display [19]. Some scholars have tried to integrate AR and BCI and conducted feasibility studies and the development of related systems [20–23]. AR is a form of Virtual Environment (VE), which is mainly divided into Video see-through (VST) AR and Optical see-through (OST) AR. Both use Head-Mounted Displays (HMDs) to fuse virtual images into the real world, enabling users to view and interact with virtual elements and the real world in real time in the same field of view [24]. VST-AR combines an enclosed HMD with a head-mounted camera, while OST-AR uses a partially transmissive HMD placed in front of the user's eyes. Presenting visual stimuli to users via OST-AR can overcome the limitations of LCDs and improve the portability and flexibility of SSVEP-BCI [25].

Asynchronous pattern recognition and control of prosthetic hand. In the aspect of the visually evoked paradigm, scholars have carried out research on the graphical presentation and coding methods of visual stimulus. The representative ones include checkerboard stimulus [11], Newton's ring stimulus [26], sinusoidal coding, joint frequency-phase modulation (JFPM) method [27], joint frequency and space modulation [28], etc. However, most of the research on visually evoked paradigms is aimed at increasing the number of codes, classification accuracy, and information transfer rate. In a prosthetic hand BCI system, the user's interaction with the prosthetic hand and the environment is crucial. BCI can be divided into the synchronous system and asynchronous system according to whether there is a clock trigger signal. In a synchronous system, users rely on the received clock trigger signal to perform interactive operations [29]. Asynchronous systems do not have this limitation, and users can send or not send commands at their own will. This requires the BCI system to actively detect whether the user intends to enter a control command, that is, to detect the status of Non-intentional Control (NC) and Intentional Control (IC) [30]. To realize the user's initiative in the external device control of BCI, the construction of an asynchronous BCI system is crucial. At present, more and more scholars have begun to pay attention to this direction and have proposed asynchronous detection algorithms [31–33].

Machine vision assistance. We believe that in addition to asynchronous detection algorithms, excellent asynchronous interaction logic is more important. This determines the friendliness of the BCI system in practical applications. In most BCI systems used for external device control, researchers are more focused on allowing users to input all the commands of the BCI system [34,35]. However,

with the complexity of the interaction logic of the BCI system, too many instructions will make people unable to handle them. Therefore, the introduction of artificial intelligence (AI) can replace humans to make regular instructions under certain circumstances [36]. Object detection is one of the basic components of computer vision, which aims to detect the location and category of specific objects in an image through a camera [37]. Object detection has become the basis of many vision tasks, and it also provides more possibilities for the interaction of BCI systems.

There are some related studies, (1) in terms of user intent decoding methods: Jie Pan et al. improved a CCA method based on phase constraints. It fixes the phase of the sinusoidal reference signal by visual latency estimated from calibration data to better characterize the temporal characteristics of SSVEP [38]. Qingguo Wei et al. proposed a training data-driven CCA algorithm, which can be used to design spatial filters to improve the performance of SSVEP-based BCI [39]. To rationally use the fundamental frequency and harmonic components of SSVEP, Xiaogang Chen et al. proposed a filter bank CCA algorithm, and the algorithm was verified on a 40-target character input system [40]. In recent years, deep learning methods have also been introduced into SSVEP-BCI. Kwak et al. used a convolutional neural network (CNN) to perform pattern recognition on the EEG of SSVEP under dynamic stimulation conditions, with an accuracy rate of 94.03% [41]. Waytowich et al. constructed a compact CNN to perform pattern recognition on the EEG of SSVEP with twelve stimulation modalities and obtained an average accuracy of 80% [42]. The study found that many SSVEP decoding methods based on deep learning did not reflect a better recognition effect than CCA and its derivatives, but increased the complicated training process. (2) In terms of asynchronous control of the prosthetic hand: Parini et al. utilized a spatial filter based on a common spatial pattern (CSP) to maximize/minimize the variance of IC/NC signals and then classify the two [32]. Bin Xia et al. used the CCA algorithm to calculate the ratio between the second-largest coefficient and the largest coefficient and then classified the IC/NC status by comparing the relationship between predefined thresholds and ratio values [33]. Kaori et al. implemented the classification of IC/NC states by combining multiset CCA spatial filtering with a support vector machine (SVM) [43]. Xin Zhang et al. used CNN to identify IC/NC states and also achieved good results [31]. (3) In terms of AR-SSVEP paradigm enhancement and multi-technology integration: Yufeng Ke et al. designed a high-speed online AR-SSVEP paradigm, and verified the validity of this paradigm online by controlling the tasks of the manipulator [44]. Lingling Chen et al. proposed a robotic arm asynchronous control system based on SSVEP in an AR environment. The integration of AR-SSVEP paradigm enhancement technology and asynchronous control technology was realized on the control task of the manipulator [45]. Xiaogang Chen et al. combined AR-SSVEP paradigm enhancement technology and machine vision technology to design and implement a robotic arm control system, proving that the introduction of machine vision technology has played a positive role in the brain-controlled robotic arm system [46]. To sum up, although related research has made exciting progress, most of the research has not solved the above three levels of problems at the same time: firstly, AR-SSVEP paradigm enhancement technology to improve user intent decoding accuracy; Secondly, the prosthetic hand asynchronous control technology to optimize the flexibility of human-computer interaction; Finally, machine vision assistance technology to improve the efficiency of human-computer interaction.

Therefore, a novel brain-controlled prosthetic hand method based on AR-SSVEP enhancement, asynchronous control, and machine vision assistance was proposed. The major contributions of the paper can be summarized in three aspects. (1) An AR visual stimulus paradigm and (2) an EEG asynchronous decoding algorithm in the case of AR fusion were proposed. (3) The prosthetic hand manipulation interaction strategy was better constructed using object detection and behavior judgment. Specifically, eight continuous control modes and stimulation modes were set up for the prosthetic hand with 8 DOF. Based on this, an asynchronous visual stimulus paradigm in the AR environment was constructed, and the stimulus frequency and stimulus presentation method of each stimulus

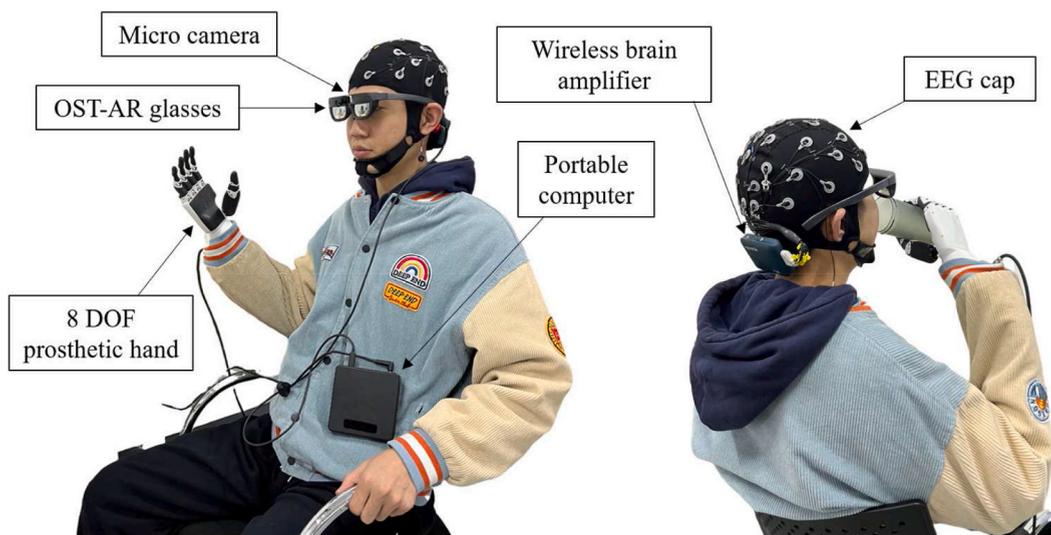


Fig. 1. The experimental platform. The experimental platform consists of a 32-channel EEG cap, an EEG amplifier, a pair of OST-AR glasses, a portable computer, an 8-DOF prosthetic hand, and a micro camera. All components are connected by wired or wireless means.

mode was designed. In addition, to achieve asynchronous decoding of EEG signals, we proposed an asynchronous pattern classifier based on center extended canonical correlation analysis and support vector machine (Center-ECCA-SVM). Then, with the object detection and behavior judgment model based on the YOLOv4 algorithm, the intelligent start-stop and stimulus paradigm switching of the visual stimulator was realized. Thus, the smooth control logic of the prosthetic hand BCI system was improved. The remainder of this paper is organized as follows. Section 2 describes the materials and methods, mainly including experimental materials, experimental paradigms, data processing steps, and related algorithms. The results are presented in Section 3. Remarks and discussions are presented in Section 4, followed by the conclusion in Section 5.

2. Materials and methods

2.1. Subjects and experimental platform

Twelve subjects (9 males, and 3 females, aged 23–28 years) in good physical condition with normal corrected visual acuity in both eyes participated in this experiment. Before the experiment, the subjects were informed about the content and procedure of the experiment, and simply wear AR glasses to get familiar with the visual stimulus interface. All experiments were performed in a quiet room without strong electromagnetic interference. Written informed consent was obtained from each participant before the experiment. The Institutional Review Board of Xi'an Jiaotong University approved the proposed experiment, and all experiments were conducted following the Declaration of Helsinki.

The experimental platform consists of a 32-channel EEG cap, an EEG amplifier (Neuracle, NeuSen W32), OST-AR glasses (Shixiang Technology, G350), a portable computer (Intel i5-1135G7, 2.4 GHz), an 8-DOF prosthetic hand (Inspire Robot Technology, RH56DFX-2R), and a micro camera (1080p, 30fps). The display refresh rate of the OST-AR glasses is 60 Hz, the resolution is 2560×1440 . The experimental platform is shown in Fig. 1. EEG is acquired by an EEG amplifier and sent to a portable computer via a wireless local area network (WLAN). The portable computer sends visual stimuli to AR glasses via the USB 3.0 protocol. The prosthetic hand receives control commands from the portable computer through the serial communication interface.

2.2. Mode design and visual stimulus

A flexible prosthetic hand with 8 DOF was selected as the control object of the BCI system. The DOF includes: (1) The flexion/extension DOF of the index finger, middle finger, ring finger, and little finger (4 DOF). (2) The flexion/extension DOF and sideways swing DOF of the thumb (2 DOF). (3) The pitch DOF and swing DOF of the wrist (2 DOF). We set up 8 control modes, namely, Grasp (Fig. 2 (a)), Put down (Fig. 2 (b)), Pinch (Fig. 2 (c)), Point (Fig. 2 (d)), Fist (Fig. 2 (e)), Palm push (Fig. 2 (f)), Hold pen (Fig. 2 (g)), and Initial (Fig. 2 (h)).

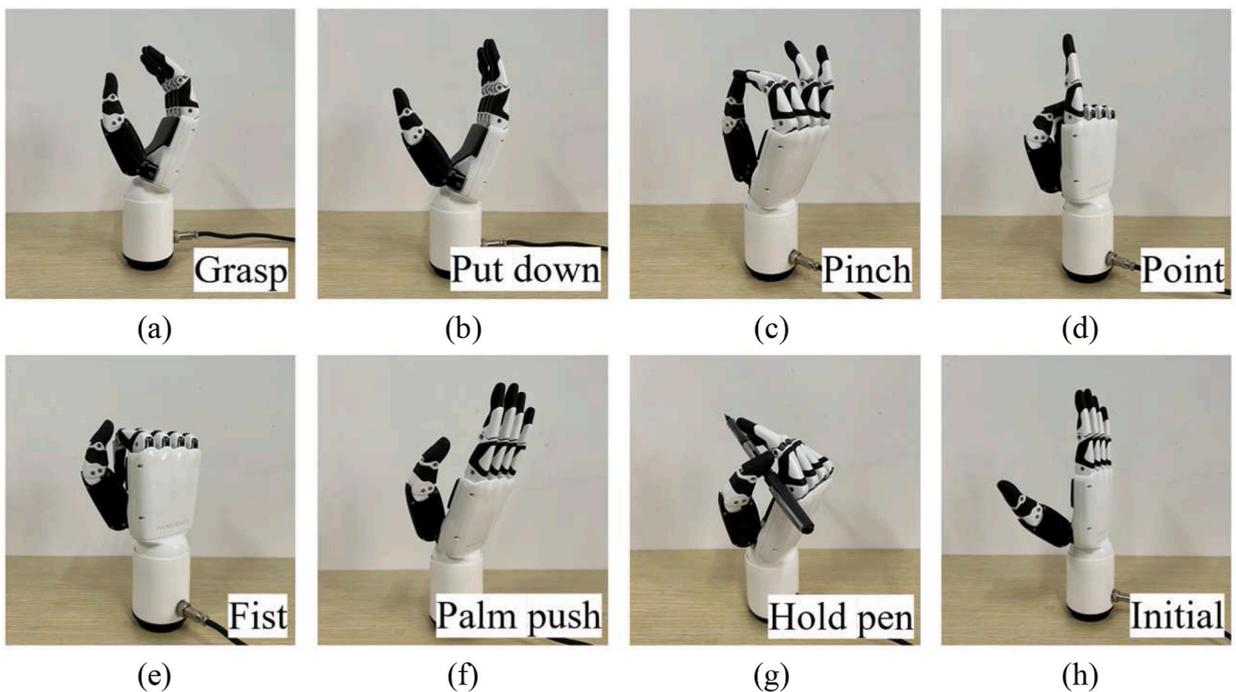


Fig. 2. The final pose state of the prosthetic hand in each control mode. There are 8 control modes, Figure (a) shows Grasp, Figure (b) shows Put down, Figure (c) shows Pinch, Figure (d) shows Point, Figure (e) shows Fist, Figure (f) shows Palm push, Figure (g) shows Hold pen and Figure (h) shows Initial.

Initial (Fig. 2 (h)). Each control mode corresponds to the stimulus mode with the same name, so there are 8 stimulus modes in total. The final pose state of the prosthetic hand in each control mode is shown in Fig. 2.

The images of each stimulus mode are presented in AR glasses using white-colored squares (aspect ratio 2:1), arranged in a 2×4 array, the layout method referred to previous research [18]. Silhouettes and names of hand movements are centered in squares to provide visual labels. The first-person view of the visual stimulus interface is shown in Fig. 3. The visual stimulus interface is presented to the user via optical see-through head-mounted displays (OST-HMD). The virtual interface is 45 cm away from the human eye. The user can directly see the environment and control objects through the area near the "+" symbol. We delineate patterns of visual stimulus interface regions. The area of the stimulus image array was set as the IC area, and the area near the "+" symbol where the environment could be directly seen was set as the NC area. In this paper, the NC mode was named "Center Mode".

Transparency-based sinusoidal modulation [47] was applied to the stimulus images to achieve flickering. The mathematical model can be described as:

$$F_i = \frac{1}{2} \times M_i \times \left(\sin \left(2\pi f_i \times \frac{n}{f_s} + \varphi_i \right) + 1 \right) \quad (1)$$

In the formula, $n = 1, 2, 3 \dots$, which represents the sequence of refresh frames of AR glasses. f_s represents the display refresh rate of the AR glasses, and n/f_s represents the n -th refresh time of the screen. F_i represents the RGB matrix of the i -th stimulus image after sinusoidal modulation. M_i represents the RGB matrix of the i -th stimulus image before sinusoidal modulation. f_i represents the frequency of the sinusoidal modulation of the i -th stimulus mode. φ_i represents the phase of the sinusoidal modulation of the i -th stimulus.

We take 11 Hz, 12 Hz, 13 Hz, 14 Hz, 15 Hz, 16 Hz, 17 Hz, 18 Hz as the modulation frequency, and take 0π 、 0.5π 、 π 、 1.5π 、 0π 、 0.5π 、 π 、 1.5π as the initial phase of each stimulus mode.

2.3. Experimental setup

In the experiments, the EEG cap and amplifier were used for EEG acquisition. Electrode distribution followed the International 10–20 system [48]. The reference electrode (Ref) was located on the scalp near the parietal lobe. The ground electrode (GND) was located on the scalp near the prefrontal lobe. The P3, Pz, P4, PO3, PO4, O1, Oz, and O2 channels near the occipital lobe were selected as the EEG acquisition channels. The sampling frequency was 1000 Hz. The electrode impedance was less than 15 k Ω .

There are two experiments, the first experiment carried out the comparison and evaluation of technical indicators by setting a control group and an experimental group. The second experiment was mainly for specific tasks in real-life scenarios, and the purpose was to verify the effectiveness and convenience of the developed method and the brain-controlled prosthetic hand system.

2.3.1. The first experiment

There were two paradigms in the first part experiment, the experiment in the AR environment was set as the experimental group, and the experiment in the LCD environment was set as the control group. The purpose was to compare and verify the effectiveness of SSVEP signal decoding under AR visual stimulus. In addition, to verify the asynchronous SSVEP pattern recognition algorithm proposed in this paper, the experimental content about EEG acquisition in Center mode was added to the experimental group. The experimental paradigm is shown in Fig. 4.

Experimental group: Each subject wore an EEG cap and OST-AR glasses, and sat in a chair naturally and relaxed, facing in any direction. They gazed at stimulus images in AR glasses or the external environment according to the cues in the experimental

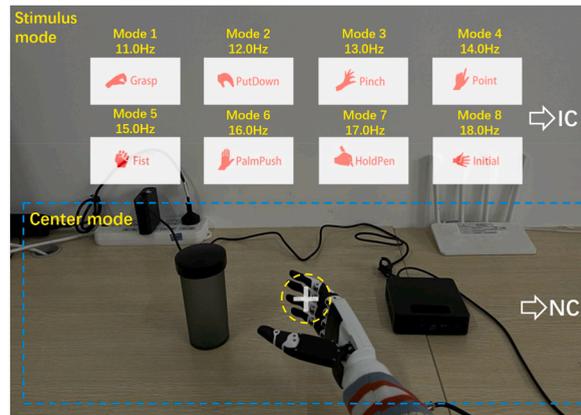


Fig. 3. The first-person view of the visual stimulus interface. The images of each stimulus mode are presented using white-colored squares (aspect ratio 2:1), arranged in a 2×4 array. Silhouettes and names of hand movements are centered in squares to provide visual labels. The user can directly see the environment and control objects through the area near the "+" symbol. The upper half of the interface is the IC mode area, and the lower half is the NC mode area.

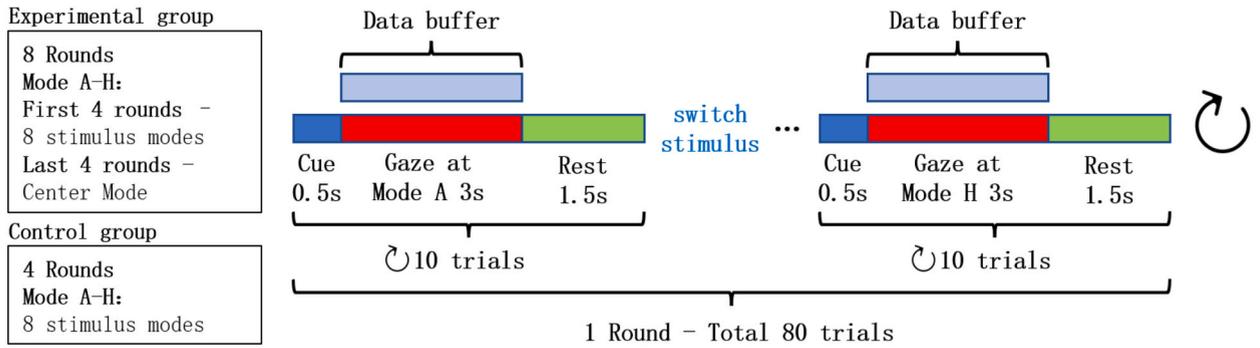


Fig. 4. The experimental paradigm. There are two paradigms in the experiment, the experiment in the AR environment is set as the experimental group, and the experiment in the LCD environment is set as the control group. The blue part represents a visual cue in the interface that reminds the subject. The red part represents that the subject needs to follow the cue to look at the corresponding mode. The green part represents rest. Patterns A-H have different meanings in the experimental and control groups, see notes in the figure for details. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

paradigm. Blink freely during the experiment. Each subject completed 8 rounds of the experiment, with a 5-min rest between each round. Each round of experiments consisted of 80 trials. In the first 4 rounds, they were asked to watch 8 stimulus modes in turn, with 10 trials for each pattern. Before each trial, there was a cue of the stimulus label (duration 0.5s). In the last 4 rounds, subjects were asked to focus on the prosthetic hand through a see-through area centered on the "+" symbol. The stimulus time of each trial was 3s. There was a 2s rest period between trials. The first 4 rounds of experiments represent the experimental data of 8 stimulus modes (that is, the data of the IC state), and the last 4 rounds of experiments represent the experimental data of the Center mode (that is, the data of the NC state). IC and NC status data are not mixed together.

Control group: Each subject sat relaxed in a chair and looked at the LCD screen of a 13-inch laptop. The screen was placed 45 cm away from the subject's eyes. The visual stimulus pictures on the LCD screen were arranged in the same way as in the AR paradigm. And the details are shown in Fig. 5. Subjects gazed at stimulus images in the LCD according to cues in the experimental paradigm. The subjects were free to blink during the experiment. Each subject completed 4 rounds of experiments, with a 5-min rest between each round. Each round of experiments consisted of 80 trials. Subjects were asked to watch 8 stimulus modes in turn, with 10 trials for each pattern. Before each trial, there was a cue of the stimulus label (duration 0.5s). The stimulus time of each trial was 3s. There was a 2s rest period between trials.

2.3.2. The second experiment

We set 4 common life scene tasks, namely, Pouring water - Drinking water (Fig. 6 (a)), Opening the door (Fig. 6 (b)), Taking a pen - Writing (Fig. 6 (c)), and Typing (Fig. 6 (d)), as shown in Table 1. Four subjects were randomly selected from the 12 subjects to participate in the second experiment. And the brain-controlled prosthetic hand system developed in this paper was used to perform these 4 tasks in sequence. To facilitate the evaluation of the performance of the method and system, the sequence of the prosthetic hand patterns in each task was set, and the subjects were informed in advance. The sequence of prosthetic hand modes prescribed in each task is shown in Table 1 ("Hold" means that the user communicates the "action hold" command to the prosthetic hand through the NC mode, that is, the user only observes the environment and the prosthetic hand through the AR glasses, and does not change the action

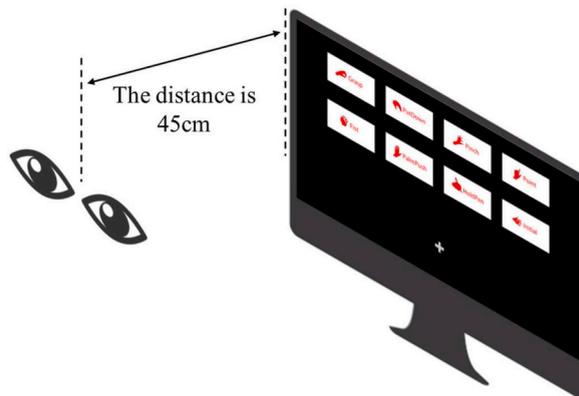


Fig. 5. Schematic diagram of the visual stimulation interface of the LCD screen and the spatial position of the human eye. The visual stimulus pictures on the LCD screen are arranged in the same way as in the AR paradigm.

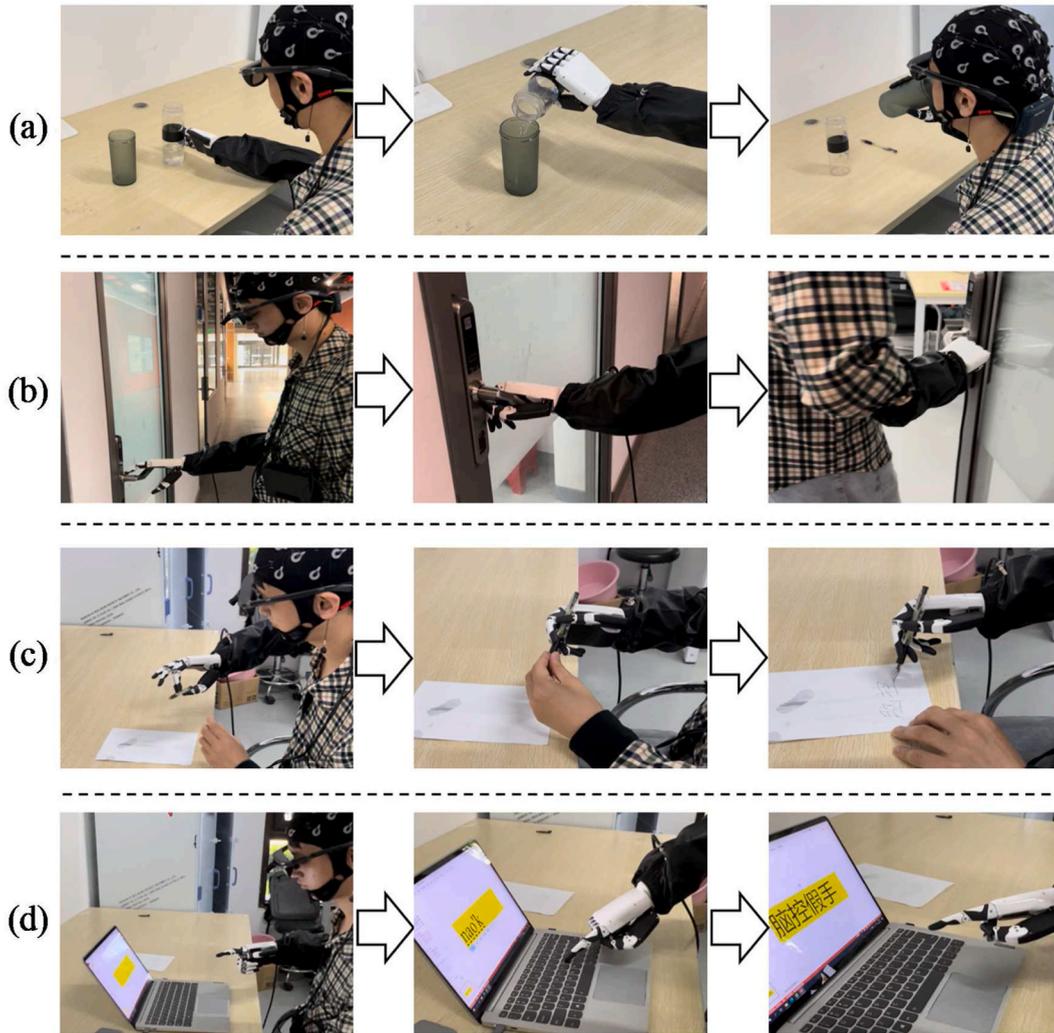


Fig. 6. The second experimental scenario. Figure (a) shows the task scenario of Pouring water - Drinking water. Figure (b) represents the task scenario of Opening the door. Figure (c) shows the task scenario of Taking a pen - Writing. Figure (d) represents the task scenario of Typing.

Table 1
Introduction of 4 real-life scenario tasks.

Task	Description	Instruction Sequence
(1) Pouring water - Drinking water	The user controls the prosthetic hand to pick up the cup and pour a fixed amount of water into another cup, then put down the original cup, pick up the cup filled with water and complete the drinking action, and finally put the cup back in place.	Hold – Grasp – Hold – Put down – Hold – Grasp – Hold – Put down - Initial
(2) Opening the door	The user controls the prosthetic hand to grab the door handle, open the door, then push the door to enter, and finally release the door handle.	Hold – Fist – Hold – Put down - Initial
(3) Taking a pen - Writing	The user controls the prosthetic hand to pick up the pen on the table, maintains a good pen-holding posture with the assistance of the left hand, then writes the prescribed Chinese characters on the paper, and finally puts the pen back on the table.	Hold – Pinch – Hold – Put down – Hold - Hold Pen – Hold – Put down - Initial
(4) Typing	The user controls the prosthetic hand to perform a “Point” gesture, and taps the keyboard to type out the prescribed Chinese characters.	Hold – Point – Hold - Initial

of the prosthetic hand. This mode can appear when the two actions of the prosthetic hand switch, or the user’s subjective intention). The experimental scene is shown in Fig. 6. The time-consuming of each task was recorded, and if the “cup” or “pen” dropped halfway, the task was judged to be failed.

2.4. Data preprocessing

Signal preprocessing included (1) removal of trend term and (2) 8–40 Hz Butterworth band-pass filter. The preprocessing process was designed to remove noise, artifacts, and power-frequency interference in the EEG [49].

2.5. Center-ECCA-SVM algorithm for pattern recognition

2.5.1. ECCA algorithm

ECCA [50,51] is the abbreviation of extended canonical correlation analysis, which is an algorithm based on CCA [52], and additionally adds the individual SSVEP signals of each stimulus mode as individual reference templates. Then the comprehensive correlation coefficient between the SSVEP signal to be classified and the standard reference template and the personal reference template is obtained. The ECCA algorithm can realize SSVEP stimulus pattern recognition with reference to individual characteristics.

First, the algorithm needed to obtain the user's personal reference template $\bar{V}^{(q)}$ and the standard reference template $Y^{(q)}$ for each stimulus pattern. $\bar{V}^{(q)}$ and $Y^{(q)}$ were obtained by averaging the pre-collected samples of each stimulus mode of the subject.

Then, according to the calculation process of the CCA algorithm, optimization and dimension reduction was performed between the SSVEP signal to be classified X , the personal reference template $\bar{V}^{(q)}$ and the standard reference template $Y^{(q)}$, respectively. The optimization criterion was to maximize the canonical correlation coefficient ρ_{max} . Six spatial filter coefficients were obtained by singular value decomposition (SVD): $W_X^1, W_{\bar{V}^{(q)}}^1, W_X^2, W_{Y^{(q)}}^1, W_{\bar{V}^{(q)}}^2, W_{Y^{(q)}}^2$. The optimization formula is as follows:

$$\arg \max_{W_X^1, W_{\bar{V}^{(q)}}^1} \rho_{max}(X, \bar{V}^{(q)}) \arg \max_{W_X^2, W_{Y^{(q)}}^1} \rho_{max}(X, Y^{(q)}) \arg \max_{W_{\bar{V}^{(q)}}^2, W_{Y^{(q)}}^2} \rho_{max}(\bar{V}^{(q)}, Y^{(q)}) \rho_{max}(S_1, S_2) = \frac{W_{S_1}^T cov(S_1, S_2) W_{S_2}}{\sqrt{(W_{S_1}^T cov(S_1, S_1) W_{S_1}) (W_{S_2}^T cov(S_2, S_2) W_{S_2})}} \quad (2)$$

Among them, the spatial filter coefficients $W_X^1, W_X^2, W_{\bar{V}^{(q)}}^2$, and $W_{Y^{(q)}}^1$ were selected, and the correlation coefficient vector $\hat{r}^{(q)}$ was obtained according to the following formula:

$$\hat{r}^{(q)} = \begin{bmatrix} r_1^{(q)} \\ r_2^{(q)} \\ r_3^{(q)} \\ r_4^{(q)} \end{bmatrix} = \begin{bmatrix} \rho(W_X^{2T} X, W_{Y^{(q)}}^1 Y^{(q)}) \\ \rho(W_X^{1T} X, W_X^{1T} \bar{V}^{(q)}) \\ \rho(W_X^{2T} X, W_X^{2T} \bar{V}^{(q)}) \\ \rho(W_{\bar{V}^{(q)}}^2 \bar{V}^{(q)}, W_{\bar{V}^{(q)}}^2 \bar{V}^{(q)}) \end{bmatrix} \quad (3)$$

Then, the comprehensive correlation coefficient $r^{(q)}$ of the SSVEP signal to be classified X in each stimulus mode was obtained. As shown in the following formula:

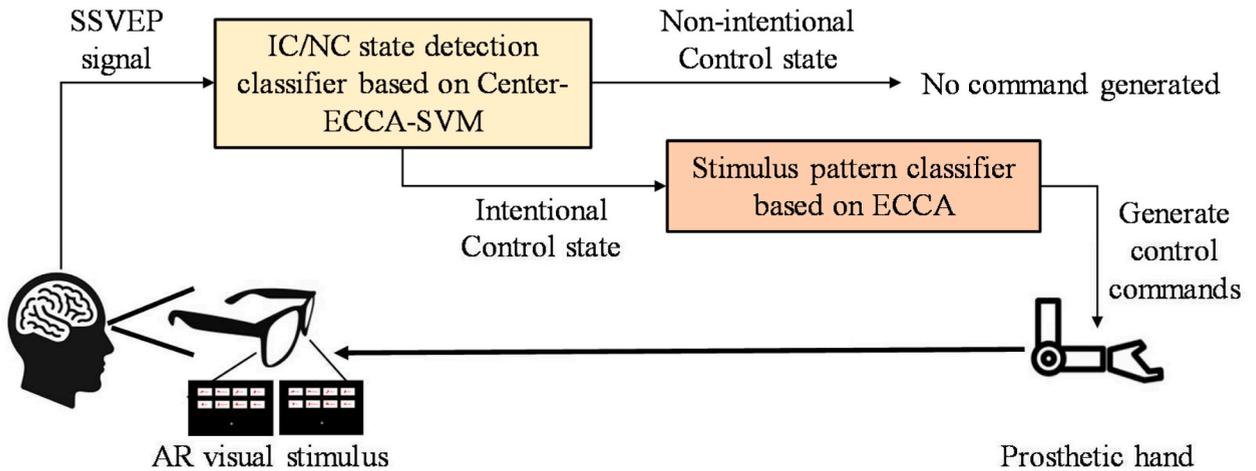


Fig. 7. The asynchronous paradigm strategy. When the BCI system recognizes the current EEG signal as the NC mode, it suspends the recognition of the stimulus mode and avoids interfering operations with external devices. When it is recognized as IC mode, the system sends control commands to the prosthetic hand according to the recognition result of the stimulus mode.

$$r^{(q)} = \sum_{u=1}^4 \text{sign}(r_u^{(q)}) \cdot (r_u^{(q)})^2 \tag{4}$$

The final pattern recognition result τ was obtained according to the following formula, which means that the recognition result was the τ -th stimulation pattern.

$$\tau = \underset{q}{\text{arg max}} r^{(q)}, q = 1, 2, 3, \dots, Q \tag{5}$$

where Q represents the number of stimulus modes in the SSVEP paradigm.

2.5.2. Asynchronous paradigm strategy

In the application of prosthetic hand manipulation, the BCI system needed to have a mode without control command output. This means that the user conveyed the operation intention of “no command output” to the BCI system. In daily use, this mode was indispensable and occupied a high proportion. In asynchronous BCI systems, this mode was usually named NC mode. Instead, the mode in which the user intends to enter commands was named IC mode [53]. When the BCI system recognized the current EEG signal as the NC mode, it suspended the recognition of the stimulus mode and avoided interfering operations with external devices. When it was recognized as IC mode, the system sent control commands to the prosthetic hand according to the recognition result of the stimulus mode. The details are shown in Fig. 7.

Attentional orienting refers to the active and selective allocation of attention to the current specific visual field by the human eye. Different attentional orienting has different effects on processing in the visual cortex of the brain. Attention can be divided into overt attention and covert attention according to whether directional behavior can be observed through head or eye movements. When the human eye receives periodic visual stimuli with overt attention or covert attention, SSVEP features at corresponding frequencies are generated in the EEG. Compared with covert attention, overt attention contributes more to the perceptual effect in the center of the retina, and the SSVEP feature magnitude is higher [54].

Each visual stimulus appeared in the user’s field of vision. And the unfixed stimulus image could also affect the user’s covert attention at the same time. When the user was in the IC state (setting the stimulus mode of fixation as A), overt attention in stimulus mode A resulted in a spectral energy boost at this frequency in the EEG. When the user was in the NC state, the control object and the environment became the overt attention of the user, and the covert attention of the 8 stimulus images also had additive effects on the EEG.

Therefore, referring to the above characteristics, we proposed an asynchronous pattern classifier based on Center-ECCA-SVM.

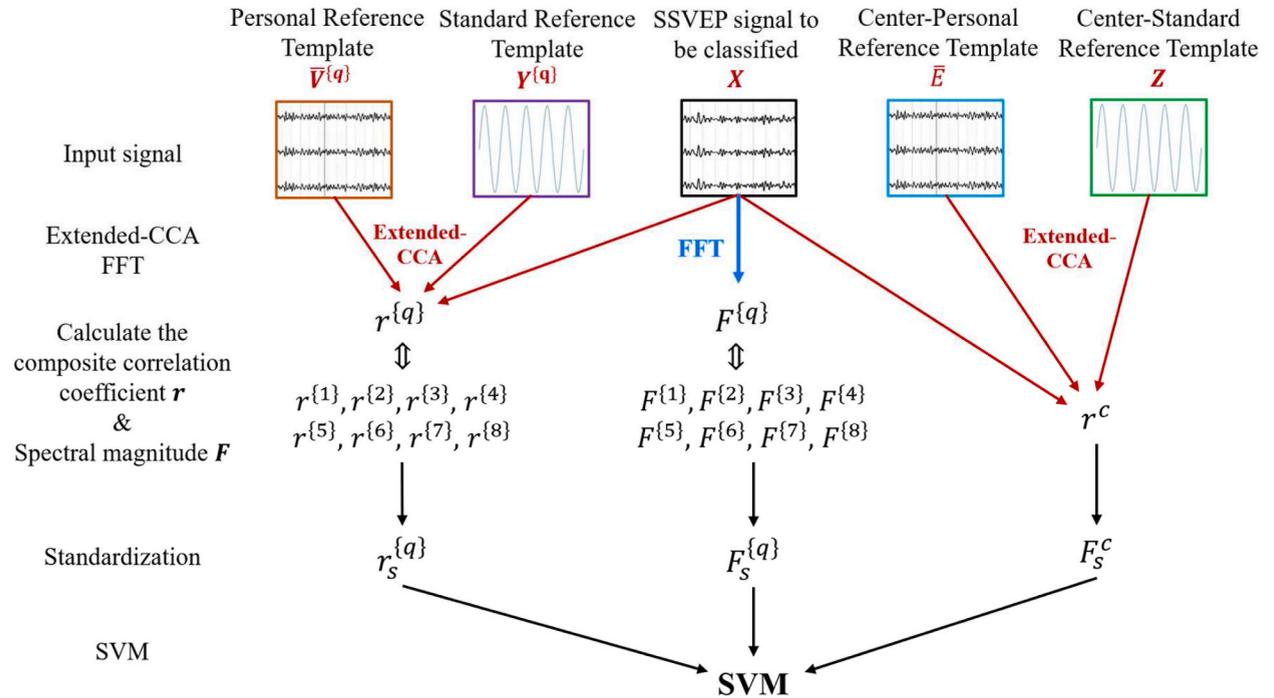


Fig. 8. The flowchart of the Center-ECCA-SVM algorithm. The process described in the figure mainly reflects the calculation process of the input feature vector of the SVM model, also known as the feature extraction process.

2.5.3. Center-ECCA-SVM algorithm

The Center-ECCA-SVM algorithm added a personal reference template for covert attention (named Center-personal reference template in this paper) and a standard reference template for covert attention (named Center-standard reference template in this paper) to the ECCA algorithm. The flowchart of the Center-ECCA-SVM algorithm is shown in Fig. 8.

Center-personal reference template \bar{E} represents that when the user was in the NC state (when the user gazed at the environment and the prosthetic hand through the "+" area), the EEG signal was affected by the covert attention of the 8 stimulus images. This phenomenon resulted in the mixed superposition of multiple weak SSVEP features. The Center-personal reference template was obtained by averaging the pre-collected EEG samples in Center mode.

Center-standard reference template Z was obtained by superimposing the standard reference template signals of each stimulus mode. The formula is shown below:

$$Z = \begin{bmatrix} \sum_{q=1}^{q=Q} \cos(\omega^{(q)} \bullet m) \\ \sum_{q=1}^{q=Q} \sin(\omega^{(q)} \bullet m) \\ \vdots \\ \sum_{q=1}^{q=Q} \cos(L \bullet \omega^{(q)} \bullet m) \\ \sum_{q=1}^{q=Q} \sin(L \bullet \omega^{(q)} \bullet m) \end{bmatrix}, \quad \omega^{(q)} = 2\pi f^{(q)} / f_s \quad (6)$$

In the formula, Z represents the Center-standard reference template. Q represents the number of reference templates (number of stimulus modes). $\omega^{(q)}$ represents the angular frequency of the q -th stimulus mode. m represents the time series, $m = 1, 2, 3, \dots, M$, where M was the number of sampling points for a single trial. L represents the number of harmonics of the Center-standard reference template. $f^{(q)}$ represents the stimulus frequency of the q -th stimulus mode. f_s represents the signal sampling frequency.

The specific pattern recognition process of the Center-ECCA-SVM algorithm is as follows:

- (1) The ECCA operation was performed on the SSVEP signal to be classified X with a personal reference template $\bar{V}^{(q)}$ and standard reference template $Y^{(q)}$, and the comprehensive correlation coefficient $r^{(q)}$ of each stimulus mode was obtained.
- (2) The ECCA operation was performed on the SSVEP signal to be classified X with Center-Personal Reference Template \bar{E} and Center-Standard Reference Template Z , and the comprehensive correlation coefficient r^c of Center mode was obtained.
- (3) Fast Fourier transform (FFT) was performed on the SSVEP signal to be classified X , and the spectral amplitude $F^{(q)}$ at the stimulus frequency corresponding to each stimulus mode was obtained.
- (4) $r^{(q)}$, r^c , $F^{(q)}$ was standardized (mean equal to 0, standard deviation equal to 1) to obtain $r_s^{(q)}$, $F_s^{(q)}$, F_s^c .
- (5) $r_s^{(q)}$, $F_s^{(q)}$, F_s^c as input to the SVM model and then trained.

The algorithm involved 8 stimulus patterns, so the algorithm constructed an asynchronous pattern classifier by obtaining 9 kinds of comprehensive correlation coefficients and the spectral amplitudes of 8 kinds of frequencies.

2.6. Algorithm training and statistical analysis

The training process of the Center-ECCA-SVM algorithm is as follows:

- (1) *Feature extraction process*: According to the algorithm process, the comprehensive correlation coefficient and the spectral amplitude of each stimulus frequency were obtained to form the feature vector of each sample. The feature vectors of the training set samples were then normalized. The expectation and standard deviation of the normalization process were then applied to the test set samples to obtain the normalized feature vectors of the test set samples.
- (2) *The training process of SVM*: The input data of SVM was $N \times F$ standardized feature vectors (N represents the number of training samples, and F represents the number of features). The kernel function of SVM adopted Radial Basis Function (RBF). The grid parameter optimization method based on cross-validation was adopted to determine the optimal C parameter and gamma parameter. The variation range of parameter C was $[2^{-6}, 2^6]$. The variation range of the parameter gamma was $[2^{-6}, 2^6]$. The parameter tuning process adopted three-fold cross-validation. The exponential step size of meshing for both parameters was 1. SVM model training was performed using MATLAB's LIBSVM toolbox.

We used two training forms to explore the performance of the algorithm, respectively training the SVM model with different proportions of training samples. Because IC mode could be subdivided into 8 stimulus modes, while Center mode had no subdivision. Therefore, to reduce the time-consuming of subjects in the experiment, the proportion of NC training samples was appropriately reduced. The training samples are divided into IC/NC state disparate samples (4:1) and balanced samples (2:1).

Before SVM training, the EEG samples of the 8 stimulus modes were assigned 8 labels respectively. EEG samples in Center mode were assigned one label. There were a total of 9 classification labels. In the model testing stage, 8 stimulus modes were output as IC mode, and Center mode was output as NC mode.

We separately trained and tested the model on the EEG data of all subjects. We trained and tested the model according to 4-fold cross-validation [55]. The data analysis time of the classifier was 2.0s, and the electrodes selected 8 channels (P3, Pz, P4, PO3, PO4, O1, Oz, O2) in the occipital lobe area.

In addition, considering the actual application of the prosthetic hand system, it may cause danger if the NC mode was misidentified as the IC mode. For example, when the user manipulated the prosthetic hand to keep holding the cup in the NC state, if it was mistaken for a put-down instruction, the cup would fall. However, if the IC state was mistakenly recognized as the NC state, the system keeps the original state. Although it was a misidentification, it was less dangerous.

Therefore, in this paper, we set a novel metric to judge the performance of the classifier, named “tolerant classification accuracy”. Under this metric, the misidentification of IC state as NC state was tolerated as correct, as a rough measure of the safety level of the asynchronous pattern classifier for the prosthetic hand system.

2.7. Intelligent start-stop method of visual stimulator

For the peripheral control of the prosthetic hand, the friendliness of human-computer interaction is crucial. (1) It should be considered that wearing AR glasses for a long time and receiving visual stimuli will interfere with the user’s daily life. (2) And the asynchronous pattern recognition algorithm cannot achieve 100% accuracy, so an intelligent system switch is necessary. Therefore, we realized the intelligent start-stop of the visual stimulator based on object detection and behavior judgment. When wearing AR glasses, users could avoid the continuous flickering of stimulating images, and comfort was improved. The misrecognition probability of

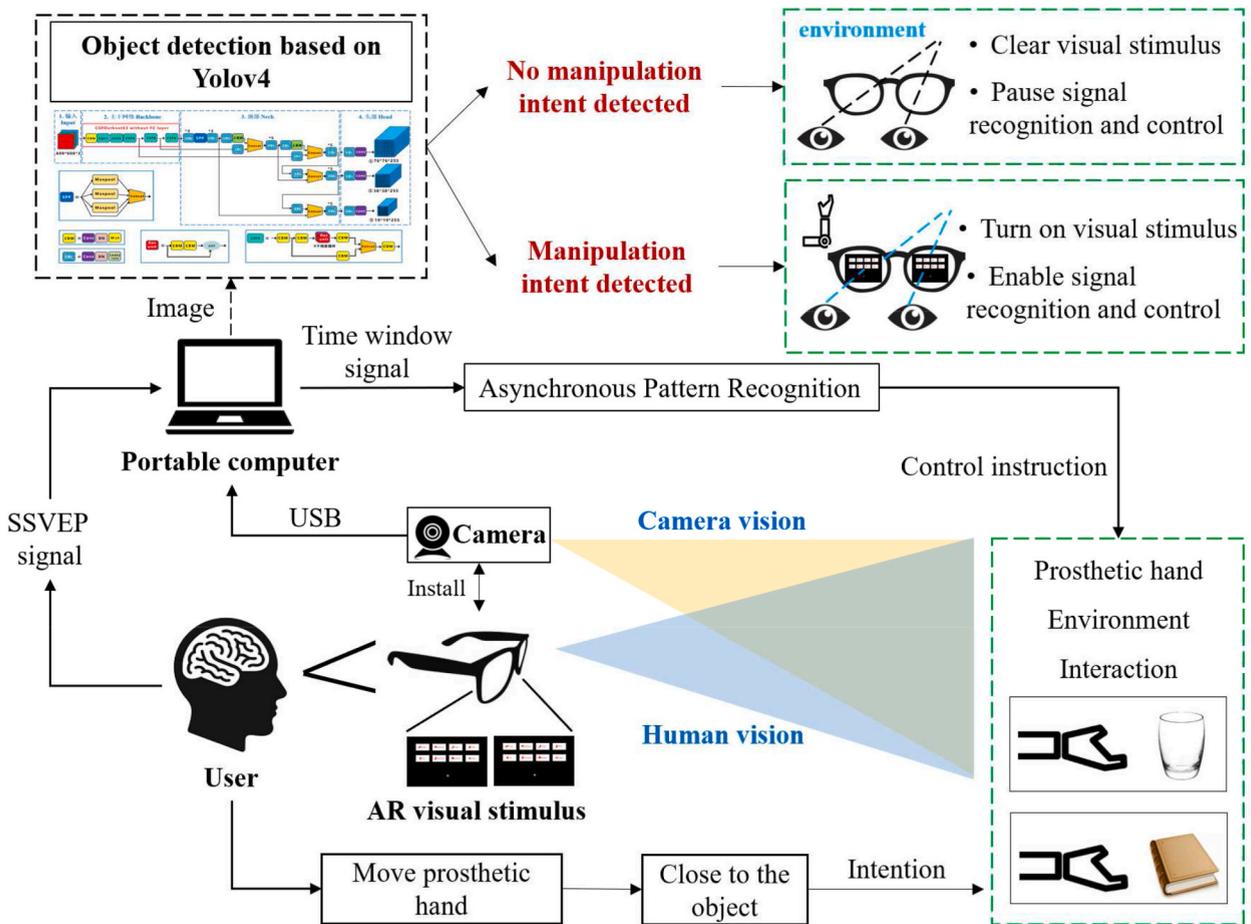


Fig. 9. The smart start-stop strategy of the visual stimulator. A miniature camera is integrated on the AR glasses, and the image is transmitted to the portable computer in real time for object detection. When the user wants to manipulate the prosthetic hand, he needs to bring the prosthetic hand close to the grasped object. If the user does not want to control the prosthetic hand, just move the prosthetic hand out of the field of view. If it is detected that the prosthetic hand and the interactive object enter the field of view of the camera at the same time, the host computer program performs behavior judgment and changes the standby state of the visual stimulator.

asynchronous pattern recognition could be reduced.

The smart start-stop strategy of the visual stimulator is shown in Fig. 9. A miniature camera was integrated into the AR glasses, and the image was transmitted to the host computer in real-time for object detection. When the user wanted to manipulate the prosthetic hand to perform a task, he needed to use the arm and elbow joints to drive the prosthetic hand close to the grasped object. If the user did not want to control the prosthetic hand, just move the prosthetic hand out of the field of view. If it was detected that the prosthetic hand and the grasped object enter the camera field of view at the same time, the host computer program performed behavioral judgment and changed the standby state of the visual stimulator.

The object detection model adopted YOLOv4 [56]. The process of model construction and training is as follows: (1) data set acquisition, (2) data set calibration, (3) data augmentation [57], (4) dividing the data set and test set, (5) k-means clustering [58] to obtain the prior frame of the data set, and (6) YOLOv4 network training. Considering the computing power load, we trained and verified the YOLOv4 and YOLOv4-tiny [56] models.

The image dataset was obtained by shooting videos of various angles, action switching, and interaction with various objects of the prosthetic hand and intercepting them in frames. The original samples are 6340 frame pictures. The manual dataset calibration software used LABELIMG [59]. Data augmentation generated twice the number of samples by random contrast, brightness, and image flipping. The training set and test set accounted for 85% and 15% of the total samples, respectively, and were divided randomly. The YOLOv4 and YOLOv4-tiny model was built with the DARKNET framework [56]. YOLOv4 and YOLOv4-tiny models run on CPU (Intel i5-1135G7).

3. Result

3.1. Comparison of EEG features between AR and LCD paradigms

In the previous section, we divided the experimental group and the control group. After averaging, data preprocessing, and fast Fourier transform (FFT), the spectral amplitude and signal-noise ratio (SNR) of the average EEG signal of each stimulus mode were obtained. The comparison of SSVEP features of AR and LCD visual stimulus paradigms is shown in Table 2, and the histogram is shown in Fig. 10.

The EEG signal evoked by the LCD paradigm (control group) had an average spectral amplitude of 1.3091 μ V and an average SNR of 11.9537 dB at the stimulus frequency. The EEG signals evoked by the AR paradigm (experimental group) had an average spectral amplitude of 1.5371 μ V and an average SNR of 12.3745 dB at the stimulus frequency. Compared with the control group, it increased by 17.41% and 3.52% respectively.

3.2. Comparison of EEG decode between AR and LCD paradigms

To verify the performance of the AR paradigm in EEG decoding, the ECCA algorithm was used as a stimulus pattern classifier to decode the EEG signals of multiple subjects under the AR and LCD paradigms. At different data analysis times (0.3s–3.0s), the average classification accuracy and ITR of the two paradigms are shown in Table 3, and the comparison histogram is shown in Fig. 11 (a) and Fig. 11 (b).

Both the LCD paradigm and the AR paradigm achieved the best accuracy in the data analysis time of 3s, which were $94.24 \pm 4.72\%$ and $96.71 \pm 3.91\%$, respectively. The classification accuracy of stimulus patterns for both paradigms improved with the increase of data analysis time. When the data analysis time is less than 0.8s, the accuracy of the LCD paradigm is better than that of the AR paradigm, but the accuracy of both is relatively low. When the data analysis time is longer than 1.0s, the AR paradigm is better than the LCD paradigm.

Both paradigms obtained the highest ITR in the data analysis time of 1.5s, which are $60.19 \pm 12.34\text{bit}/\text{min}$ and $67.47 \pm 14.00\text{bit}/\text{min}$, respectively. The same as the accuracy rate, when the data analysis time is less than 0.8s, the ITR of the LCD paradigm is better than that of the AR paradigm. When the data analysis time is longer than 1.0s, the AR paradigm is better than the LCD paradigm.

Table 2

The comparison of SSVEP features of AR and LCD visual stimulus paradigms.

Stimulus mode	AR paradigm		LCD paradigm	
	Spectral magnitude (μ V)	SNR (dB)	Spectral magnitude (μ V)	SNR (dB)
Grasp/11 Hz	2.1355	16.1002	1.7575	12.1402
Put down/12 Hz	1.4673	12.3245	1.4223	10.5597
Pinch/13 Hz	1.7631	12.0741	1.3483	12.3182
Point/14 Hz	1.4182	9.2222	0.8893	12.6861
Fist/15 Hz	1.7145	11.3432	1.6611	13.308
Palm push/16 Hz	1.6687	12.7218	1.2505	12.107
Hold pen/17 Hz	1.2466	12.3544	0.9474	10.3119
Initial/18 Hz	0.8826	12.8555	1.1966	12.1981
Average value	1.5371	12.3745	1.3091	11.9537

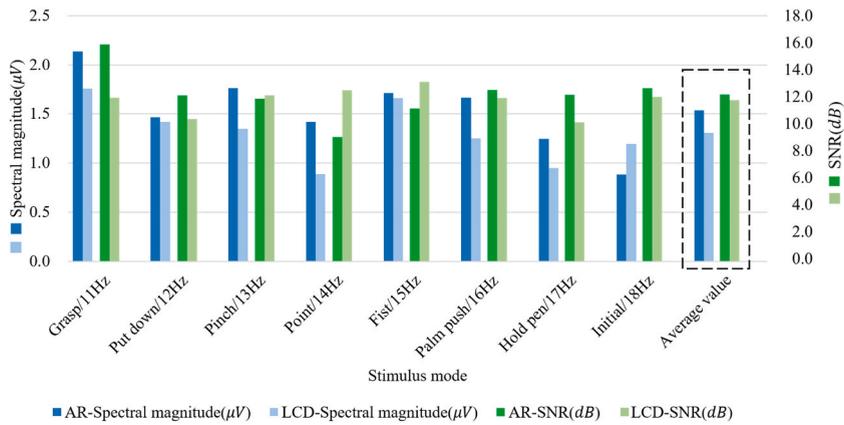


Fig. 10. Histogram of SSVEP feature contrast under AR and LCD visual stimulus paradigms. The dark blue and light blue legends represent the spectral magnitudes at eigenfrequencies in the AR and LCD paradigms, respectively, as measured by the principal axes. The dark and light green legends represent the SNR at eigenfrequencies in the AR and LCD paradigms, as measured by the secondary axis. The black dashed box represents the mean value. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3

The average classification accuracy and ITR of the two paradigms at different data analysis time (0.3s–3.0s).

Data analysis time (s)	LCD paradigm		AR paradigm	
	Accuracy (%)	ITR (bit/min)	Accuracy (%)	ITR (bit/min)
0.3	27.29 ± 5.76	8.49 ± 5.08	26.20 ± 4.74	7.95 ± 4.26
0.5	45.90 ± 9.45	29.16 ± 9.29	42.26 ± 3.87	24.01 ± 5.44
0.8	64.99 ± 10.03	49.99 ± 14.52	63.89 ± 7.80	48.90 ± 12.78
1.0	70.92 ± 8.91	52.56 ± 13.79	72.33 ± 11.31	56.52 ± 18.93
1.2	75.94 ± 9.66	53.94 ± 15.06	81.42 ± 9.49	64.31 ± 15.89
1.5	85.70 ± 7.47	60.19 ± 12.34	89.25 ± 8.28	67.47 ± 14.00
1.8	88.52 ± 6.29	56.43 ± 10.27	92.74 ± 6.16	64.04 ± 10.04
2.0	90.23 ± 5.95	54.34 ± 8.57	94.84 ± 4.92	62.15 ± 7.79
2.2	90.90 ± 5.06	51.21 ± 7.02	95.46 ± 5.53	58.83 ± 8.33
2.5	92.54 ± 3.81	48.15 ± 6.58	96.03 ± 4.73	53.68 ± 6.69
3.0	94.24 ± 4.72	43.21 ± 5.74	96.71 ± 3.91	46.81 ± 5.01

3.3. Comparison of EEG decode between ECCA and other algorithms

To verify the superiority of the ECCA algorithm, under the AR-SSVEP paradigm, CCA, filter bank canonical correlation analysis algorithm (FBCCA) [40], task-related component analysis algorithm (TRCA) [60], and the sum of squared correlations algorithm (SSCOR) [61] were selected to compare the decoding performance. Firstly, the hyperparameters of each algorithm were set to ensure its optimal performance, and then the classification effect was tested on the EEG signals of the AR-SSVEP discrete stimulation mode. Table 4 represents the hyperparameter selection and tuning range settings for the above algorithms.

Fig. 12 (a) and Fig. 12 (b) respectively show the average decoding accuracy and ITR of various algorithms in different data analysis time. It shows that with the increase of data analysis times, the average correct rate of all algorithms shows an increasing trend. At the same time, due to the common influence of the average correct rate and the data analysis time, the ITR of all algorithms shows a trend of increasing first and then decreasing. Moreover, since the increase of data analysis time will reduce the real-time performance of system control, it is not possible to blindly increase the data analysis time to improve classification accuracy. The SSVEP-BCI system needs to select an appropriate data analysis time. In detail, under the data analysis time of 0.8–2.0s, the ECCA (proposed in this paper) algorithm has better average classification accuracy and ITR than the FBCCA and CCA algorithms. Under the data analysis time of 2.2s–3.0s, the performance of the FBCCA algorithm is slightly better than ECCA. The TRCA and SSCOR algorithms are inferior in the average classification accuracy and ITR. To sum up, the ECCA algorithm has shown excellent classification performance in the discrete EEG signal recognition of the AR-SSVEP paradigm, especially in the case of short data analysis time, it has the best decoding performance. By comprehensively considering the classification accuracy rate, ITR, and real-time interaction of prosthetic hand control scenarios in different data analysis times, this paper selects the data analysis time of 2s for follow-up research.

3.4. Asynchronous pattern recognition results

The Center-ECCA-SVM asynchronous pattern classifier was divided into two training sample proportions: disparate samples and balanced samples. The data analysis time of the classifier was taken as 2.0s. There were 9 categories of pattern recognition, including 8

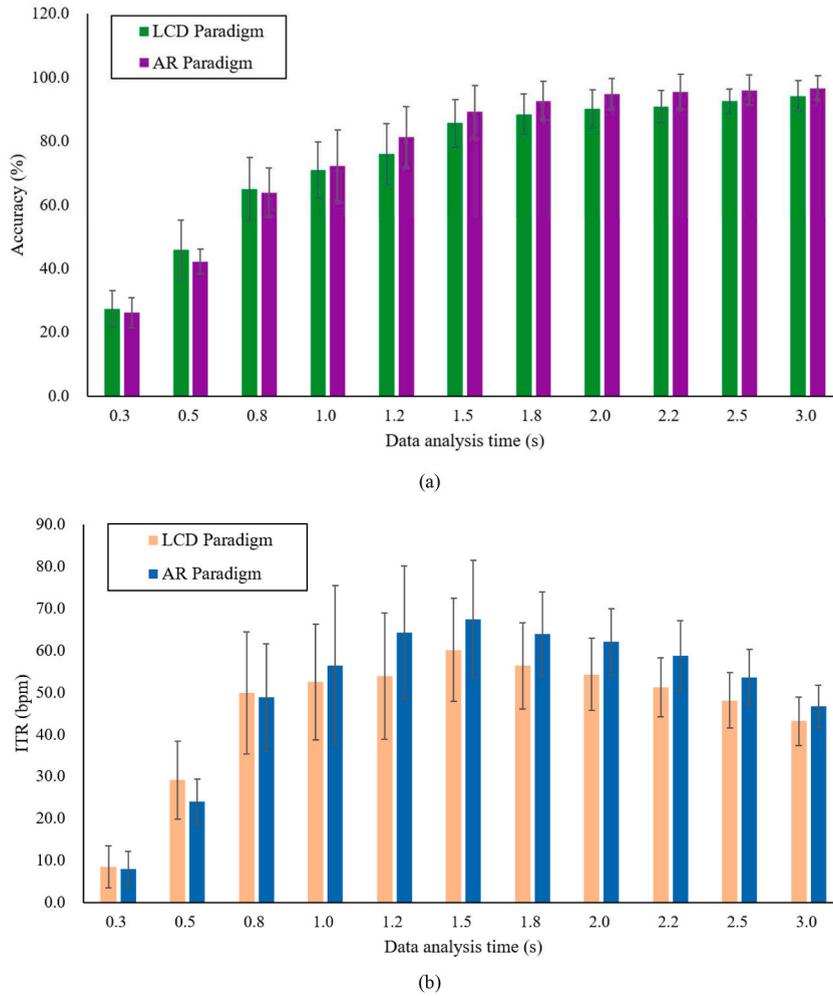


Fig. 11. Histogram of the average classification accuracy and ITR of the two paradigms under different data analysis time (0.3s–3.0s). Figure (a) shows the comparison of classification accuracy, with the green and purple legends representing the LCD and AR paradigms, respectively. Figure (b) represents the comparison of ITR, with orange and blue color legends representing LCD and AR paradigms, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4
Hyperparameter selection and tuning range settings for each algorithm.

Algorithm	Settings
CCA	Select the number of template harmonics: 1~4
FBCCA	Select the number of Filter Bank sub-bands: 1~5, the number of template harmonics: 1~4
TRCA	Select the number of Filter Bank sub-bands: 1~5
SSCOR	Select the number of Filter Bank sub-bands: 1~5
ECCA	Select the number of template harmonics: 1~4

categories of IC status and 1 category of NC status. The pattern recognition accuracy of the classifiers with the two sample proportions under the normal metric and the tolerant metric is shown in [Table 5](#).

The comparison histogram is shown in [Fig. 13](#). Classifiers trained on balanced samples achieved the best classification performance on multiple subjects. The average accuracy was $94.66 \pm 3.87\%$ under the normal metric and $97.40 \pm 2.78\%$ under the tolerant metric. Subject S03 obtained the best classification results, the accuracy under the two metrics were 98.13% and 100%, respectively.

3.5. Evaluation of object detection model

The YOLOv4 and YOLOv4-tiny models were trained, and the training results are shown in [Fig. 14 \(a\)](#) and [Fig. 14 \(b\)](#) respectively.

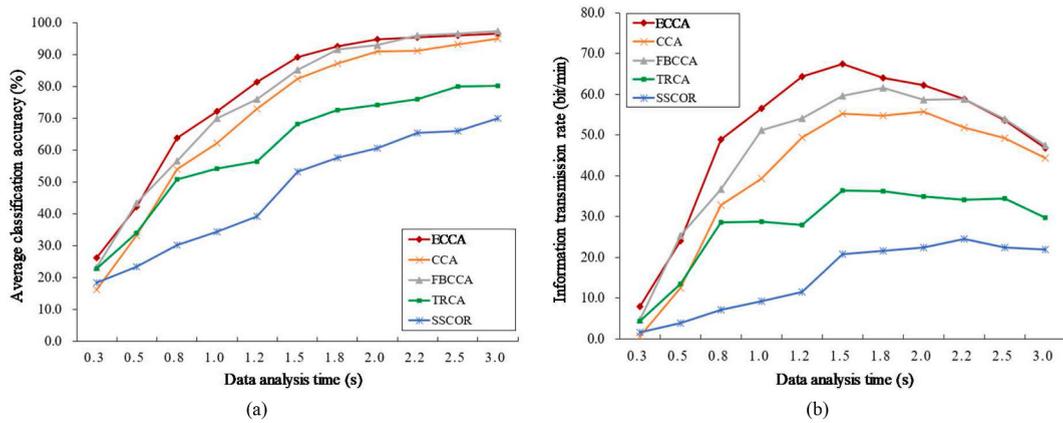


Fig. 12. Figure (a) represents the average decoding accuracy of various algorithms in different data analysis time. Figure (b) represents the average ITR of various algorithms in different data analysis time.

Table 5

The pattern recognition accuracy of the classifiers with the two sample proportions under the normal metric and the tolerant metric (for all subjects who only participated in the first experiment).

Subject	Disparate samples		Balanced samples	
	Normal metric	Tolerant metric	Normal metric	Tolerant metric
S01	97.00%	99.00%	97.50%	99.17%
S02	95.25%	98.75%	94.58%	98.13%
S03	97.75%	99.50%	98.13%	100.00%
S04	87.50%	94.00%	87.92%	93.96%
S05	95.00%	96.75%	96.88%	99.58%
S06	94.50%	98.50%	97.50%	98.33%
S07	96.50%	98.00%	95.21%	97.71%
S08	85.75%	90.00%	89.58%	92.29%
Average value	93.66% ± 4.50	96.81% ± 3.26	94.66% ± 3.87	97.40% ± 2.78

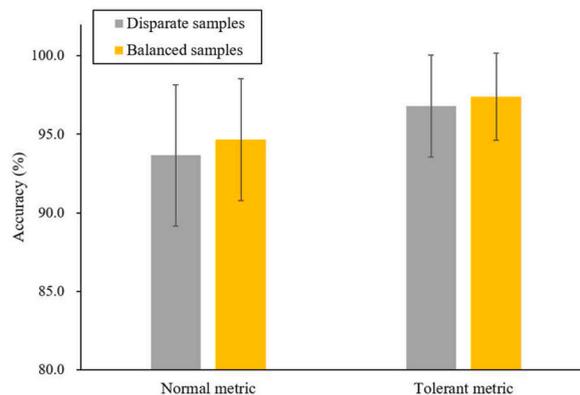


Fig. 13. Histogram of the pattern recognition accuracy of the classifier with two sample proportions under the normal metric and the tolerant metric. The gray legend represents the classification accuracy under Disparate samples, and the orange legend represents the classification accuracy under Balanced samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Mean Average Precision (MAP) refers to the mean precision under different Recalls. The model could draw a PR curve (Precision-Recall curve) for the detection results of each type of object, and then calculate the Average Precision (AP) (the area under the PR curve, the higher the AP value, the better the model performance). MAP is the average of AP values for each category [62].

Loss is obtained from the loss function. The loss function is used to evaluate the degree of deviation between the predicted value and the true value. The better the loss function, the better the performance of the model [63]. The YOLOv4 model uses CIoU Loss as its loss function [56].

After 4000 iterations of the neural network, the MAP of the YOLOv4 model reached 94.2%, and the loss dropped to 1.2656. The

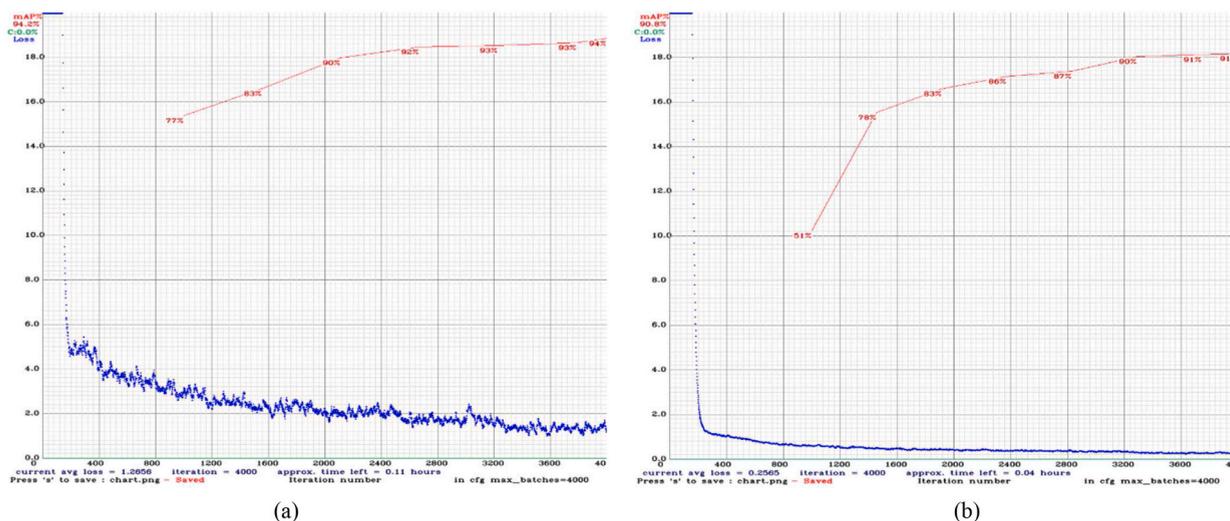


Fig. 14. The training results of YOLOv4 and YOLOv4-tiny models. Figure (a) represents the Loss and MAP changes of the YOLOv4 model over 4000 iterations. Figure (b) represents the Loss and MAP changes of the YOLOv4-tiny model over 4000 iterations.

MAP of the YOLOv4-tiny model reaches 90.8%, and the loss drops to 0.2565. The YOLOv4-tiny model has lower loss and less fluctuation.

We choose the YOLOv4-tiny model for online scene verification, and the real-time detection speed is measured by frame per second, which refers to the number of image frames processed by the object detection program per second [56]. Fig. 15 (a) and Fig. 15 (b) shows the scene where the user manipulates the prosthetic hand to interact with objects, which is displayed by the computer's graphical interface. The green detection box represents the prosthetic hand, and the blue box represents the interactive object. When the user brings the prosthetic hand close to the object, both detection boxes turn red. The average detection speed of the YOLOv4-tiny model in real-time is 25.29fps, and the average confidence level of the prosthetic hand is 96.4%. Fig. 16 (a) and Fig. 16 (b) shows the start-stop of the stimulus interface from the user's first perspective.

3.6. Experiment results of real-life scenarios

The completion time of subjects in each task scenario is shown in Table 6. Except for one subject, the other subjects successfully completed the 4 life-scenario tasks, with an average time-consuming of $39.6 \pm 6.0s$, $24.5 \pm 10.1s$, $50.5 \pm 6.0s$, and $26.2 \pm 11.2s$, respectively. One of the subjects failed both the Pouring water - Drinking water task and the Taking a pen - Writing task. The reason is that the two commands "Hold" and "Put down" were incorrectly recognized during the task, resulting in the drop of the cup and pen.

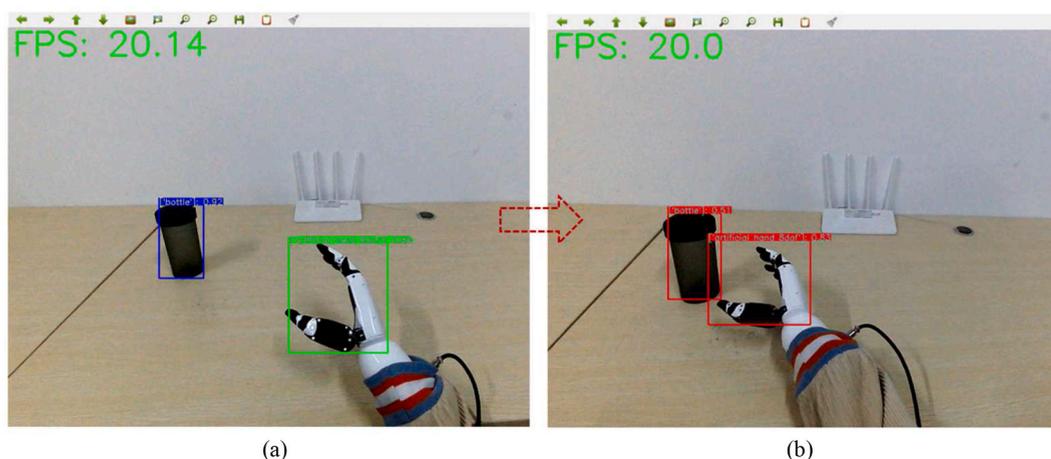


Fig. 15. The computer graphics interface for the scene where the user manipulates the prosthetic hand to interact with the object. The green detection box represents the prosthetic hand, and the blue box represents the interactive object (Fig. 15 (a)). When the user brings the prosthetic hand close to the object, both detection boxes turn red (Fig. 15 (b)). The frame per second is displayed in the upper left corner. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

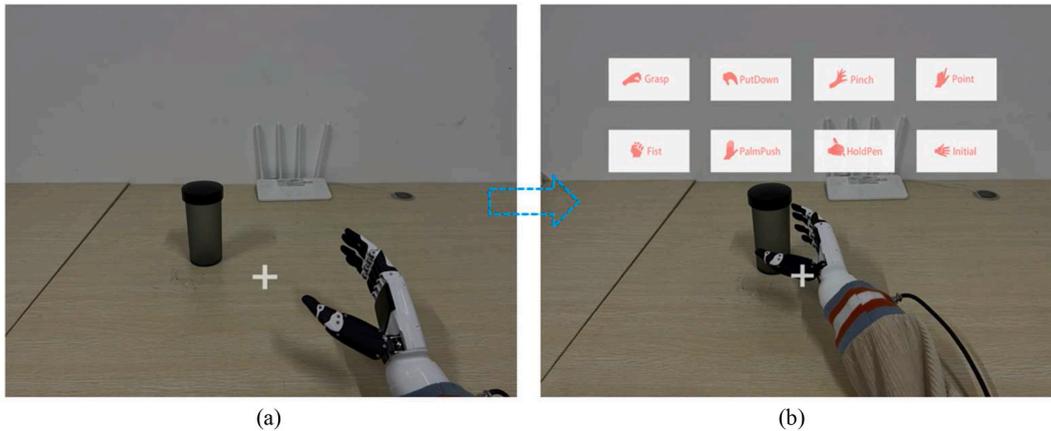


Fig. 16. The start-stop of the stimulus interface from the user's first perspective. When the prosthetic hand approaches the interactive object (the interactive behavior is detected) (Fig. 16 (a)), the stimulus interface is turned on, and the signal detection and control module of the BCI system operates at the same time (Fig. 16 (b)).

Table 6

The task completion time in each scenario.

Number	Pouring water - Drinking water	Opening the door	Taking a pen - Writing	Typing
1	35.1s	16.2s	44.3s	18.6s
2	37.3s	20.1s	50.9s	20.5s
3	46.4s	22.7s	56.3s	23.0s
4	fail	39.1s	fail	42.8s
Average	$39.6 \pm 6.0s$	$24.5 \pm 10.1s$	$50.5 \pm 6.0s$	$26.2 \pm 11.2s$

Since there was no chance of correction in this situation, the experiment was deemed a failure. In summary, the average time spent on each task was within the acceptable range for daily life, the number of misrecognition during the experiment was kept at a low level, and most of them were self-correcting. The 4 scene tasks designed in the experiment cover common hand movements in daily life, and the practical application effects of the AR visual stimulation paradigm, asynchronous EEG decoding algorithm, and interaction strategy described in this paper were tested through experiments. It proved the effectiveness and convenience of the brain-controlled prosthetic hand system developed in this paper.

4. Discussions

Firstly, an ablation experiment analysis of the ECCA-AR method was performed. Secondly, the importance of human-computer interaction experience from the perspective of the practicality of brain-controlled prosthetic hands was discussed. In more detail, it was elaborated from three aspects of AR-SSVEP paradigm enhancement, asynchronous pattern recognition and control, and machine vision assistance. Finally, the experimental details, limitations, and future research directions of this study were discussed.

4.1. Analysis of ablation experiments

The purpose of the ablation experiments is to validate the contributions made by the paradigm improvements and method improvements. Six conditions were set up, namely (1) CCA-AR, (2) FBCCA-AR, (3) TRCA-AR, (4) SSCOR-AR, (5) ECCA-AR and (6) ECCA-LCD. Details are shown in Table 7. Under the unified AR-SSVEP paradigm, comparing the decoding accuracy and ITR between various methods (that is, the first 5 conditions), to verify the contribution of the method improvement to the system. Under the unified ECCA method, comparing the decoding accuracy and ITR between the two paradigms (that is, the latter two operating conditions), in order to verify the contribution of the paradigm improvement to the system. The decoding accuracy and ITR were obtained from the experimental data of all subjects participating in experiment one.

By comparing the first five conditions, it was found that the classification accuracy and ITR of ECCA-AR (proposed in this paper) were the highest when the data analysis time was 2s, indicating that the ECCA method was superior to other methods. It has contributed to the brain-controlled prosthetic hand system in decoding ability and ITR. By comparing the latter two conditions, it was found that the classification accuracy and ITR of ECCA-AR were the highest when the data analysis time was 2s, indicating that the AR-SSVEP paradigm proposed in this paper was superior to the conventional LCD-SSVEP paradigm. From a paradigmatic perspective, EEG's ability to represent user intent was facilitated, and the system's ITR was also improved.

Table 7
The average classification accuracy and ITR under various conditions.

Number	Conditions	Accuracy (%)	ITR (bit/min)
1	CCA-AR	90.90	55.79
2	FBCCA-AR	93.10	58.63
3	TRCA-AR	74.21	35.39
4	SSCOR-AR	60.63	22.75
5	ECCA-AR	94.84	62.15
6	ECCA-LCD	90.23	54.34

4.2. Interaction and application in AR-BCI

Considering the practicality of the BCI-based prosthetic hand, especially as an alternative device for disabled patients, there are generally 4 requirements: (1) Excellent user intent recognition accuracy. (2) Smooth and seamless human-machine interaction process. (3) Broad application generalization ability. (4) Excellent portability and affordable price.

Many published papers investigate various BCI system paradigms [2,64] and recognition algorithms [65,66]. Most of them focus on improving the accuracy of intent recognition, which is a very meaningful research direction. But the interactive experience of the BCI system is equally important. Many rigorous BCI experimental paradigms can indeed achieve good recognition accuracy. However, in practical applications, subjects' mental states, environmental changes, and body shaking are all potential interference factors [67]. In addition, the interaction logic and strategy of the BCI system also play an important role. However, few studies have been carried out in this area.

This paper focuses more on the interaction of the BCI system. As a visually evoked paradigm, SSVEP is a BCI paradigm with excellent generality and accuracy [9]. For the prosthetic hand control system, the LCD-based SSVEP-BCI system has defects in interaction, which cannot consider both visual stimulus and control objects [15]. Therefore, this study uses augmented reality technology to construct an AR visual stimulus paradigm. Due to the importance of asynchronous performance in BCI systems for equipment control, we investigate asynchronous pattern recognition algorithms. In addition, the continuous visual stimulus will inevitably cause the user's disgust, and will also reduce the stability of the BCI system. Therefore, the automatic intelligent start-stop of the BCI system is also worth studying.

4.3. Rationality of the AR visual stimulus paradigm

AR can integrate real environment and virtual information and be perceived by human eyes. As a head-mounted display, OST-AR glasses can convey stimulus information to the human eye just like LCD. In addition, compared with VST-AR, OST-AR glasses bring users a more portable and friendly interactive experience. Most of the existing AR-SSVEP studies are based on VST-AR [68], and the AR-SSVEP system for alternative prosthetic hand control is not involved.

We design the AR visual stimulus paradigm for the application scenario of prosthetic hand control. The results of the feature comparison experiment showed that under the AR paradigm, the average spectral amplitude of the EEG signal at the stimulus frequency was 1.5371 μ V, and the average SNR was 12.3745 dB. Compared with the LCD paradigm, it increased by 17.41% and 3.52% respectively. In this experiment, the feature level of the AR paradigm is higher than that of the LCD paradigm, which is different from previous related papers. We speculate that the reasons may be: (1) The LCD paradigm of this paper uses a 13-inch laptop, rather than the desktop display (above 20 inches) of the related paper. The lack of portability of a desktop computer is contrary to the prosthetic hand system. Compared with the 13-inch LCD, the OST-HMD brings a larger visual perception area to the subjects. And according to the principle of visual imaging, the image is projected to the left eye and the right eye respectively, and the stimulus intensity is stronger. (2) The excellent AR paradigm design brings a stronger sense of immersion to the subjects, improves the overt attention of each stimulus mode, weakens the covert attention from other stimulus modes and environments, and enhances the SSVEP characteristics.

In addition, we designed a decoding comparison between the LCD paradigm and the AR paradigm at different data analysis times (0.3s–3.0s). It can be seen from the results that the two paradigms obtained the best accuracy in the data analysis time of 3s (94.24 \pm 4.72% and 96.71 \pm 3.91%, respectively). With the increase of data analysis time, the classification accuracy of the stimulus mode of the two paradigms increased, and the trend slowed down after 2.0s. ITR first increased and then decreased, and both reached the peak at 1.5s (60.19 \pm 12.34bit/min and 67.47 \pm 14.00bit/min, respectively). The experimental results conform to the common characteristics of previous studies [40,69]. The change of ITR is caused by the mutual restriction of the accuracy rate and the data analysis time. Although the accuracy rate has a positive impact on the ITR, the increase of the data analysis time will reduce the ITR. It can be seen from the histogram that when the data analysis time is less than 0.8s, the performance of the LCD paradigm is better than that of the AR paradigm. But at this time, the accuracy and ITR of both are low, which cannot support the normal operation of the prosthetic hand system. When the data analysis time is longer than 0.8s, the AR paradigm has better performance. Although the ITR is the highest at 1.5s, the accuracy still has a large range of increase. Excessive data analysis time will also reduce the real-time performance of the system.

4.4. Asynchronous pattern recognition logic and algorithm performance analysis

The NC mode indicates that the user conveys the operation intention of “no command output” to the BCI system. This model is indispensable and is one of the main ways to implement asynchronous BCI systems [53]. For example, in a synchronization system, the user needs to manipulate the prosthetic hand to hold the water cup for 10 s, during which the BCI system will continue to perform EEG recognition. If there are only 8 motion control modes, the user needs to keep staring at the Grasp stimulus. If it is mistakenly recognized as a command such as put down or a palm push, the water glass will fall. In the asynchronous system, when the BCI system recognizes the current EEG signal as NC mode, there will be no command output. In published studies, a stimulus mode was assigned to the NC mode. This approach does not fundamentally improve the level of interaction. The user still needs to keep looking at the stimulus image representing the NC mode to output the NC state command. Especially for the application scenario of prosthetic hand control, the user is required to keep looking at the prosthetic hand to interact with the environment, rather than staring at the stimulus image all the time.

Covert attention often exists as a distractor in SSVEP-BCI based on overt attention. This inspired us to implement asynchronous systems. In our study, an asynchronous algorithm can be constructed by the effect of covert attention of stimulus images on EEG in “Center mode”. The Center mode just allows users to directly gaze at the environment and control objects, which is very consistent with the interaction logic of the BCI system.

From the experimental results, under the data analysis time of 2.0s, the Center-ECCA-SVM asynchronous pattern classifier trained on balanced samples obtained better accuracy on multiple subjects (normal metric: $94.66 \pm 3.87\%$, tolerant metric: $97.40 \pm 2.78\%$). Among them, the accuracy rate of subject S03 reached 100% under the tolerant metric. This algorithm can distinguish the user’s IC and NC states in the AR paradigm under the premise of ensuring smooth interaction. Using the sliding time window, the user can actively switch the control state of the prosthetic hand under the continuous visual stimulus. For the better performance of the classifier under balanced samples, we speculate that the reason is that: IC status can be subdivided into 8 stimulus modes. The Center mode has no subdivision, but due to changes in environmental information and the movement of the user’s line of sight, the triggering factors of the signal are more complex. The more training samples in the Center mode, the more diverse NC mode features can be learned by the classifier. Additionally, we speculate that as the Center mode samples increase further, there may be more samples of stimulus mode that are misclassified as Center mode. The accuracy under the tolerant metric is improved, while the accuracy under the normal metric is decreased. This will be verified in future studies.

4.5. Enhancement of interactive experience by system intelligent start-stop

For the prosthetic hand system: (1) Wearing AR glasses for a long time and receiving visual stimuli will confuse users. (2) The asynchronous pattern recognition algorithm cannot be 100% accurate. The intelligent system switch can help users wear AR glasses for a long time and enhance the user interaction experience. In our research, based on the YOLOv4 algorithm, the object detection and behavior judgment of prosthetic hands and interactive objects was realized, which is used as the basis for intelligent switches.

It can be seen from the training results that after 4000 iterations, both the YOLOv4 model and the YOLOv4-tiny model tend to be stable. The MAP of the YOLOv4-tiny model is 90.8%, which is 3.4% lower than the YOLOv4 model. However, the loss of the YOLOv4-tiny model is lower, and the fluctuation is smaller, which may be because the YOLOv4-tiny model is shallower and has a smaller number of parameters, so the model is less difficult to fit. The YOLOv4-tiny model tends to be stable after about 3200 iterations, and the training time of 4000 iterations is reduced by 5 times compared with the YOLOv4 model, so it has excellent CPU usage and recognition speed, as well as good detection performance.

During the verification process of the online scene, the average detection speed of the YOLOv4-tiny model in real-time reached 25.29fps, and the average confidence of the prosthetic hand reached 96.4%. This is sufficient as a command basis for the prosthetic hand system to start-stop. In addition, we can further limit the CPU occupancy in the object detection program and use the performance more for the EEG signal decoding process. In our study, the standby state of the BCI system can be woken up at any time, which can be understood as a system mode rather than a system shutdown at the power level.

4.6. Details in the experiment

We recruited 12 subjects to participate in both experiments, and none of the subjects had done similar visual BCI experiments. In this study, we put more emphasis on the interaction and practical application of the BCI system, so we did not conduct the experimental deployment as rigorously as in previous studies [44]. We will make the slightly depressing BCI experiment a little bit more routine, and here we do the following: (1) Considering that when the user wears the prosthetic hand to complete daily activities, he cannot concentrate as highly as the experimental process. Therefore, we reminded the subjects to allow slight body movements and blink at will during the experiment. (2) The real-life environment cannot have less interference than the EEG experimental environment, so we asked the subjects whether they needed to listen to music. Five of the 12 subjects accepted the offer. We did not analyze and study this, and it is not clear whether music influenced the experimental results. But we think this is more towards situations that are common in everyday life. (3) We did not ask the subjects to refrain from speaking. During the experiment, 3 subjects had short and low-pitched communication with the experimental supervisor, mostly about questions and suggestions about the experimental paradigm. These experimental data are still included in the experimental data analysis. (4) One of the subjects reported that the SSVEP stimulator made her feel a certain degree of visual fatigue and discomfort. This may be one of the reasons why she failed both the Pouring water - Drinking water task and Taking a pen - Writing task.

4.7. Limitations and future studies

There are still some limitations worthy of further study. Twelve subjects were recruited for the experimental study, all healthy subjects. Intuitively, SSVEP is a visually evoked EEG signal, and experiments on healthy subjects have good guiding significance [70]. However, online experimental verification of disabled patients and feasibility study of system wearing are also necessary, and we will carry out this study in the future. At the same time, we will continue to increase the number of subjects and expand the research, such as the study of individual differences and gender differences.

In this paper, AR glasses are used as the carrier of visual stimulus, and the portable computer is used as the host computer to complete data processing and control. This does not achieve the most lightweight system structure. There are already AR glasses equipped with computing chips and operating systems, such as the Microsoft HoloLens series [71]. However, due to the limitation of computing power, such AR glasses cannot fully carry the BCI system in this paper. In future research, a higher level of portability can be achieved by improving the computing power of AR glasses chips or optimizing programs.

Studies have shown that the high-frequency SSVEP paradigm can alleviate visual fatigue to a certain extent and can expand more instruction sets for the BCI system [72]. However, the high-frequency stimulus has higher requirements on the refresh rate of the display device. And because the EEG signal characteristics induced by the high-frequency stimulus are relatively weak, the signal decoding needs to be improved [73]. In the future, more attention can be paid to the research of high-frequency SSVEP to further improve the comprehensive performance of the BCI system. In addition, the layout of stimulus pictures in the AR-SSVEP paradigm [18], the joint modulation scheme of the stimulator [28], and the high-precision decoding method based on spatial filters [74] are the key directions of research, and further exploration is required.

5. Conclusion

The AR-based visual stimulus paradigm, asynchronous pattern recognition method, and system intelligent start-stop interaction method have been studied in this paper, respectively. In the AR paradigm, the average SSVEP feature amplitude of multiple subjects was increased by 17.41% compared with the LCD paradigm, and the average stimulus mode recognition accuracy was increased by 2.62%. Under the data analysis time of 2s, the Center-ECCA-SVM classifier achieves excellent asynchronous decoding performance (Normal metric: $94.66 \pm 3.87\%$, Tolerant metric: $97.40 \pm 2.78\%$). The YOLOv4-tiny model also brings excellent recognition speed and accuracy to the intelligent start-stop of the system. Moreover, the brain-controlled prosthetic hand helped the subjects to complete 4 kinds of daily life tasks in the real scene, and the time-consuming were all within an acceptable range, which verified the effectiveness and practicability of the system. In summary, this study builds a smooth prosthetic hand manipulation and interaction logic and provides support for BCI areas for alternative prosthetic control, and movement disorder rehabilitation programs.

Funding

This work is supported in part by grants from the Science and Technology Supporting Xinjiang Project (No. 2020E0259) and partially supported by the National Key Research and Development Program of China (No. 2017YFB1300303).

Ethics statement

This studies involving human participants were reviewed and approved by the institutional review board of Xi'an Jiaotong University (No. 20211452), and all experiments were conducted in accordance with the declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Data availability statement

Data will be made available on request.

CRedit authorship contribution statement

Xiaodong Zhang: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. **Teng Zhang:** Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Yongyu Jiang:** Data curation, Software, Writing – original draft. **Weiming Zhang:** Validation. **Zhufeng Lu:** Software. **Yu Wang:** Validation. **Qing Tao:** Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J.R. Wolpaw, et al., Brain-computer interface technology: a review of the first international meeting, *Ieee T Rehabil Eng* 8 (2) (Jun 2000) 164–173, <https://doi.org/10.1109/Tre.2000.847807> (in English).
- [2] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, *Sensors-Basel* 12 (2) (Feb 2012) 1211–1279, <https://doi.org/10.3390/s120201211> (in English).
- [3] A. Constantine, V. Asanza, F.R. Loayza, E. Pelaez, D. Peluffo-Ordonez, BCI system using a novel processing Technique based on electrodes selection for hand Prosthesis control, *IFAC-PapersOnLine* 54 (15) (2021) 364–369, <https://doi.org/10.1016/j.ifacol.2021.10.283> (in English).
- [4] G. Pfurtscheller, et al., "The hybrid BCI", *Front Neurosci-Switz* 4 (2010), 4210.3389/fnpro.2010.00003.
- [5] G. Lange, C.Y. Low, K. Johar, F.A. Hanapiyah, F. Kamaruzaman, Classification of electroencephalogram data from hand Grasp and Release movements for BCI controlled Prosthesis, *Proc Tech* 26 (2016) 374–381, <https://doi.org/10.1016/j.protcy.2016.08.048> (in English).
- [6] R. Alazrai, H. Alwanni, M.I. Daoud, EEG-based BCI system for decoding finger movements within the same hand, *Neurosci. Lett.* 698 (Apr 17 2019) 113–120, <https://doi.org/10.1016/j.neulet.2018.12.045> (in English).
- [7] A.A. Frolov, et al., Post-stroke rehabilitation training with a Motor-Imagery-based brain-computer interface (BCI)-Controlled hand exoskeleton: a Randomized controlled Multicenter trial, *Front Neurosci-Switz* 11 (Jul 20 2017), 40010.3389/fnins.2017.00400.
- [8] E. Buch, et al., Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke, *Stroke* 39 (3) (Mar 2008) 910–917, <https://doi.org/10.1161/Strokeaha.107.505313> (in English).
- [9] S. Amiri, A. Rabbi, L. Azinfar, R. Fazel-Rezai, A Review of P300, SSVEP, and Hybrid P300/SSVEP Brain- Computer Interface Systems, 2013.
- [10] M. Labecki, R. Kus, A. Brzozowska, T. Stacewicz, B.S. Bhattacharya, P. Suffczynski, Nonlinear origin of SSVEP spectra—a combined experimental and modeling study, *Front. Comput. Neurosci.* 10 (2016) 129.
- [11] F.B. Vialatte, M. Maurice, J. Dauwels, A. Cichocki, Steady-state visually evoked potentials: focus on essential paradigms and future perspectives, *Prog Neurobiol* 90 (4) (Apr 2010) 418–438 [Online]. Available: <Go to ISI>://WOS:000276770000002.
- [12] R. Zerafa, T. Camilleri, O. Falzon, K.P. Camilleri, To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs, *J. Neural. Eng.* 15 (5) (2018) 051001.
- [13] Y. Zhang, S.Q. Xie, H. Wang, Z. Zhang, Data analytics in steady-state visual evoked potential-based brain-computer interface: a review, *IEEE Sensor. J.* 21 (2) (2020) 1124–1138.
- [14] X. Duart, E. Quiles, F. Suay, N. Chio, E. García, F. Morant, Evaluating the effect of stimuli color and frequency on SSVEP, *Sensors-Basel* 21 (1) (2020) 117.
- [15] J. Mu, D.B. Grayden, Y. Tan, D. Oetomo, Comparison of steady-state visual evoked potential (SSVEP) with LCD vs. LED stimulation, *Ieee Eng Med Bio* (2020) 2946–2949 [Online]. Available: <Go to ISI>://WOS:000621592203073.
- [16] M. Nakanishi, Y.J. Wang, X.G. Chen, Y.T. Wang, X.R. Gao, T.P. Jung, Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis, *Ieee T Bio-Med Eng* 65 (1) (Jan 2018) 104–112, <https://doi.org/10.1109/Tbme.2017.2694818>.
- [17] X.G. Chen, Y.J. Wang, M. Nakanishi, X.R. Gao, T.P. Jung, S.K. Gao, High-speed spelling with a noninvasive brain-computer interface, *P Natl Acad Sci USA* 112 (44) (Nov 3 2015) E6058–E6067, <https://doi.org/10.1073/pnas.1508080112>.
- [18] X.C. Zhao, C.Y. Liu, Z.X. Xu, L.P. Zhang, R. Zhang, SSVEP stimulus layout effect on accuracy of brain-computer interfaces in augmented reality glasses, *IEEE Access* 8 (2020) 5990–5998, <https://doi.org/10.1109/Access.2019.2963442>.
- [19] M.K. Bekele, R. Pierdicca, E. Frontoni, E.S. Malinverni, J. Gain, A survey of augmented, virtual, and mixed reality for cultural heritage, *Acem J Comput Cult He* 11 (2) (Jun 2018), 710.1145/3145534.
- [20] K. Takano, N. Hata, K. Kansaku, Towards intelligent environments: an augmented reality-brain-machine interface operated with a see-through head-mount display, *Front. Neurosci.* 5 (2011) 60.
- [21] M. Wang, R. Li, R. Zhang, G. Li, D. Zhang, A wearable SSVEP-based BCI system for quadcopter control using head-mounted device, *IEEE Access* 6 (2018) 26789–26798.
- [22] H. Si-Mohammed, et al., Towards BCI-based interfaces for augmented reality: feasibility, design and evaluation, *Ieee T Vis Comput Gr* 26 (3) (Mar 1 2020) 1608–1621, <https://doi.org/10.1109/Tvcg.2018.2873737>.
- [23] J. Faller, et al., A Feasibility Study on SSVEP-Based Interaction with Motivating and Immersive Virtual and Augmented Reality, 2017.
- [24] R.T. Azuma, A survey of augmented reality, *Presence-Virtual Aug* 6 (4) (Aug 1997) 355–385, <https://doi.org/10.1162/pres.1997.6.4.355>.
- [25] J.Y. Hong, et al., See-through optical combiner for augmented reality head-mounted display: index-matched anisotropic crystal lens, *Sci Rep-Uk* 7 (Jun 5 2017), 275310.1038/s41598-017-03117-w.
- [26] J. Xie, G.H. Xu, J. Wang, F. Zhang, Y.Z. Zhang, Steady-state motion visual evoked potentials produced by oscillating Newton's rings: implications for brain-computer interfaces, *PLoS One* 7 (6) (Jun 19 2012) e3970710, 1371/journal.pone.0039707.
- [27] M. Nakanishi, Y.J. Wang, Y.T. Wang, Y. Mitsukura, T.P. Jung, A high-speed brain speller using steady-state visual evoked potentials, *Int. J. Neural Syst.* 24 (6) (Sep 2014), 145001910.1142/S0129065714500191.
- [28] Z.Y. Wang, H.L. Hu, X.F. Chen, T. Zhou, T.H. Xu, A novel SSVEP-based brain-computer interface using joint frequency and space modulation, *Ieee Conf Comput* (2020) 906–911 [Online]. Available: <Go to ISI>://WOS:000593830400150.
- [29] C.H. Han, K.R. Muller, H.J. Hwang, Brain-switches for asynchronous brain-computer interfaces: a systematic review, *Mar, Electronics-Switz* 9 (3) (2020). .3390/electronics9030422.
- [30] A A Nooh, J Yunus, S M Daud, A review of asynchronous electroencephalogram-based brain computer interface systems, *International Conference on Biomedical Engineering and Technology IPCBEE[C]* (2011).
- [31] X. Zhang, et al., A convolutional neural network for the detection of asynchronous steady state motion visual evoked potential, *Ieee T Neur Sys Reh* 27 (6) (Jun 2019) 1303–1311, <https://doi.org/10.1109/Tnsre.2019.2914904>.
- [32] S. Parini, L. Maggi, A.C. Turconi, G. Andreoni, A robust and self-paced BCI system based on a four class SSVEP paradigm: algorithms and protocols for a high-transfer-rate direct brain communication, *Comput. Intell. Neurosci.* (2009) 864564, <https://doi.org/10.1155/2009/864564>.
- [33] B. Xia, X. Li, H. Xie, W.L. Yang, J. Li, L.H. He, Asynchronous brain-computer interface based on steady-state visual-evoked potential, *Cogn Comput* 5 (2) (Jun 2013) 243–251, <https://doi.org/10.1007/s12559-013-9202-7> (in English).
- [34] K.S. Hong, M.J. Khan, Hybrid brain-computer interface techniques for improved classification accuracy and increased number of commands: a review, *Front Neurobotics* 11 (Jul 24 2017), 3510.3389/fnbot.2017.00035.
- [35] R. Abiri, S. Borhani, E.W. Sellers, Y. Jiang, X.P. Zhao, "A comprehensive review of EEG-based brain-computer interface paradigms," *J. Neural. Eng.* 16 (1) (Feb 2019). ARTN01100110.1088/1741-2552/aaf12e.
- [36] A.R. Zangene, A. Abbasi, K. Nazarpour, Estimation of lower limb kinematics during squat task in different loading using sEMG activity and deep recurrent neural networks, *Sensors-Basel* 21 (23) (Dec 2021). ARTN 777310.3390/s21237773.
- [37] Z.Q. Zhao, P. Zheng, S.T. Xu, X.D. Wu, Object detection with deep learning: a review, *Ieee T Neur Net Lear* 30 (11) (Nov 2019) 3212–3232, <https://doi.org/10.1109/Tnnls.2018.2876865> (in English).
- [38] J. Pan, X. Gao, F. Duan, Z. Yan, S. Gao, Enhancing the classification accuracy of steady-state visual evoked potential-based brain-computer interfaces using phase constrained canonical correlation analysis, *J. Neural. Eng.* 8 (3) (Jun 2011) 036027, <https://doi.org/10.1088/1741-2560/8/3/036027>.
- [39] Q.G. Wei, S. Zhu, Y.J. Wang, X.R. Gao, H. Guo, X. Wu, A training data-driven canonical correlation analysis algorithm for designing spatial filters to enhance performance of SSVEP-based BCIs, *Int. J. Neural Syst.* 30 (5) (May 2020). Artn205002010.1142/S0129065720500203.
- [40] X.G. Chen, Y.J. Wang, S.K. Gao, T.P. Jung, X.R. Gao, Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface, *J. Neural. Eng.* 12 (4) (Aug 2015). Artn04600810.1088/1741-2560/12/4/046008.
- [41] N.S. Kwak, K.R. Muller, S.W. Lee, A convolutional neural network for steady state visual evoked potential classification under ambulatory environment, *PLoS One* 12 (2) (Feb 22 2017) e017257810, 1371/journal.pone.0172578.

- [42] N. Waytowich, et al., Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials, *J. Neural. Eng.* 15 (6) (Dec 2018). ARTN06603110.1088/1741-2552/aae5d8.
- [43] K. Suefusa, T. Tanaka, Asynchronous brain-computer interfacing based on mixed-coded visual stimuli, *IEEE Trans. Biomed. Eng.* 65 (9) (Sep 2018) 2119–2129, <https://doi.org/10.1109/TBME.2017.2785412>.
- [44] Y.F. Ke, P.X. Liu, X.W. An, X.Z. Song, D. Ming, An online SSVEP-BCI system in an optical see-through augmented reality environment, *J. Neural. Eng.* 17 (1) (Feb 2020). ARTN01606610.1088/1741-2552/ab4dc6.
- [45] L.L. Chen, et al., Adaptive asynchronous control system of robotic arm based on augmented reality-assisted brain-computer interface, *J. Neural. Eng.* 18 (6) (Dec 2021). ARTN06600510.1088/1741-2552/ac3044.
- [46] X.G. Chen, X.S. Huang, Y.J. Wang, X.R. Gao, Combination of augmented reality based brain-computer interface and computer vision for high-level control of a robotic arm, *Ieee T Neur Sys Reh* 28 (12) (Dec 2020) 3140–3147, <https://doi.org/10.1109/Tnsre.2020.3038209> (in English).
- [47] X. Chen, Z. Chen, S. Gao, X. Gao, A high-itr ssvep-based bci speller, *Brain-Computer Interfaces* 1 (3–4) (2014) 181–191.
- [48] R.W. Homan, J. Herman, P. Purdy, Cerebral location of international 10–20 system electrode placement, *Electroencephalogr. Clin. Neurophysiol.* 66 (4) (1987) 376–382, [https://doi.org/10.1016/0013-4694\(87\)90206-9](https://doi.org/10.1016/0013-4694(87)90206-9).
- [49] A.-B.R. Suleiman, T.A.-H. Fatehi, Features Extraction Techniques of EEG Signal for BCI Applications, Faculty of Computer and Information Engineering Department College of Electronics Engineering, University of Mosul, Iraq, 2007.
- [50] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, T.-P. Jung, A high-speed brain speller using steady-state visual evoked potentials, *Int. J. Neural Syst.* 24 (6) (2014) 1450019.
- [51] Y. Wang, M. Nakanishi, Y.-T. Wang, T.-P. Jung, Enhancing detection of steady-state visual evoked potentials using individual training data, 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2014) 3037–3040, 2014: Ieee.
- [52] G. Bin, X. Gao, Z. Yan, B. Hong, S. Gao, An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method, *J. Neural. Eng.* 6 (4) (2009) 046002.
- [53] M.-C. Schaeffer, T. Aksenova, Switching Markov decoders for asynchronous trajectory reconstruction from ECoG signals in monkeys for BCI applications, *J. Physiol. Paris* 110 (4) (2016) 348–360.
- [54] M.M. Müller, P. Malinowski, T. Gruber, S. Hillyard, Sustained division of the attentional spotlight, *Nature* 424 (6946) (2003) 309–312.
- [55] P. Burman, A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods, *Biometrika* 76 (3) (1989) 503–514.
- [56] A. Bochkovskiy, W. Chien-Yao, M.L. Hong-Yuan, YOLOv4: Optimal Speed and Accuracy of Object Detection[Z]. Cornell University Library, Ithaca, arXiv.org, 2020.
- [57] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, *Journal of big data* 6 (1) (2019) 1–48.
- [58] A. Gulhane, P.L. Paikrao, D. Chaudhari, A review of image data clustering techniques, *Int. J. Soft Comput. Eng.* 2 (1) (2012) 212–215.
- [59] Tzatalin. Labelling. Git code. <https://github.com/tzatalin/labelling>.
- [60] Q. Sun, M. Chen, L. Zhang, C. Li, W. Kang, Similarity-constrained task-related component analysis for enhancing SSVEP detection, *J. Neural. Eng.* 18 (4) (Jun 4 2021), <https://doi.org/10.1088/1741-2552/abfdfa>.
- [61] G.R. Kiran Kumar, M. Ramasubba Reddy, Designing a sum of squared correlations framework for enhancing SSVEP-based BCIs, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (10) (Oct 2019) 2044–2050, <https://doi.org/10.1109/TNSRE.2019.2941349>.
- [62] R. Padilla, S.L. Netto, E.A. Da Silva, A survey on performance metrics for object-detection algorithms, in: 2020 International Conference on Systems, Signals and Image Processing, IWSSIP), 2020, pp. 237–242. IEEE.
- [63] K. Janocha, W.M. Czarnecki, On Loss Functions for Deep Neural Networks in Classification, 2017 *arXiv preprint arXiv:1702.05659*.
- [64] M. Hamed, S.H. Salleh, A.M. Noor, Electroencephalographic motor imagery brain connectivity analysis for BCI: a review, *Neural Comput.* 28 (6) (Jun 2016) 999–1041, https://doi.org/10.1162/NECO_a_00838 (in English).
- [65] A. Ravi, N.H. Beni, J. Manuel, N. Jiang, "Comparing user-dependent and user-independent training of CNN for SSVEP BCI", *J. Neural. Eng.* 17 (2) (Apr 2020). ARTN02602810.1088/1741-2552/ab6a67.
- [66] F. Fahimi, Z. Zhang, W.B. Goh, T.S. Lee, K.K. Ang, C.T. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI", *J. Neural. Eng.* 16 (2) (Apr 2019). ARTN02600710.1088/1741-2552/aaf3f6.
- [67] M. Billinger, et al., Is it significant? Guidelines for reporting BCI performance, in: *Towards Practical Brain-Computer Interfaces*, Springer, 2012, pp. 333–354.
- [68] H. Si-Mohammed, F.A. Sanz, G. Casiez, N. Roussel, A. Lécuyer, Brain-computer interfaces and augmented reality: a state of the art, in: *Graz Brain-Computer Interface Conference*, 2017.
- [69] Y. Zhang, et al., Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIs, in: *International Conference on Neural Information Processing*, Springer, 2011, pp. 287–295.
- [70] I. Volosyak, H. Cecotti, D. Valbuena, A. Graser, Evaluation of the Bremen SSVEP based BCI in real world conditions, *IEEE International Conference on Rehabilitation Robotics* (2009) 322–331, 2009: IEEE.
- [71] S. Park, S. Bokijonov, Y. Choi, Review of microsoft hololens applications over the past five years, *Appl. Sci.* 11 (16) (2021) 7259.
- [72] M.H. Chang, H.J. Baek, S.M. Lee, K.S. Park, An amplitude-modulated visual stimulation for reducing eye fatigue in SSVEP-based brain-computer interfaces, *Clin. Neurophysiol.* 125 (7) (2014) 1380–1391.
- [73] I. Volosyak, D. Valbuena, T. Luth, T. Malechka, A. Graser, BCI demographics II: how many (and what kinds of) people can use a high-frequency SSVEP BCI? *Ieee T Neur Sys Reh* 19 (3) (2011) 232–239.
- [74] C.M. Wong, B. Wang, Z. Wang, K.F. Lao, A. Rosa, F. Wan, Spatial filtering in SSVEP-based BCIs: unified framework and new improvements, *IEEE Trans. Biomed. Eng.* 67 (11) (Nov 2020) 3057–3072, <https://doi.org/10.1109/TBME.2020.2975552>.