Shared Three-Dimensional Robotic Arm Control Based on Asynchronous BCI and Computer Vision

Yajun Zhou, Tianyou Yu, *Member, IEEE*, Wei Gao, Weichen Huang, Zilin Lu, Qiyun Huang, and Yuanqing Li, *Fellow, IEEE*

***Abstract*— *Objective:* A brain-computer interface (BCI) can be used to translate neuronal activity into commands to control external devices. However, using noninvasive BCI to control a robotic arm for movements in three-dimensional (3D) environments and accomplish complicated daily tasks, such as grasping and drinking, remains a challenge. *Approach:* In this study, a shared robotic arm control system based on hybrid asynchronous BCI and computer vision was presented. The BCI model, which combines steady-state visual evoked potentials (SSVEPs) and blink-related electrooculography (EOG) signals, allows users to freely choose from fifteen commands in an asynchronous mode corresponding to robot actions in a 3D workspace and reach targets with a wide movement range, while computer vision can identify objects and assist a robotic arm in completing more precise tasks, such as grasping a target automatically. *Results:* Ten subjects participated in the experiments and achieved an average accuracy of more than 92% and a high trajectory efficiency for robot movement. All subjects were able to perform the reach-grasp-drink tasks successfully using the proposed shared control method, with fewer error commands and shorter completion time than with direct BCI control. *Significance:* Our results demonstrated the feasibility and efficiency of generating practical multidimensional control of an intuitive robotic arm by merging hybrid asynchronous BCI and computer vision-based recognition.**

***Index Terms*— Asynchronous brain-computer interface (BCI), electroencephalography (EEG), electrooculography (EOG), robotic arm, computer vision.**

Manuscript received 22 March 2023; revised 13 July 2023; accepted 21 July 2023. Date of publication 27 July 2023; date of current version 7 August 2023. This work was supported in part by the Technology Innovation 2030-Major Project 2022ZD0208900; in part by the Key Research and Development Program of Guangdong Province, China, under Grant 2018B030339001; and in part by the Key Realm Research and Development Program of Guangzhou, China, under Grant 202007030007. *(Corresponding author: Yuanqing Li.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Sichuan Provincial Rehabilitation Hospital, China, under Approval No: CKLL-2018008.

Yajun Zhou, Weichen Huang, Zilin Lu, and Qiyun Huang are with the Research Center for Brain-Computer Interface, Pazhou Lab, Guangzhou 510330, China, and also with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China.

Tianyou Yu and Yuanqing Li are with the Research Center for Brain-Computer Interface, Pazhou Lab, Guangzhou 510330, China, also with the School of Automation Science and Engineering, South China University of Technology, Guangzhou, 510640, China, and also with South China Brain-Computer Interface Technology Co., Ltd, Guangzhou 510320, China (e-mail: auyqli@scut.edu.cn).

Wei Gao is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China, also with the School of Software, South China Normal University, Guangzhou 510631, China, and also with the Pazhou Laboratory, Research Center for Brain-Computer Interface, Guangzhou 510330, China.

Digital Object Identifier 10.1109/TNSRE.2023.3299350

I. Introduction

INDIVIDUALS who suffer from severe neuromuscular disorders or motor system impairments, such as amyotrophic lateral sclerosis (ALS), brainstem stroke, or spinal cord injury (SCI), typically experience a loss of upper limb function, including the ability to perform daily activities such as reaching and grasping objects. However, most individuals with these conditions can still utilize their brain to control external devices by producing function-related neural processes [1], [2], [3], [4], [5], [6]. Robots, which are generally developed as neuroprosthetic devices, have the potential to enhance the efficiency of environmental interaction for these individuals [7], [8], [9]. Nonetheless, coordinating the high number of degrees of freedom (DOFs) to achieve anthropomorphic control remains a significant challenge for these severely paralyzed patients.

Brain-computer interface (BCI) technology, which utilizes several different input signals, has been developed in recent years and may serve as the foundation for a new generation of robotic control [10]. In 1999, Chapin et al. [11] first demonstrated that a robotic manipulator could be controlled directly by cortical neuron ensembles. Since then, numerous types of studies have successfully converted animal or human motor cortical activities into intentional movements of a mechanized arm [12], [13], [14], [15]. Neuronal activities were recorded from specific parts of cortical areas, such as the primary motor (M1) cortex, ventrolateral (VL) thalamus, or posterior parietal cortex (PPC). Although these invasive strategies provide neural signals with high quality, they carry risks associated with invasive surgical procedures. Therefore, noninvasive BCI through recording tools, such as electroencephalography (EEG) or magnetoencephalography (MEG), for visual arm or prosthetic control remains a major research topic [16], [17], [18] because of the short preparation time and its ability to be applied without surgery.

EEG-based BCI models have been introduced to the intensional manipulation of several kinds of robots, such as

quadcopters with motor imagery (MI) [16], small humanoid robots with P300 [19] or MI [20], and mobile robots [21], [22] or wheelchairs with steady-state visual evoked potentials (SSVEPs) [23], [24]. However, to our knowledge, few studies have focused on noninvasive BCI-based robotic arm control, which expanded the traditional robotic movement from one-dimensional space to two- or even three-dimensional (3D) space. The desired multidimensional movement for robot arms can be generated by using a step-by-step strategy. For example, Chen et al. [25] proposed an SSVEP-based robotic arm control system in which the 3D trajectories for a task were decomposed into a discrete sequence with each command corresponding to a movement along one specific dimension under the Cartesian coordinate system. The increase in control dimensions may aggravate the users' operational burden if they have to switch frequently between intensional commands in different dimensions to keep pace with the ongoing movement of robotic arms. In addition, motor imagery-based BCIs may hardly satisfy the high number of DOFs needed to handle complicated scenarios with robotic actions such as grasping and drinking [26]. Therefore, multiple commands with quicker and more accurate manipulation should be provided by the BCI-based robotic system.

Over the past two decades, SSVEPs have been attracting increasing interest among researchers due to their high communication rates and low user training requirements [27], [28], [29]. SSVEPs can be evoked when a subject focuses on a visual stimulus that flickers steadily on a screen at a specific frequency [27], [30]. In typical SSVEP-based BCI, the flicker frequency is unique for each object, and the target object can be recognized by applying efficient methods, such as the canonical correlation analysis (CCA)-based algorithm [31]. For example, Nakanishi et al. [32] achieved averaged information transfer rates (ITRs) of 325.33 bits min^{-1} via an extended CCA-based method [32]. However, for an EEG-based BCI system, working in asynchronous mode is more necessary in actual applications because subjects can decide the onset time of sending a command by themselves. A hybrid strategy, such as combining EEG signals with electrooculography (EOG) signals, could be a possible solution to address this problem because the latter is easier to detect and less fluctuant than the former [33], [34], [35]. More importantly, eye movements (e.g., blinking, changing in gaze direction, and fixating) are usually performed intentionally and naturally.

In order to enhance the efficacy of BCI-based robotic arm control, we proposed a novel control strategy that focused on the integration of hybrid BCI and computer vision. Specifically, the hybrid BCI model detected users' manipulation intentions in 3D space in an unknown environment and worked in an asynchronous mode in which both recorded SSVEP and EOG signals contributed to the final decision. The asynchronous control approach allows subjects to move a robotic arm at their own pace. When objects enter the camera's field of view, the computer vision module from the Intel RealSense D435 camera helps the robotic arm make precise and efficient movements for high-level tasks, such as grasping and drinking. Both offline and online experiments were conducted to demonstrate the practicability of the proposed system.

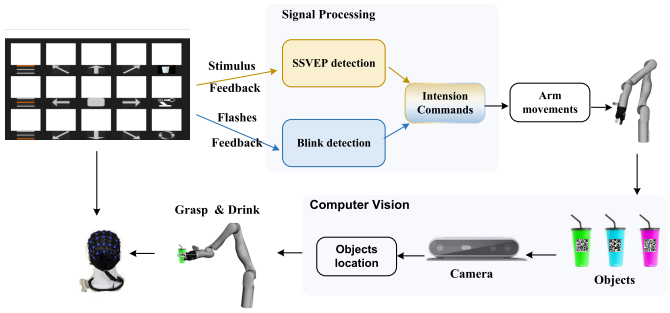


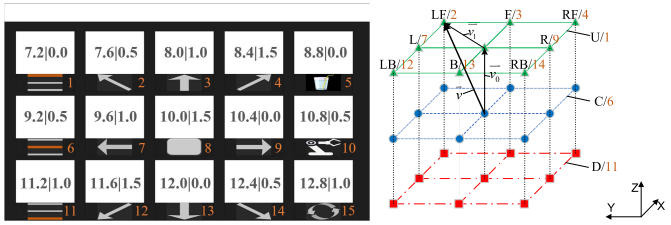
Fig. 1. Schematic of the BCI robot control system.

II. Materials and Methods

A. System Description

As depicted in Fig 1, the proposed hybrid asynchronous BCI-based robot control system is comprised of three subsystems: a hybrid asynchronous BCI model, a robotic arm, and a computer vision model. The program of the hybrid asynchronous BCI system was developed on one computer, while the robotic arm subsystem and the computer vision subsystem were deployed on the other computer. The two computers worked independently of each other, and mutual communication between them was accomplished over the TCP/IP protocol.

1) Hybrid Asynchronous BCI: A fifteen-target hybrid BCI subsystem combining SSVEPs and EOGs was developed and run in an asynchronous mode to select one of these corresponding commands to be acted upon by the robotic arm. The main user interface was a 3 × 5 stimulation matrix containing 15 buttons, each corresponding to a command. As depicted in Fig 2(a), a periodic sinusoidal stimulation paradigm was employed, in which each button flickered with a specific frequency and phase to evoke the SSVEPs. Therefore, the amplitude of the stimuli was modulated accordingly [29], [36]. We also presented a cue for eye blinks to evoke EOG signals by highlighting the buttons with their size transitorily reduced in a random sequence, which maintained ongoing SSVEP stimulation synchronously. To connect the BCI commands with the robotic arm's 3D movement, we designed a new strategy called 3D vector synthesis, which is described below. The three buttons in the first column (marked as "1", "6" and "11") indicate the upward, centering, and downward movements in the vertical direction of the end effector of the robotic arm, respectively (abbreviated as "U", "C", and "D", respectively). Another eight buttons with different arrows (marked as "2", "3", "4", "7", "9", "12", "13", "14") allowed the robot to move toward eight corresponding directions along the horizontal plane (namely, "left front", "front", "right front", "left", "right", "left back", "back" and "right back", which were abbreviated as "LF", "F", "RF", "L", "R", "LB", "B", "RB", respectively). To control the robotic arm movement in 3D space, a participant should first determine its movement trend (e.g., "U") in the vertical direction and then choose one of the eight movements in the horizontal plane. Therefore, the command direction in 3D space is the synthetic vector of the motion vector in the vertical direction and the motion vector



(a) (b)

Fig. 2. (a) The main user interface of the proposed system, with the numbers next to the icons representing the indices corresponding to each button. The values displayed on the interface indicate the frequencies and phases used for each button during the SSVEP stimulation. (b) Illustration of the relationship between the BCI commands and the 3D direction of the end-effector of the robotic arm.

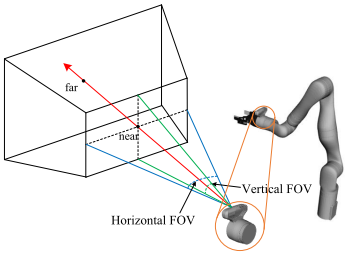
in the horizontal plane, as described below.

$$\overrightarrow{v} = \overrightarrow{v_0} + \overrightarrow{v_1} \tag{1}$$

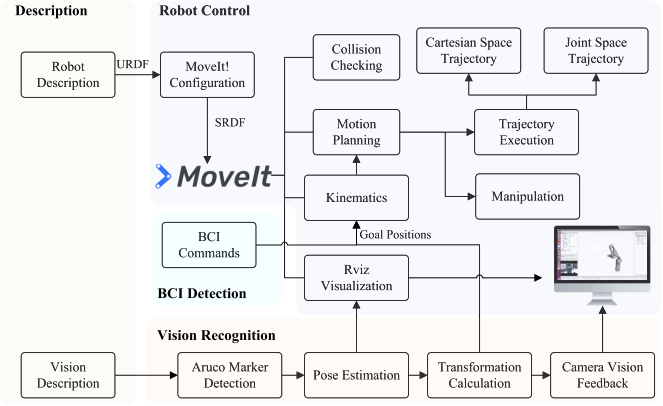
The additional four buttons including one of the third one in the second row and the three in the last column (marked as 8, 5, 10, 15) indicate the following commands: drinking (Dk), stopping the robot (Sp), moving the robot to the initial position (Hm) and changing the control mode of the robot between step movement and continuous movement (Ch). During the step movement control, a robot would move toward an appointed direction for 30 mm and stop to waiting for a next command, while it would keep steady movement in a speed of 3 mm/s toward a direction until receiving a next command during the continuous movement control. The parameter of moving speed was determined based on the average performance of the subjects prior to their official participation in the online test with the robotic arm.

2) Robotic Arm: A 7-DOF (i.e., each corresponding to a joint actuator) human-like robotic arm named Gen3 (Kinova Robotics Inc., Canada) with two fingers was used as the BCI actuator for all subjects. The robotic arm workspace was a cube with a 0.5-m side length and the hand orientation fixed in Cartesian space towards the end-effector. Visual boundaries were set for the workspace of the robotic arm as part of the control software to avoid collision with the tabletop. The robot was run on a platform of a robot operating system (ROS) to receive control commands from the BCI subsystem. Note that in order to ensure the independent display of the robotic arm's camera screen and the visual stimulation for the BCI and for the sake of operational convenience, we chose to develop the two subsystems on separate computers. This subsystem realized three functions: BCI-guided control, pose recognition, and vision-guided control. BCI-guided control describes when the robotic arm was moved with commands that were issued by the subjects through the hybrid asynchronous BCI subsystem. Vision-guided control describes when the robotic arm automatically executed the predetermined actions (e.g., grasping a water cup and moving it near a subjects' mouth) once the coordinates of the target objects were determined by the following computer vision subsystem.

3) Computer Vision: For perception data, RGB-D images of resolution 640 × 480 were captured with an Intel RealSense D435 camera statically mounted on the wrist of a robot



(a)



(b)

Fig. 3. (a) Illustration of FOV of camera for detecting the objects. (b) A schematic representation of the ROS architecture and flowchart for vision-guided robotic arm control.

arm. The camera streamed to a computer through USB 3.0. After the robot successfully moved into the capture region of the target under BCI-guided control, the camera detected the quick response (QR) code, which is a type of matrix barcode pasted on the target object, and then provided the location information for the robotic arm subsystem. Fig. 3(a) depicts a schematic representation illustrating the field of view (FOV) for QR code recognition. The FOV was quantified as $69.4 \pm 3°$ horizontally and $42.5 \pm 3°$ vertically. Notably, the precision of depth perception was influenced by factors such as QR code dimensions and ambient illumination. The most favorable range for optimal recognition was empirically determined to lie within 0.3 to 0.8 meters.

4) Shared Control: Shared control was achieved by integrating BCI-guided control and vision-guided control, especially for complex manipulation tasks of the robotic arm. As shown in Fig. 3(b), the control architecture for the robotic arm sub-system in the ROS framework comprises several components and modules that work together to enable robot control and interaction. The control system initiation involves defining the robot's physical structure and accurate kinematics using a URDF (Unified Robot Description Format) file for robot description. Subsequently, the MoveIt! configuration module was utilized to generate an SRDF (specify the robot description file) file, which provided high-level information such as groups, joint limits, and allowed collisions, enabling customization of the MoveIt! package. The package incorporated different kinematics solver (e.g., the Kinematics and Dynamics Library (KDL) was applied in this study), allowing for both forward and inverse kinematics calculation, as well

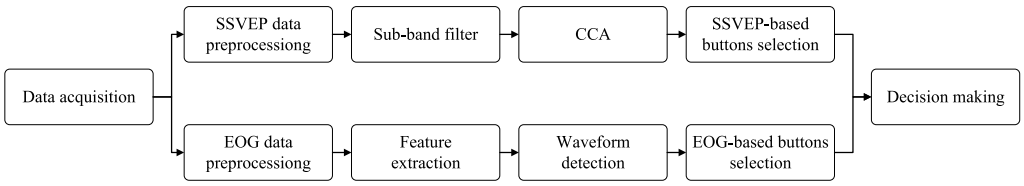


Fig. 4. Flowchart of data processing for EEG and EOG signals. At the end of each trial stimulation, these two different data are acquired, and the corresponding processing procedures are executed simultaneously. An FBCCA method consists of a subband filter and a CCA process, and blink-related waveform detection is implemented for the preprocessed SSVEP and EOG signals, respectively. With the selected buttons of the two data types, decision making can be used to determine the final output.

as a planning interface for trajectory generation. Note that the goal positions required for kinematics calculations can be generated from either BCI detection for BCI-guided control (refer to the forthcoming *EEG and EOG Data Analysis* section for details) or vision recognition for vision-guided control. These planned trajectories were then executed by commanding the robotic arms joint positions or velocities in either Cartesian space or joint space. Moreover, the resulting trajectories were visualized in RViz, offering real-time monitoring and debugging capabilities.

As for the vision-guided control component, the ArUco detection library was ultilized to identify the track markers with QR codes in the camera feed, allowing for precise localization of objects of interest. Subsequently, the poses of these markers were transformed from the camera frame to the robots base frame, enabling their integration into the MoveIt! motion planning pipeline. To achieve this, a ROS node was implemented, which subscribed to the ArUco markers associated with each target object. Furthermore, the recognized objects can be visualized in RViz, providing real time feedback to the subjects.

B. EEG and EOG Data Analysis

1) EEG Data Analysis: In this study, a filter-bank canonical correlation analysis (FBCCA), which was established for character spelling in our previous study [36], was applied as the SSVEP detection method. For the sake of completeness, we provide a brief overview of the method in this section. As illustrated in Fig. 4, during offline analysis and online testing for SSVEPs, data epochs consisting of signals from twenty-four channels were extracted by the stimulus program. The FBCCA method was then applied for extracting frequency-related features by taking into account both template data that modulating visual stimulation and individual calibration data, as well as the capture of specific harmonic information from SSVEP [29], [37], [38]. Specifically, the EEG data epoch was first decomposed into N subbands of harmonic components using bandpass filters (i.e., Butterworth infinite impulse response (IIR)) [39]. For the nth subband component of the SSVEP epoch, the lower-bound frequency was calculated as $6 + (n-1) \times 7$ Hz, while the upper-bound frequency was fixed at 48 Hz. Next, we calculated the correlation coefficients between each of the subband components and sine-cosine reference signals by employing standard canonical correlation analysis (CCA) [40], and integrate them into a vector $\boldsymbol{r}_{n,k}$ of the nth subband and the kth target. Considering two multidimensional variables $\boldsymbol{A}$ and $\boldsymbol{B}$ and their linear combinations $\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{U}$ and $\boldsymbol{y} = \boldsymbol{B}^T\boldsymbol{V}$, CCA calculates the weight vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ that maximize the correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$ by solving the following optimization problem:

$$\max_{\boldsymbol{U},\boldsymbol{V}} \rho(\boldsymbol{x},\boldsymbol{y}) = \frac{E\left[\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{B}^T\boldsymbol{V}\right]}{\sqrt{E\left[\boldsymbol{U}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{U}\right]E\left[\boldsymbol{V}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{V}\right]}} \tag{2}$$

where the maximum value of ρ is the maximum canonical correlation with respect to $\boldsymbol{U}$ and $\boldsymbol{V}$. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ were selected as i) $\boldsymbol{A_1} = \boldsymbol{X}_n$ and $\boldsymbol{B_1} = \boldsymbol{Y}_{f_k}$, ii) $\boldsymbol{A_2} = \boldsymbol{X}_n$ and $\boldsymbol{B_2} = \hat{\boldsymbol{X}}_{n,k}$, iii) $\boldsymbol{A_3} = \boldsymbol{Y}_{f_k}$ and $\boldsymbol{B_3} = \hat{\boldsymbol{X}}_{n,k}$, iv). In this study, $\boldsymbol{X}_n$, $\hat{\boldsymbol{X}}_{n,k}$ are the test EEG data epoch and the individual calibration data of the kth target, respectively. $\boldsymbol{Y}_{f_k}$ represents the sine-cosine template data epoch with frequency f_k and is created as follows:

$$\boldsymbol{Y}_{f_k} = \begin{bmatrix} \sin(2\pi f t + \phi) \\ \cos(2\pi f t + \phi) \\ \vdots \\ \sin(2\pi N_h f t + \phi) \\ \cos(2\pi N_h f t + \phi) \end{bmatrix} \tag{3}$$

where ϕ is the initial phase and N_h is the number of harmonics (in this study, $N_h = 3$).

A vector of correlation coefficients was then obtained using these three pairs of weight vectors [29], [38], i.e., $\boldsymbol{r}_{n,k}(1) = \rho(\boldsymbol{X}_n^T\boldsymbol{U}_{\boldsymbol{A_1B_1}}, \boldsymbol{Y}_{f_k}^T\boldsymbol{V}_{\boldsymbol{A_1B_1}})$, $\boldsymbol{r}_{n,k}(1+l) = \rho(\boldsymbol{X}_n^T\boldsymbol{U}_{\boldsymbol{A_1B_1}}, \hat{\boldsymbol{X}}_{n,k}^T\boldsymbol{V}_{\boldsymbol{A_lB_l}})$ for $l = 1, 2, 3$. Finally, the weighted feature of the kth target was used to identify the target.

$$p_k = \sum_{n=1}^{N} w(n) \cdot \sum_{i=1}^{4} sign(\boldsymbol{r}_{n,k}(i)) \cdot \boldsymbol{r}_{n,k}(i)^2 \tag{4}$$

where $sign()$ is utilized to maintain the discriminating information, while $w(n)$ denotes the weight coefficient for different SSVEP harmonic components as follows:

$$w(n) = n^{-a} + b, \; n \in \left[1 \; N\right] \tag{5}$$

where a and b are constants determined using a grid-search method during offline analysis (1 and 0, respectively, in this study).

Accordingly, we calculated the averaged correlative features in the control state as $p_{control}$ and in the idle state p_{idle}, after collecting the calibration data for SSVEP. Then, two thresholds were defined for asynchronous detection, namely $\delta_1 = \beta_1 \cdot p_{control}$ and $\delta_1 = \beta_2 \cdot p_{control}/p_{idle}$, respectively. In this study, β_1 and β_2 represent empirical constant weights within 0 to 1, with δ_1 and δ_2 fluctuate within the range of [0.5 0.8], and [1.05 1.25], respectively.

2) EOG Data Analysis: The EOG signals between 100 and 500 ms after cue onset were extracted and filtered to a frequency range of 1 to 10 Hz using a 2nd-order Butterworth bandpass filter. Next, the filtered epoch passed multithreshold waveform detection (as described in a previous study [36]), which implied that there was a clearer detected peak and valley in a waveform with a blink than in a nonblink waveform. We implemented a check criterion based on the distance between the peak and valley, as well as the regressive value obtained using an SVM model for each epoch. If these values exceeded the corresponding predefined thresholds (δ_d and δ_y as specified in [36]), which were derived from calibration EOG data specific to each subject, the corresponding button was considered to trigger an intended blink action. Additionally, the response time of individuals, defined as the difference between the time of peak onset during the online test and the predefined value after calibration, proved to be a crucial parameter in distinguishing between attended blinks, nonblinks, and unattended blinks. Therefore, the satisfied buttons for our check criterion were then considered the final EOG candidates if the response time for each of them was in a presupposed range (−160 to 160 ms in this study).

3) Decision Making: The results of intended blinking detection were considered initially in this phase. If the EOG data from a specific trial did not meet the blink detection criteria, the system concluded that no character was generated. Otherwise, further decision making would be conducted by applying the following three criteria after acquiring the two candidate sets:

1) Criterion I: $[r_1, r_2, \ldots, r_m] \cap [s_1, s_2] \neq \emptyset$.
2) Criterion II: $p_{k_{max}} > \delta_1$.
3) Criterion III: $p_{k_{max}}/p_{k_{second}} > \delta_2$.

where $[r_1, r_2, \ldots, r_m]$ are the candidates after blink detection and $[s_1, s_2]$ are the candidates with the top two values for p_k. $p_{k_{max}}$ and $p_{k_{second}}$ are the maximum and the largest value of p_k, respectively.

We decide whether to adopt criteria I based on the subjects' response to SSVEPs, as the multi-task for robot control requires a relatively fast response to EOG cue onset and SSVEP detection has the advantage of maintaining high accuracy and short response time simultaneously compared with EOG detection. If a subject achieved relatively unsatisfactory accuracy (e.g., beneath 92%) in the SSVEP offline experiment, all the above three criteria were applied for decision making. That is, once criteria I, II, and III are satisfied, the candidate with the largest value of p_k among $[r_1, r_2, \ldots, r_m]$ is recognized as the target. Otherwise, no result is output. However, if satisfactory accuracy for SSVEP detection is generated for one subject, only criterion II and criterion III would be considered, which means that the final target is the candidate with the largest value of p_k among all fifteen buttons. Otherwise, no result is output at the end of the current trial.

C. Experimental Flow

All subjects were first performed the BCI training and cue spelling session without actual control of the robotic arm. The BCI training session was designed to help the subjects learn and improve their ability to efficiently modulate SSVEPs and blink their eyes simultaneously with cue onset. Then, two online experiments corresponding to robotic arm control were conducted.

1) BCI Training and the Online Free Spelling Session: First, the subjects were required to participate in the training phase in which the SSVEP dataset (60 trials in the control state and 15 trials in the idle state) and the EOG dataset (10 trials in the control state and 10 trials in the idle state) were collected to calibrate thresholds used in the asynchronous mode. Note that the subjects were allowed to perform basic physiological activities including eye movements while avoiding severely strenuous events. The thresholds for each subject were then determined (see above *EEG and EOG Data Analysis section for details) after the above calibration datasets were acquired.* Next, each subject needed to conduct an online free spelling experiment in which 30 trials were selected (each character was designed to be chosen for two trials) in the asynchronous mode. The spelling results were presented to the subjects in a visual feedback fashion, as each corresponding target was surrounded by a red square.

2) BCI-Controlled Robot 3D Reach Task: After the free spelling test, we assessed whether each subject could manipulate the robotic arm to reach for a ping-pong ball that has a diameter of 3 cm and was connected with a support lever. The ping-pong ball was placed at a random position in 3D space, but the position was kept the same for all subjects. The subjects needed to maneuver the robotic arm within a narrow range of approach angles while avoiding touching the lever below, as the hand aperture was slightly larger than the target ball. A successful 'reach' trial occurred when a subject controlled the arm to touch the target ball from the start position and envelop it with its fingers.

First, the robot 3D reach task required subjects to control the robotic arm in the step movement mode from three different start positions (i.e., the start positions in Fig. 8) to the target position (i.e., the location of touching the ping-pong ball). Each position was executed for two runs. In the step movement mode, people utilized the hybrid asynchronous BCI subsystem to send a new movement command, taking into account the current position of the robotic arm. Upon receiving each command, the robotic arm would move step by step towards the target position. We familiarized the participants with the task for approximately 10 min, during which the participants learned the physical space in which the directional commands would be effective in moving the robotic arm close enough to the target ping-pong ball. Specifically, all subjects first chose one of the three commands for the vertical direction according to the height of the end effector of the robotic arm from the target and then selected one of the eight commands with different directions in the horizontal plane. Second, in addition to the step movement mode, the subjects also needed to move the robotic arm by sending BCI commands in the continuous movement mode to finish the 3D reach task from the first start position, as above, to the target position for two runs. In the continuous movement mode, the robotic arm would maintain a consistent speed and change its direction promptly

upon receiving a new BCI command. It would stop until the command of "Sp" was generated by subjects.

3) BCI-Controlled Robot Reach-Grasp-Drink Tasks: Subjects conducted the reach-grasp-drink task in the shared control mode. First, they were asked to execute this task for a single target object (e.g., a water cup) in which the robotic arm moved from the first start position (Fig 12(a)) toward the target object. Once the target entered the viewing field of the camera, the participants issued a grasp command, and the robotic arm finished the remaining actions with the help of vision-guided control, including grasping the cup, moving to the drinking position, and placing the cup back onto the desktop. Each subject completed two runs of this task. Then, a reach-grasp-drink task for multi objects with shared BCI control was executed, in which participants selected the drinking button through the hybrid asynchronous BCI sub-system after the robot guiding the camera to observe the objects, as in the task for single object, and chose the button corresponding to the target cups from the three cups shown in a subpage (see Fig 10). If a wrong object was selected or the robotic arm finished the drink action, they can select a return button to the main window. The robotic arm would cease its movement, return to the initial position and wait for the subjects to issue a new BCI command. Note the selected frequencies corresponding to three objects from left to right and a return button were 7.6 Hz, 9.6 Hz, 11.6 Hz and 12.8 Hz, respectively. The corresponding phases were 0.5π, 1.0π, 1.5π, 1.0π, respectively. Similarly, the robot would implement the remaining actions of drinking, as in the single object grasp task. One run by each subject for each target object. A QR code was attached to each subject's chest to identify the location of his/her mouth.

Furthermore, an additional control experiment for the robot reach-grasp-drink task in the step movement control with direct BCI control was conducted with all subjects. In this experiment, the participants steered the robotic arm, which was controlled directly with their selected BCI commands, to gradually approach the target. They then sent a grasp and drink command until the robotic arm moved to the point where the finger could hold the target water cup. The robot would subsequently execute the following actions of drinking. Here, the start position of the robot was the same as that in the task for a single object. Two runs of the task were performed by each participant.

D. Subjects and Data Acquisition

Ten healthy subjects (all males; age range: 20-28 years old) were recruited for this study. Prior to participating in the offline and online experiments, all subjects provided written informed consent for their data to be used and published. This study was also approved by our cooperative institution of the Ethics Committee of Sichuan Provincial Rehabilitation Hospital (approval number: CKLL-2018008).

A liquid-crystal screen (resolution: 1,920 × 1,080 pixels, refresh rate: 60 Hz) was applied to present stimuli for subjects. Also, a amplifier Synamps2 with 64 channels at a sampling rate of 250 Hz was used to record EEG and EOG data after stimuli. The impedances of all electrodes were maintained below 10 kΩ. The ground electrode was placed on the forehead

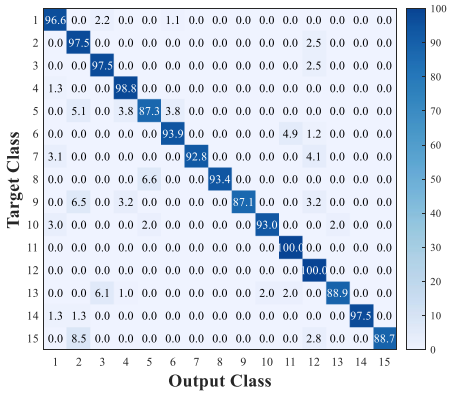


Fig. 5. Confusion matrix across all subjects with 6-fold cross-validation using 2 s-length data segment from offline training. The color bars indicate the classification accuracies, with the number on the diagonal representing the correct output.

(GND), while the reference electrode was placed on the right mastoid (A2). EEG data were acquired from 24 channels located in the occipital region, which included CP1, CPz, CP2, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, CB1, O1, Oz, O2, and CB2. EOG data were acquired from one channel on the forehead (Fp1).

III. Results

A. Experiment I: Performance of Offline SSVEP Training and the Online Free Spelling Task

The confusion matrix of the fifteen targets recognition in the offline SSVEP training task is illustrated in Fig. 5. All subjects achieved the accuracy above 80% for each target. Post-hoc comparison with Bonferroni correction suggested no significant difference between the targets from each other ($p > 0.05$), which revealed the appropriated frequency distribution for SSVEP stimulation. Moreover, we compared the averaged classification accuracy across all subjects with data length at 2000 ms between the applied FBCCA method with several commonly used methods such as the minimum energy combination (MEC) [41] and the classical canonical correlation analysis (CCA) [42]. The results demonstrated that FBCCA-based method outperformed other methods (FBCCA vs. CCA vs. MEC, 95.67 ± 4.46% vs. 80.33 ± 9.22% vs. 52.00 ± 21.44%), with one-way repeated ANOVA measures depicted significant main effects ($F(2, 27) = 304.46$, $p < 0.001$). In terms of the performance in the online free spelling task, table I depicts the classification accuracy, FPR in the idle state, and ITR in the control state across all subjects with the same data length (2.5 s, which consisted of 2.0 s of the stimulation and 0.5 s of the trial shift). The ITR metric used to evaluate the integrated performance of BCI systems was calculated as follows:

$$ITR = 60(\log_2 M + P\log_2 P + (1-P)\log_2(\frac{1-P}{M-1}))/T \quad (6)$$

where T is the mean response time (2.5 s in this study) to select a command, M is the number of classes (15 in this study), and P is the accuracy. The average classification

TABLE I
RESULTS OF AN ONLINE FREE SPELLING TEST WITH THE HYBRID ASYNCHRONOUS BCI SUB-SYSTEM

Subject	Epoch length (s)	Accuracy (%)	FPR (%)	ITR (bits/min)
S1	2.5 (2.0+0.5)	100.0	0.83	117.2
S2	2.5 (2.0+0.5)	84.9	1.67	81.6
S3	2.5 (2.0+0.5)	89.2	1.67	90.1
S4	2.5 (2.0+0.5)	93.8	0.83	100.1
S5	2.5 (2.0+0.5)	93.8	1.67	100.1
S6	2.5 (2.0+0.5)	83.3	0.83	78.6
S7	2.5 (2.0+0.5)	90.9	0.83	93.6
S8	2.5 (2.0+0.5)	100.0	0.0	117.2
S9	2.5 (2.0+0.5)	93.8	1.67	100.1
S10	2.5 (2.0+0.5)	93.8	2.50	100.1
Mean ± SD	-	92.4 ± 5.5	1.25 ± 0.71	97.9 ± 12.8

accuracy in the online free spelling task was 92.4% ± 5.5%, resulting in an ITR value of 97.9% ± 12.8% bits/min across all subjects, as well as a low FPR value of 1.25% ± 0.71%. The maximal and minimal ITR values were 117.02 bits/min (subjects S1 and S8) and 78.6 bits/min (subject S6). The online results of the free spelling task demonstrated the feasibility of implementing BCI-controlled robot systems.

B. Experiment II: Performance of the BCI-Controlled Robot 3D Reach Task

Table II illustrates the results of the robot 3D reach task using step movement control. All subjects successfully completed the robot 3D reach task, which consisted of three sessions, each corresponding to a different start position of the robotic arm. The average completion time needed to reach target from start position 1, start position 2 and start position 3 were 1.92 ± 0.63 minutes, 1.58 ± 0.47 minutes and 1.59 ± 0.48 minutes, where the numbers of commands issued by the subjects were 22.10 ± 4.98, 18.50 ± 4.19 and 17.70 ± 2.05, respectively. Furthermore, the numbers of wrong commands for the three sessions were 1.10 ± 1.71, 1.05 ± 1.67, and 1.00 ± 0.79, respectively. One-way repeated-measure ANOVA tests followed by Bonferroni post hoc tests revealed that the numbers of incorrect commands for the three sessions did not differ from each other ($F(2, 57) = 0.02$, $p = 0.97$). For all individuals, subject S8 achieved the best performance with 0 total error commands reported as well as the shortest total complete-time of 7.11 minutes and the least total number of issued commands of 104, while the performance of subject S2 was worst with the most total error commands of 24, the longest total complete-time of 14.72 minutes, and the highest total number of issued commands of 151. To provide a more intuitive demonstration of the experimental process, Fig. 6 depicts a representative trial of subject S9 guiding the robotic arm to move from the first initial point and ultimately touch the target ping-pang ball.

To evaluate the efficiency of the robot's moving trajectory from the different start positions to the end positions in 3D space, the trajectory efficiency (TE) was also recorded, which can be defined as follows:

$$TE = tr_0/tr_1 \tag{7}$$

where tr_0 and tr_1 are the straight-line distance and actual moving distance from the start position to the end position,

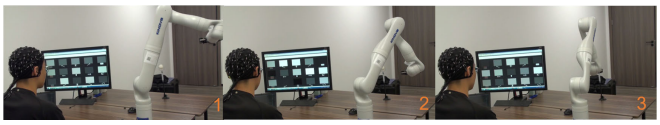


Fig. 6. Subject S9 finished the BCI-controlled robot 3D reach task in the control mode of step movement. Three sequential snapshots (from 1 to 3) show subject S9 successfully guiding the Gen3 robotic arm to move from the start position (1), close to the target ping-pang ball (2), and to touch the ball (3).

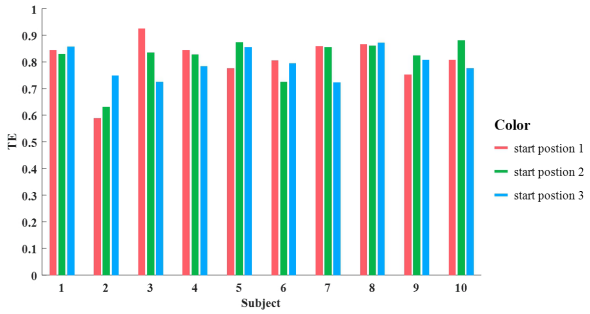


Fig. 7. TE of the robot 3D reach task for each subject. The theoretical value of TE was 1. The one-way repeated-measure ANOVA test revealed no significant difference in the three different start positions (F(2, 57) = 0.25, p = 0.78).

respectively. The value of TE varies from 0 to 1, and a larger value indicates a more optimized trajectory. As shown in Fig. 7, the average TE values across all subjects for three different sessions were 0.80 ± 0.10, 0.81 ± 0.10, and 0.79 ± 0.07.

We also mapped the position profile of the end-effector of the robotic arm generated through ROS to the 3D position dimension (3D axes) to represent the arm trajectory pathway. Fig. 8 presents the trajectory pathway reconstruction of the robot 3D reach task completed by subject S8. He successfully completed three trials, where they moved the robot from three different start positions toward the same target (i.e., the ping-pong ball) with no inaccurate commands and TE values of 0.87, 0.90, and 0.87.

Table III lists the results of the robot 3D reach task using continuous movement control. Further statistical analysis using a paired t-test revealed that the subjects reported fewer issued commands ($p < 0.05$) and smaller TE values ($p < 0.001$) after finishing the task in the mode of continuous movement than in the mode of step movement. However, there were no significant differences in terms of the duration of the task ($p > 0.05$) and the number of wrong commands ($p > 0.05$) between these two modes (Fig. 9). This phenomenon consists with the fact that an extremely short task duration was not expected from the subjects, as they were naive to the BCI-controlled manipulation. Additionally, highly attention was objectively required to the visual stimulation on the screen and the ongoing movement of the robotic arm simultaneously during the continuous movement control. The results showed that people could finish the BCI-controlled robot 3D reach task in the control mode of either step movement or continuous movement, with acceptable accuracy even from different initial robot positions which correspond to different trajectories with specific sets of distances and directions. Nevertheless, the

TABLE II
RESULTS OF THE ROBOT 3D REACH TASK IN THE STEP CONTROL MODE

Subject	Position 1			Position 2			Position 3		
	N	E	T	N	E	T	N	E	T
S1	23, 23	0, 1	1.85, 1.62	22, 18	2, 0	1.91, 1.47	15, 15	0, 0	1.22, 1.05
S2	31, 38	4, 7	3.00, 3.55	34, 20	6, 3	2.92, 1.63	21, 18	2, 2	1.82, 1.68
S3	16, 17	0, 0	1.33, 1.47	17, 18	0, 2	1.67, 1.52	17, 22	1, 2	1.80, 1.63
S4	23, 21	1, 1	2.12, 1.82	20, 17	1, 0	1.95, 1.42	17, 17	2, 1	1.33, 1.60
S5	23, 24	1, 1	2.73, 2.15	17, 14	0, 0	1.78, 0.95	19, 16	0, 1	1.95, 1.85
S6	21, 17	0, 0	1.33, 1.10	22, 16	4, 0	1.42, 0.97	16, 20	0, 1	0.93, 1.62
S7	17, 22	1, 0	1.28, 2.37	15, 18	0, 1	1.42, 2.37	21, 19	1, 2	2.48, 2.70
S8	20, 20	0, 0	1.55, 1.35	18, 16	0, 0	1.43, 0.95	16, 16	0, 0	0.88, 0.95
S9	20, 23	0, 2	1.62, 1.97	17, 18	0, 0	1.32, 1.45	16, 17	1, 1	1.27, 1.45
S10	21, 22	1, 2	2.18, 2.03	16, 17	0, 2	1.40, 1.67	18, 18	1, 2	1.80, 1.65
Mean ± SD	22.10 ± 4.98	1.10 ± 1.71	1.92 ± 0.63	18.5 ± 4.19	1.05 ± 1.67	1.58 ± 0.47	17.70 ± 2.05	1.00 ± 0.79	1.59 ± 0.48

"N", "E" and "T" denote the total number of issued commands, final execution error number, and duration (or completion time) for each attempt, respectively. Two runs were completed for each start position.

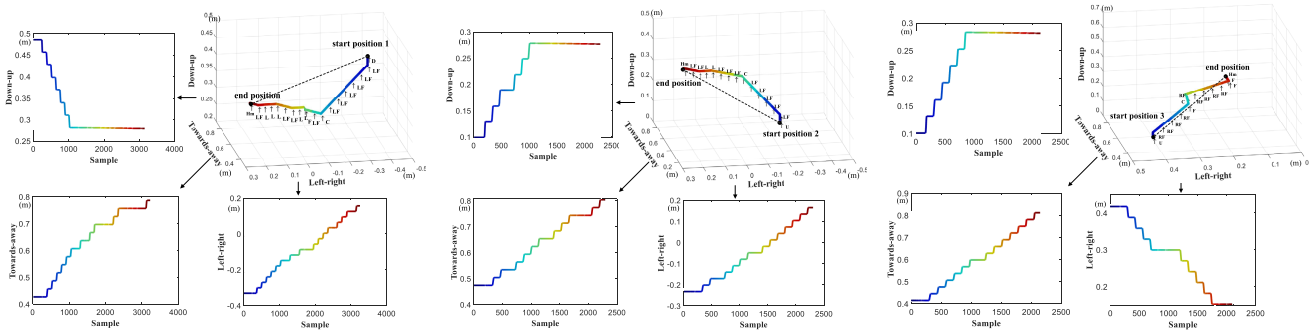


Fig. 8. Trajectory samples from three-dimensional view and single axial view of subject S8 finishing three successful trials during the robot 3D reach task. The colored solid lines represent the actual moving trajectory of the robotic arm following each BCI command (marked with arrows beneath the line), while the gray dashed lines indicate the shortest path between the start positions and the end positions.

TABLE III
RESULTS OF THE ROBOT 3D REACH TASK IN THE CONTINUOUS CONTROL MODE

Subject	Position 1		
	N	E	T
S1	15, 17	0, 0	1.48, 1.75
S2	17, 23	3, 6	1.70, 2.12
S3	18, 18	2, 1	2.73, 2.70
S4	13, 13	1, 1	1.46, 1.50
S5	18, 16	1, 0	1.47, 1.47
S6	27, 11	3, 1	2.80, 2.03
S7	20, 14	1, 1	3.00, 1.65
S8	16, 16	1, 1	1.90, 1.91
S9	15, 14	2, 1	1.73, 1.92
S10	18, 12	2, 0	2.02, 1.50
Mean ± SD	16.55 ± 3.75	1.40 ± 1.39	1.96 ± 0.48

"N", "E" and "T" denote the total number of issued commands, final execution error number, and duration for each attempt, respectively. The task of start position 1 was executed for two runs.

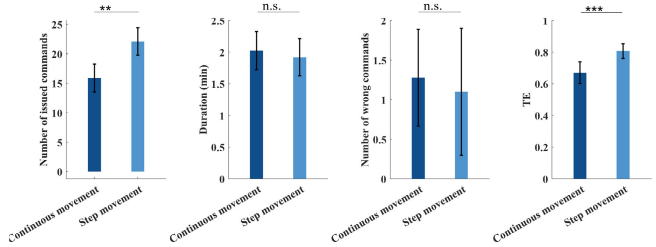


Fig. 9. Comparison of the online experiment performance results using step control and continuous control for robot 3D movement. The black asterisks ** and *** denote that there are statistically significant differences at the 0.05 and 0.001 levels, respectively, while n.s. represents no significant difference. The error bars show the standard errors of the means.

subjects stated that it was more difficult to move the robotic arm using continuous movement due to the stricter requirements of mental concentration.

C. Experiment III: Performance of the BCI-Controlled Robot Reach-Grasp-Drink Task

Table IV lists the results of the robot reach-grasp-drink task for a single object (e.g., a water cup) and multiple objects (three water cups) with the shared BCI control strategy. Generally, all subjects completed the tasks of BCI-controlled robot movement and sent a grasp drink command. On average, they were able to finish this task for a single object in 0.99 ± 0.35 minutes of BCI-guided control by issuing 9.40 ± 2.14 commands, as well as 0.35 ± 0.59 incorrect commands. In terms of the task for multiple objects, the average number of issued commands was 11.50 ± 5.07, the average number of wrong commands was 1.27 ± 1.72, and the average completion time for BCI-guided control was 1.42 ± 0.96 minutes across all subjects. It is worth to note that the whole completion time

TABLE IV
RESULTS OF THE ROBOT REACH-GRASP-DRINK TASK THROUGH THE SHARED BCI CONTROL

Subject	Single object			Multiple objects		
	N	E	T	N	E	T
S1	8, 9	0, 0	0.50, 0.91	7, 8, 7	1, 0, 0	0.85, 0.78, 0.52
S2	12, 12	1, 1	1.27, 1.37	10, 7, 12	1, 1, 1	1.15, 0.90, 1.30
S3	7, 10	0, 0	1.30, 1.17	17, 29, 22	3, 8, 4	2.89, 4.98, 3.56
S4	8, 8	0, 1	0.60, 057	17, 12, 12	4, 2, 2	2.42, 1.32, 1.35
S5	12, 10	0, 0	0.78, 1.55	14, 11, 11	1, 0, 1	1.58, 1.40, 1.46
S6	10, 15	0, 2	0.85, 1.65	9, 7, 10	0, 0, 1	0.78, 0.53, 0.97
S7	8, 8	0, 0	1.18, 1.18	20, 8, 9	3, 0, 0	2.20, 1.18, 0.85
S8	6, 8	0, 0	0.45, 0.81	8, 9, 8	1, 0, 0	0.85, 0.85, 0.73
S9	9, 9	0, 1	0.73, 0.70	14, 11, 11	1, 1, 1	1.27, 0.97, 0.75
S10	11, 8	1, 0	1.30, 0.87	9, 8, 9	0, 0, 1	1.28, 1.35, 1.58
Mean ± SD	9.40 ± 2.14	0.35 ± 0.59	0.99 ± 0.35	11.50 ± 5.07	1.27 ± 1.72	1.42 ± 0.96

"N", "E" and "T" denote total number of issued commands, final execution error number, and duration of BCI-guided control for each attempt, respectively.

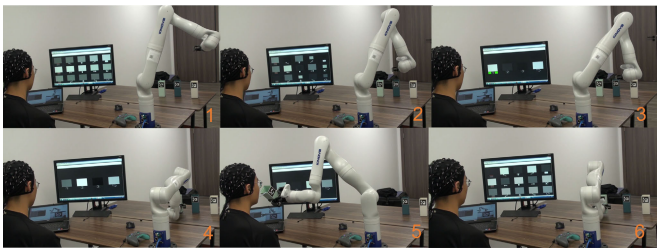


Fig. 10. Subject S9 drinking a cup of water using the Gen3 robotic arm. Six sequential snapshots (from 1 to 6) of the reach-grasp-drink task show the whole process that he successfully guiding the robotic arm close to the bottles (1 and 2), choosing the target bottle (3 and 4) by themself, drinking water (5) through a straw with the help of computer vision and placing the bottle on the table automatically by robot(6).

TABLE V
RESULTS OF THE ROBOT REACH-GRASP-DRINK TASK THROUGH DIRECT BCI CONTROL

Subject	Position 1		
	N	E	T
S1	28, 24	3, 0	2.82, 1.98
S2	24, 49	2, 8	2.55, 3.62
S3	40, 40	4, 4	6.76, 6.80
S4	31, 31	4, 4	3.20, 3.10
S5	24, 25	2, 1	3.07, 4.18
S6	26, 31	0, 2	2.48, 2.28
S7	29, 23	0, 2	2.63, 2.23
S8	23, 28	0, 0	1.37, 2.23
S9	27, 24	2, 1	2.32, 2.28
S10	27, 24	4, 2	2.58, 2.20
Mean ± SD	28.90 ± 6.83	2.25 ± 2.02	3.04 ± 1.42

"N", "E" and "T" denote the total number of issued commands, final execution error number, and duration of BCI-guided control for each attempt, respectively.

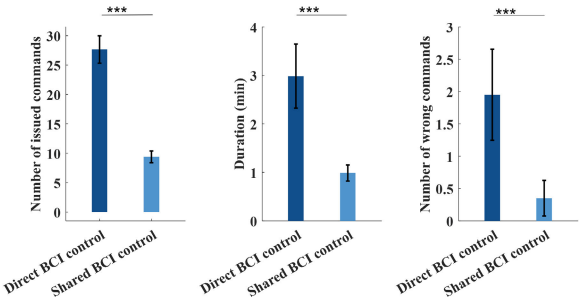


Fig. 11. Comparison of the online experimental performance results with the strategies of direct BCI control and shared BCI control. The black asterisk *** denotes that there are statistically significant differences at the 0.001 level, while n.s. denotes that there is no significant difference. The error bars show the standard errors of the means.

of one attempt for both single object and multiple objects should be the above time of BCI-guided control plus the time of vision-guided control (0.70 minutes in this study) which helps robot perform the remaining actions of drinking. Subject S8 finished the task for a single object with the least number of commands, the smallest task duration, and zero error commands, while subject S1 completed the multiobject task with the least number of commands, the smallest task duration, and one error command. Fig. 10 shows snapshots in which subject S9 successfully performed the reach-grasp-drink task with the proposed shared brain-control robotic arm system. Once the system was under vision-guided control, the robotic arm could always finish the reach and grasp tasks successfully.

Furthermore, Table V shows that all subjects achieved the average performance (i.e., the number of issued commands was 28.90 ± 6.83, the number of wrong commands was 2.25 ± 2.02, and the completion time of BCI-guided control was 3.04 ± 1.42 minutes) during the robot reach-grasp-drink task using the direct BCI control strategy. It is also worth to note that the remaining execution time of drinking for robotic arm after BCI-guided control was 0.67 munites in this study. Compared with using direct BCI control, the statistical analysis revealed that people used significantly less BCI commands with higher accuracy and shorter completion times to accomplish the robot reach-grasp-drink task using shared BCI control (Fig. 11).

To gain more insight into the task execution process through the shared BCI-controlled strategy, Fig. 12 shows the trajectory of robotic movement for subject S8 during the BCI-controlled period. He accomplished three trials of moving

the robot from the first starting position and sending a grasp command corresponding to one of the three target water cups once the target entered the vision detection area of the robot's camera. Finally, the water cup was moved near their mouth and was guided automatically by the robotic arm.

IV. Discussion

In this study, we combined hybrid asynchronous BCI and computer vision to develop a shared robotic arm control system. The hybrid asynchronous BCI model was used to select one of the fifteen commands for the robotic arm action, while computer vision was responsible for detecting target objects in the workspace. The average values of the number of issued commands, the completion time and the number of error commands during the BCI-controlled robot 3D reach task were 18.94, 1.64 minutes, and 0.90, respectively, with the exact value time depending on the complexity of the task and the performance of the subject. In addition, the shared BCI control strategy has demonstrated the advantage of shorting the completion time and improving the accuracy with less commands compared with the direct BCI control strategy. To the best of our knowledge, the proposed system is the first noninvasive BCI system that achieves robot 3D movement of brain control in an asynchronous mode and that incorporates machine vision to implement complex tasks, such as grasping and drinking.

It is crucial to achieve high accuracy with relatively satisfactory ITR and lower FPR values simultaneously for asynchronous BCI, which can make a brain-controlled robotic arm system more flexible and practical for real-life applications [25], [43]. During the BCI-guided control period, the subjects needed to focus on the task state when controlling the movement of the robot arm and needed to maintain the idle state either between two consecutive commands or during the vision-guided control period. It should be noted that pursuing satisfactory performance for online accuracy was more important than seeking a high ITR value for a brain-controlled robotic arm system. On the one hand, lower accuracy can lead to redundant robot trajectories and therefore indirectly increase the overall duration of experiments. On the other hand, controlling the robot for different actions is a highly integrated task in which participants shift frequently between the control state and the idle state. Also, this autonomy in deciding the timing of instructions of our system contributes to a short interval (0.5 s in this study) between two consecutive trials compared with other synchronous BCI robotic control systems (e.g., 4.25 s in a SSVEP-based study [44]). The comparison between the results in Table I and Table II revealed there was no significant difference in online accuracy between the two tasks, demonstrating the feasibility of utilizing the proposed asynchronous BCI for complex robotic arm control. However, the average time to select a control command in the robot relative task across all participants was 5.24 s, which revealed that participants took longer intervals between two consecutive commands in this task than in the spelling task. This was because people chose to spend more rest time to consider the next command due to the complexity of operating in a 3D environment and shifting between the stimulation screen and the robotic arm. Moreover, the results in Table II and Table III demonstrate that by using the asynchronous BCI, all participants could successfully steer the robotic arm toward the target object not only in the step movement mode but also in the mode of continuous movement. It can also be concluded that the subjects (e.g., S1, S4, S7, and S8) with relatively high online accuracy were inclined to achieve better performance than others during the brain-controlled reach task with relatively few wrong commands and short completion times.

In addition to the online performance of asynchronous BCI, building a proper strategy that converts the brain-selected results among multiple commands into practical robot actions is also helpful to improve the efficiency of robotic arm-related tasks. Several reports have proposed goal-oriented control strategies (i.e., the BCI commands were directly connected to the targets in known situations rather than guiding the robots to move) for BCI-controlled robotic arms [44], [45], [46], [47], [48], [49]. In addition, a direct control strategy for robot movement was investigated to handle more complex tasks in which the targets were unknown. For example, Edelman et al. [50] used a motor imagery-based BCI to realize robotic arm control for continuous random target tracking in two-dimensional space. Chen et al. [25] also investigated a strategy in which a robotic arm successively follows a one-dimensional axial motion in the XYZ plane of 3D space. Compared with these, our proposed 3D vector synthesis strategy has the advantage of providing more efficient control commands in 3D space, which could lead to a higher TE value than traditional one-dimensional axial control (approximately 0.8 vs. less than 0.6). In fact, this strategy even outperformed an MI-relative robot control strategy [51] that only works in 2D space in terms of TE (0.8 vs. 0.7). Moreover, this strategy allows for a reduced number of character encodings to generate BCI commands for different directions. As a result, the task duration can even be further decreased. The benefits of using fewer characters are twofold: it leads to improved task accuracy and lowers the FPR. Additionally, subjects are not required to frequently modify the robotic arm's vertical movement direction during specific tasks. Finally, all subjects successfully finished the BCI-controlled robot 3D reach task from three start positions, demonstrating the good robustness of our system, which enables people to stably customize a flexible trajectory of the robotic arm for unknown targets.

Shared control is a good combination of goal-oriented control and direct control and allows users to freely switch between the above two levels of control [44], [51], [52], [53]. On the one hand, it can transcend the limitation of goal-oriented control that all operations are required to be fully defined in advanced. On the other hand, it requires less mental effort from subjects compared to direct control. In our case, direct control was used to explore the unconstrained movements of a robotic arm and hence enhanced the workspace of the robot. Once the objects were captured by the camera that moved along with the end-effector of the robotic am, goal-oriented control (i.e., the vision-guided control in this study) took over and executed the following actions more accurately and flexibly than direct control. This kind

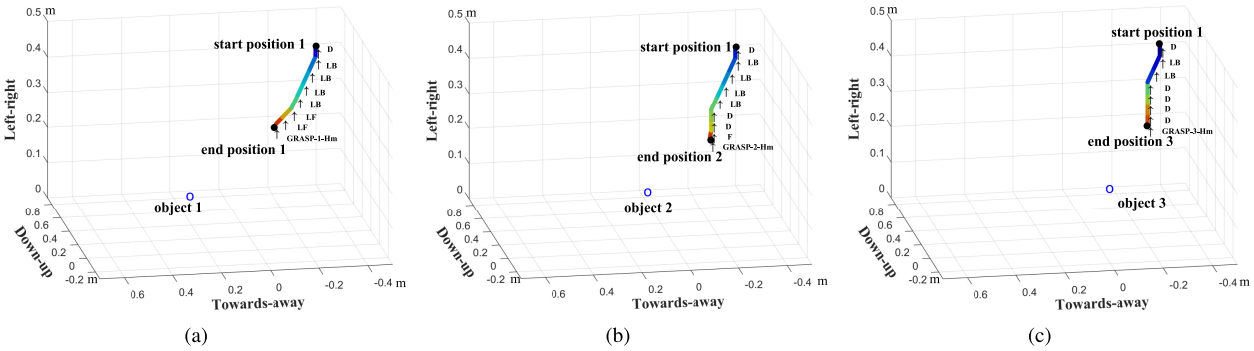


Fig. 12. Trajectory samples during BCI-guided control of subject S8 completing three successful trials in the robot 3D reach-grasp-drink task through the shared BCI control strategy.

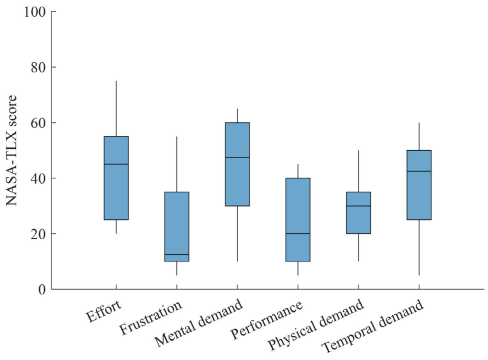


Fig. 13. NASA TLX score by dimension. The box plots represent the median values (solid bar) and IQR values (margin of box), and the whiskers denote the 95% confidence intervals.

of two-step strategy allows users to freely move the robotic arm close to the cup/cups, select the target, and drink water with the assistance of vision intelligence. The experimental results showed that all subjects finished guiding the robotic arm's movement with similar durations either through shared control (Table IV) or through direct control (Table II) due to the generally high accuracy of asynchronous BCI (> 90% for all except S2) and the similarly selected commands for optimized moving trajectories. However, there were significant differences (Fig. 11) in the performance between the shared control and the direct control, as the subjects generated fewer issued commands (9.40 ± 2.14 vs. 28.90 ± 6.83), shorter durations (0.99 ± 0.35 vs. 3.04 ± 1.42) and fewer wrong commands (0.35 ± 0.59 vs. 2.25 ± 2.02) in the reach-grasp-drink task using the former than using the latter. Therefore, the feedback from subjects after the experiments demonstrated that using the former caused much less fatigue than using the latter.

Grasp detection with visual sensors is the critical step in vision-guided control period of the shared control. In real-time experiments of shared control, 3D information about targets from a depth camera was mapped to real-world coordinates for subsequent trajectory planning. For the camera configuration, we offered a solution called eye-in-hand (i.e., the camera is attached to the end-effector of the robot) for BCI-controlled robotic arms. Compared with the other solution called eye-to-hand (i.e., the camera is placed at a fixed point in the workspace), which was used in several studies of BCI-controlled robots ([44], [51], [54]), the eye-in-hand approach applied in this study has higher recognition accuracy. Moreover, the connatural shortage of limited sight for this approach can be overcome and the range of its visual perception can even exceed the range using other approach when integrated with BCI-guided control, which gives a camera the ability to explore a 3D workspace.

After the online experiment, each subject was invited to complete a self-reported task load survey that evaluated six factors such as mental demand, physical demand, performance level, and frustration level and was measured by the NASA-TLX questionnaire scores (Fig. 13). Scores less than 50 or 60 can be regarded as acceptable according to some studies [55], [56], as a higher score indicates a higher subjective workload and vice versa [57]. Specifically, the two factors of mental demand and effort achieved higher scores than the other factors, suggesting that fatigue associated with the proposed system may be mainly caused by mental involvement and the effort needed for multiple tasks (i.e., staring at the flickering buttons and blinking eyes simultaneously with the cue onset). However, the average scores were less than 30 for the factors of the performance and frustration levels, indicating that people were overall satisfied with their task completion.

In the future, future improvements can be achieved to promote the usability of the proposed system for real-life applications in the following directions. First, the performance of the system with elderly individuals of paralyzed patients requires further investigation, as their health states may fluctuate. Second, the calibration time of SSVEP recognition can be shortened using transfer-learning techniques. Thirdly, it should be noted that QR code detection is not the ultimate solution for vision-guided control. To enhance the practicality of object recognition and mouth tracking in subjects using a camera in real-world scenarios., the proposed system would benefit from a more adaptable computer vision approach. This could involve leveraging deep learning technologies for image and depth identification, enabling more flexible and accurate analysis. Furthermore, while the TCP/IP protocol effectively prevented any loss or alteration of BCI commands during the experiments, and the transmission delay remained within an acceptable range (within 60 ms in this study), it may be necessary to consider Bluetooth transmission as an alternative

in cases where the wireless network signals are too weak. Finally, more intelligence techniques on robotics, such as path planning and avoidance of obstacles, can be integrated to improve the overall performance of the proposed system.

V. Conclusion

In this study, a novel shared control strategy that combined hybrid asynchronous BCI with computer vision was investigated to control a robotic arm. The results of the experiments demonstrated that subjects could operate the robotic arm to move to an indeterminate position in any given 3D environment with both high accuracy and low FPR values simultaneously. Furthermore, with the assistance of computer vision, people were able to to perform more complex and accurate tasks, such as grasping and drinking, with few commands and high accuracy. The BCI-guided control enabled an enhanced moving range of the robotic arm, while subsequent vision-guided control contributed to completing precise movements more efficiently, thereby reducing the user's mental effort and allowing the system to be used for longer periods of time.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] M. A. Lebedev and M. A. L. Nicolelis, "Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation," *Physiolog. Rev.*, vol. 97, no. 2, pp. 767–837, Apr. 2017.

[3] C. Pandarinath et al., "High performance communication by people with paralysis using an intracortical brain-computer interface," *eLife*, vol. 6, p. e18554, Feb. 2017.

[4] C. E. Bouton et al., "Restoring cortical control of functional movement in a human with quadriplegia," *Nature*, vol. 533, no. 7602, pp. 247–250, May 2016.

[5] Y. Pei et al., "A tensor-based frequency features combination method for brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 465–475, 2022.

[6] Y. Pei et al., "Data augmentation: Using channel-level recombination to improve classification performance for motor imagery EEG," *Frontiers Hum. Neurosci.*, vol. 15, Mar. 2021, Art. no. 645952.

[7] S. W. Brose et al., "The role of assistive robotics in the lives of persons with disability," *Amer. J. Phys. Med. Rehabil.*, vol. 89, no. 6, pp. 509–521, 2010.

[8] L. Bi, X.-A. Fan, and Y. Liu, "EEG-based brain-controlled mobile robots: A survey," *IEEE Trans. Hum.-Mach. Syst.*, vol. 43, no. 2, pp. 161–176, Mar. 2013.

[9] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 3, pp. 579–590, May 2013.

[10] G. Cheng, S. K. Ehrlich, M. Lebedev, and M. A. L. Nicolelis, "Neuroengineering challenges of fusing robotics and neuroscience," *Sci. Robot.*, vol. 5, no. 49, pp. 7–10, Dec. 2020.

[11] J. K. Chapin, K. A. Moxon, R. S. Markowitz, and M. A. L. Nicolelis, "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex," *Nature Neurosci.*, vol. 2, no. 7, pp. 664–670, Jul. 1999.

[12] J. Wessberg et al., "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates," *Nature*, vol. 408, no. 6810, pp. 361–365, Nov. 2000.

[13] M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, and A. B. Schwartz, "Cortical control of a prosthetic arm for self-feeding," *Nature*, vol. 453, no. 7198, pp. 1098–1101, Jun. 2008.

[14] L. R. Hochberg et al., "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, May 2012.

[15] S. N. Flesher et al., "A brain–computer interface that evokes tactile sensations improves robotic arm control," *Science*, vol. 372, no. 6544, pp. 831–836, May 2021.

[16] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface," *J. Neural Eng.*, vol. 10, no. 4, p. 046003, Aug. 2013.

[17] C. Escolano, J. M. Antelis, and J. Minguez, "A telepresence mobile robot controlled with a noninvasive brain–computer interface," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 42, no. 3, pp. 793–804, Jun. 2012.

[18] R. Fukuma et al., "Real-time control of a neuroprosthetic hand by magnetoencephalographic signals from paralysed patients," *Sci. Rep.*, vol. 6, no. 1, pp. 1–14, Feb. 2016.

[19] C. J. Bell, P. Shenoy, R. Chalodhorn, and R. P. Rao, "Control of a humanoid robot by a noninvasive brain–computer interface in humans," *J. neural Eng.*, vol. 5, no. 2, p. 214, 2008.

[20] Y. Chae, J. Jeong, and S. Jo, "Toward brain-actuated humanoid robots: Asynchronous direct control using an EEG-based BCI," *IEEE Trans. Robot.*, vol. 28, no. 5, pp. 1131–1144, Oct. 2012.

[21] C. Zhang, Y. Kimura, H. Higashi, and T. Tanaka, "A simple platform of brain-controlled mobile robot and its implementation by SSVEP," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–7.

[22] D. Zhang et al., "Brain-controlled 2D navigation robot based on a spatial gradient controller and predictive environmental coordinator," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 12, pp. 6138–6149, Dec. 2022.

[23] H. Li, L. Bi, and J. Yi, "Sliding-mode nonlinear predictive control of brain-controlled mobile robots," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 5419–5431, Jun. 2022.

[24] R. Chai, S. H. Ling, G. P. Hunter, Y. Tran, and H. T. Nguyen, "Brain–computer interface classifier for wheelchair commands using neural network with fuzzy particle swarm optimization," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 5, pp. 1614–1624, Sep. 2014.

[25] X. Chen, B. Zhao, Y. Wang, S. Xu, and X. Gao, "Control of a 7-DOF robotic arm system with an SSVEP-based BCI," *Int. J. Neural Syst.*, vol. 28, no. 8, Oct. 2018, Art. no. 1850018.

[26] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Non-invasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Sci. Rep.*, vol. 6, no. 1, p. 38565, Dec. 2016.

[27] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain–computer interface with high transfer rates," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1181–1186, Oct. 2002.

[28] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, and T.-P. Jung, "A high-speed brain speller using steady-state visual evoked potentials," *Int. J. Neural Syst.*, vol. 24, no. 6, Sep. 2014, Art. no. 1450019.

[29] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, Nov. 2015.

[30] G. R. Müller-Putz, R. Scherer, C. Brauneis, and G. Pfurtscheller, "Steady-state visual evoked potential (SSVEP)-based communication: Impact of harmonic frequency components," *J. Neural Eng.*, vol. 2, no. 4, pp. 123–130, Dec. 2005.

[31] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2610–2614, Dec. 2006.

[32] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, Jan. 2018.

[33] A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 741–753, Apr. 2011.

[34] J. Ma, Y. Zhang, A. Cichocki, and F. Matsuno, "A novel EOG/EEG hybrid human–machine interface adopting eye movements and ERPs: Application to robot control," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 876–889, Mar. 2015.

[35] S. He and Y. Li, "A single-channel EOG-based speller," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1978–1987, Nov. 2017.

[36] Y. Zhou, S. He, Q. Huang, and Y. Li, "A hybrid asynchronous brain-computer interface combining SSVEP and EOG signals," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 10, pp. 2881–2892, Oct. 2020.

[37] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "SSVEP recognition using common feature analysis in brain–computer interface," *J. Neurosci. Methods*, vol. 244, pp. 8–15, Apr. 2015.

[38] M. Nakanishi, Y. Wang, Y.-T. Wang, and T.-P. Jung, "A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials," *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0140703.

[39] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface," *J. Neural Eng.*, vol. 12, no. 4, Aug. 2015, Art. no. 046008.

[40] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, NY, USA: Wiley, 1962.

[41] Y. Li, J. Pan, F. Wang, and Z. Yu, "A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3156–3166, Nov. 2013.

[42] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, "An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method," *J. Neural Eng.*, vol. 6, no. 4, Aug. 2009, Art. no. 046002.

[43] J. R. Millan and J. Mourino, "Asynchronous BCI and local neural classifiers: An overview of the adaptive brain interface project," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 159–161, Jun. 2003.

[44] X. Chen, B. Zhao, Y. Wang, and X. Gao, "Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026012.

[45] A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using EEG signals," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 6570–6577.

[46] C. Yang, H. Wu, Z. Li, W. He, N. Wang, and C.-Y. Su, "Mind control of a robotic arm with visual fusion technology," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3822–3830, Sep. 2018.

[47] C. I. Penaloza and S. Nishio, "BMI control of a third arm for multitasking," *Sci. Robot.*, vol. 3, no. 20, pp. 1–12, Jul. 2018.

[48] L. Chen et al., "Adaptive asynchronous control system of robotic arm based on augmented reality-assisted brain–computer interface," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066005.

[49] B. Fang et al., "Brain–computer interface integrated with augmented reality for human–robot interaction," *IEEE Trans. Cognit. Develop. Syst.*, early access, Jul. 28, 2022, doi: 10.1109/TCDS.2022.3194603.

[50] B. J. Edelman et al., "Noninvasive neuroimaging enhances continuous neural tracking for robotic device control," *Sci. Robot.*, vol. 4, no. 31, Jun. 2019, Art. no. eaaw6844.

[51] Y. Xu et al., "Shared control of a robotic arm using non-invasive brain–computer interface and computer vision guidance," *Robot. Auto. Syst.*, vol. 115, pp. 121–129, May 2019.

[52] R. Liu, Y.-X. Wang, and L. Zhang, "An FDES-based shared control method for asynchronous brain-actuated robot," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1452–1462, Jun. 2016.

[53] L. Cao, G. Li, Y. Xu, H. Zhang, X. Shu, and D. Zhang, "A brain-actuated robotic arm system using non-invasive hybrid brain–computer interface and shared control strategy," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 046045.

[54] X. Chen, X. Huang, Y. Wang, and X. Gao, "Combination of augmented reality based brain–computer interface and computer vision for high-level control of a robotic arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 3140–3147, Dec. 2020.

[55] M. Eitrheim and A. Fernandes, "The NASA task load index for rating workload acceptability," in *Proc. Hum. Factors User Needs Transp., Control, Workplace-HFES-Europe Annu. Meeting*, 2016, pp. 26–28.

[56] P. T. McCabe, *Contemporary Ergonomics 2003*. Boca Raton, FL, USA: CRC Press, 2002.

[57] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Apr. 1988.