

Extended Control With Hybrid Gaze-BCI for Multi-Robot System Under Hands-Occupied Dual-Tasking

Hong Zeng¹, Member, IEEE, Yitao Shen, Dengfeng Sun, Xuhui Hu², Member, IEEE, Pengcheng Wen, and Aiguo Song³, Senior Member, IEEE

Abstract—Currently there still remains a critical need of human involvements for multi-robot system (MRS) to successfully perform their missions in real-world applications, and the hand-controller has been commonly used for the operator to input MRS control commands. However, in more challenging scenarios involving concurrent MRS control and system monitoring tasks, where the operator's both hands are busy, the hand-controller alone is inadequate for effective human-MRS interaction. To this end, our study takes a first step toward a multimodal interface by extending the hand-controller with a hands-free input based on gaze and brain-computer interface (BCI), i.e., a hybrid gaze-BCI. Specifically, the velocity control function is still designated to the hand-controller that excels at inputting continuous velocity commands for MRS, while the formation control function is realized with a more intuitive hybrid gaze-BCI, rather than with the hand-controller via a less natural mapping. In a dual-task experimental paradigm that simulated the hands-occupied manipulation condition in real-world applications, operators achieved improved performance for controlling simulated MRS (average formation inputting accuracy increases 3%, average finishing time decreases 5 s), reduced cognitive load (average reaction time for secondary task decreases 0.32 s) and perceived workload (average rating score decreases 15.84) with the hand-controller extended by the hybrid gaze-BCI, over those with the hand-controller alone. These findings reveal the potential of the hands-free hybrid gaze-BCI to extend the traditional manual MRS input devices for creating a more operator-friendly interface, in challenging hands-occupied dual-tasking scenarios.

Index Terms—Gaze tracking, electroencephalograph, brain-computer interface, multi-robot system, dual-tasking.

Manuscript received 12 July 2022; revised 8 November 2022 and 3 December 2022; accepted 4 January 2023. Date of publication 9 January 2023; date of current version 3 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62173089, Grant 61673105, and Grant 92148205; and in part by the Aeronautical Science Foundation of China under Grant 20171969007. (Corresponding author: Hong Zeng.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Southeast University Ethics Committee under Application No. 2019ZDSYLL001-P01, and performed in line with the Declaration of Helsinki.

Hong Zeng, Yitao Shen, Dengfeng Sun, Xuhui Hu, and Aiguo Song are with the State Key Laboratory of Bioelectronics and the Jiangsu Province Key Laboratory of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: hzeng@seu.edu.cn).

Pengcheng Wen is with the AVIC Aeronautics Computing Technique Research Institute, Xi'an 710068, China (e-mail: wpcheng@avic.com).

Digital Object Identifier 10.1109/TNSRE.2023.3234971

I. INTRODUCTION

EVEN with rapid advances in the autonomous multi-robot system (MRS), many of the real-world robotic missions still cannot be reliably assigned to fully autonomous mobile robots [1]. For instance, when employing MRS as a means of gaining essential information for supporting follow-up firefighting operation, current MRSs are limited in their world awareness and cognitive capabilities for handling complex and dynamic incidents. The human-in-the-loop MRS operation is a tenable way to overcome this limitation [2], which involves dynamic authoritative control of specific MRS activities based upon local circumstances and human expertise, as well as concurrent monitoring of mission execution and MRS statuses. To realize remote control and monitoring of MRS, a screen usually displays the MRS in the environment and critical information collected from sensors mounted on MRS, a haptic hand-controller is commonly adopted where an operator holds it with dominant hand for inputting velocity and formation commands to MRS (the haptic feedback could also be provided with the hand-controller), while the non-dominant hand is also bounded by, e.g., the operating manual, the keyboard for dealing with abnormal alarms from the monitoring program, etc. It has been shown by Kruijff et al. [3] that manipulating robots in a disaster scene is highly cognitive-demanding and difficult for operators. One of the main factors increasing the cognitive-load is that the commonly adopted user interface alone may be inadequate for such challenging dual-task scenarios [4]. In specific, although the haptic hand-controller excels at specifying continuous velocity commands for MRS, it is poor at issuing discrete MRS formation commands that are usually derived with a less natural mapping from its continuous-valued inputs. As a consequence, it may introduce extra cognitive load that prevents effective and efficient executing the mission [5].

To overcome the above limitation of current user interface for the considered hands-occupied dual-task (i.e., concurrent controlling and monitoring of MRS) scene, it is critical to provide effective supplementary inputs. On one hand, the operator's non-dominant hand that is occupied may be too busy to offer extra inputs. On the other hand, although additional regular hands-on user interfaces (e.g., mouse, keyboard, touchscreen) could provide intuitive inputs for discrete target selection, frequent switches between holding

the hand-controller and touching such hands-on interfaces by dominant hand could easily foster human errors and slow down the work pace as well [6]. In contrast to solutions above, hands-free interfaces may provide a more viable means for supplementing manual control. Considering the teleoperation of MRS in a noisy disaster scene, it requires the visual-motor integration, i.e., the operator's gazes being directed at the screen for inspections and the hands being busy at fine motor control movement, thus we only focus on the non-verbal attention-aware hands-free interactions based on brain or gaze signals.

Harnessing the gaze to infer the operator's intended on-screen commands promises to be a natural control interface for human-MRS interactions [7], [8], [9], [10]. Nevertheless, accuracy of gaze-based input is found to largely depend on the target size [11]. In real-world practice, it is difficult to design an optimal command icon size for gaze-based input due to the variations in subjects, screen sizes and etc., making the gaze modality alone fail to be reliable enough to acquire user's intended command for our study. The brain-computer interface (BCI) allows the interaction between a user and a machine by the cerebral activity recorded from the electrodes on the scalp, e.g., Electroencephalograph (EEG) signals. In particular, BCI based on the steady-state visual evoked potential (SSVEP) requires the user's eyes to gaze at and pay attention to his/her desired target associated with a specific flickering stimulus, the cerebral pattern in response to the stimulation can be recognized with high accuracy and speed for multiple targets selection [12]. Though being intuitive to provide hands-free inputs, there are also many factors that may degrade the detection performance of SSVEP in real applications [13], such as changes in psycho-physiological states, inherent background noise in EEG signals and etc. Requiring a long fixation at the target may increase the evidences from gaze or EEG data and thus can improve the user's input accuracy [10], but the interaction efficiency would probably be significantly decreased.

All in all, neither gaze-based input alone nor SSVEP-based input alone is able to realize reliable and efficient command selections. Recently, there are several efforts toward exploring the hybrid gaze-BCI that combines the gaze and EEG input modalities [14], [15], [16], [17], [18], [19]. Typical hybrid gaze-BCI sequentially utilizes the gaze to navigate the cursor to the region containing potential targets of interest, and then after the dwell time elapses, triggers the BCI paradigm to confirm the selection [15], [17]. However, not only the ultimate confirmation solely depends on the possibly unreliable BCI [18], but also the input efficiency is low, since such a hybrid interface remains in idle before the dwell time for the gaze input and the time for recording and detecting specific cerebral patterns elapse.

In this work, we take a first step toward developing a novel hybrid gaze-BCI to supplement the commonly-adopted hand-controller in a challenging MRS operation scenario. Firstly, we present an effective solution to address the reliable and efficient sensor-processing issue in hands-free on-screen commands selection using hybrid gaze-BCI. Specifically, instead

of being sequentially processed, the gaze and SSVEP sensor data are simultaneously processed to issue fused selection decisions. Secondly, we assessed the impacts of the combined user interface (the proposed hybrid gaze-BCI issuing discrete MRS formation commands supplements to a hand-controller issuing continuous velocity commands for MRS) on MRS control performance, operator's cognitive load and perceived workload with respect to the hand-controller alone (issuing both formation and velocity commands), in the context of a hands-occupied dual-task mission (concurrent control of simulated MRS and system monitoring) in a disaster environment. In summary, the contribution of our paper is two-fold:

- We provide a collaborative gaze and BCI input by simultaneous processing and fusion, other than sequential combination as well as gaze/BCI alone. It may enable an intuitive hands-free interface for reliably and efficiently selecting on-screen commands to control MRS.
- The present study is to the best of our knowledge, the first to validate the feasibility and effectiveness of hybrid gaze-BCI as a hands-free interface supplementary to a manual one for improving the human-MRS interaction in a challenging scenario, i.e., the hands-occupied dual-tasking.

II. RELATED WORK

A. BCI/Gaze-Based Input Supplementing to Manual Input

BCI-based and gaze-based interfaces have been widely studied for recovery or replacement of a lost ability for the disabled [15], [20], [21]. Yet there has been little previous work dedicated to investigate such hands-free interfaces supplementary to conventional hands-on interfaces for augmenting control capability of healthy operators.

For users with motor impairments who can only employ EEG or electrooculogram (EOG) signals as the primary inputs for interaction with the environment, EEG-based BCIs and hybrid EEG/EOG-based BCIs with paradigms such as motor imagery (MI), event-related potentials (ERP) and so on have demonstrated their great values [15], [20], [21]. However, for a wider BCI population, e.g., healthy operators who mostly rely on normal manual controls, as pointed out by Xu et al. [22], BCI should be more appropriately used as an auxiliary control modality where it is more convenient or there are no other alternatives. Under such circumstances, the MI-based and ERP-based BCIs would suffer from significant performance degradation, due to the interference from the limb/hand movement, ocular, facial muscle related artifacts. By contrast, a large body of literature [23], [24], [25], [26], [27] have reported that the attention-aware SSVEP-based BCIs could not only produce high accuracy and efficiency for multi-target selection, but also perform more robustly with the above artifacts, than MI-based and ERP-based BCIs. Therefore, our work has also exploited the SSVEP paradigm to facilitate the BCI-based input.

The eye movement signals have also been capitalized on to realize hands-free intuitive interfaces for selecting the moving

targets [7], [8], [9] or still targets [15]. In particular, our study concentrates on selecting still targets, and the development of such gazed-based interface suffers from Midas Touch problem [10], i.e., not every location we fixate is related to the target. Early efforts for overcoming such a problem relies on detecting unnatural customized gaze “gestures” (such as blink timings, dwell times, or complex eye movement patterns), which may soon lead to user fatigue [10]. Recent studies have resorted to develop decoding algorithms to identify the intended target from natural gaze signals [28]. Our work further extends the existing work by fusing the accumulated evidences of target zone from natural gaze signals with those of target frequency from simultaneously recorded SSVEP signals.

B. BCI/Gaze-Based Input Under Dual-Tasking

In contrast to the single-task scene, the concurrent task in company with the one controlled by BCI or gaze in the dual-task paradigm would require the same cognitive resources simultaneously, so available resources have to be split and thus extra mental workload introduces to operators [29]. For this reason, the dual-task paradigm has been widely adopted [29], [30], [31] to simulate the highly cognitive-demanding operation scenarios in real world, where the performance measurements of the primary task reflect the operator’s manipulation ability under high cognitive load, and those of a secondary task provide objective information about the operator’s cognitive load. Most previous studies have investigated the BCI-based and gaze-based control performance in the single-task paradigm [7], [8], [9], [32], [33], [34], [35], whereas few efforts have been made in a more demanding dual-task paradigm where the operator’s both hands are involved until very recently.

The study closely related to our effort is [36], which investigates the BCI control of a robot for human augmentation under dual-tasking. In [36], the authors use goal-oriented action imagery EEG signals to extend healthy participants to control a supernumerary robotic limb for grasping/releasing an object, while simultaneously balancing a ball placed on a board held with their hands. Nevertheless, since both the imagined hand grasping/releasing and the actual hands balancing movement activate the same brain regions for such two tasks [37], the recognition accuracy of two-class goal-action imagery EEG signals has been greatly affected, half of the participants achieved poor dual-tasking performance. By contrast, our work utilizes the SSVEP which has been verified to be relatively robust to limb/hand movement, ocular, muscle related artifacts [23], [24], [25], [26], [27], enabling a better reliability. Besides, the SSVEP BCI could provide more commands than the goal-action imagery one, which is beneficial for controlling MRS of multiple freedoms.

III. HUMAN-MRS INTERACTION TEST-BED OVERVIEW

We developed a simulation platform of mobile grounded robots with Gazebo and robot operating system (ROS) on a PC running Ubuntu operating system, as a human-MRS

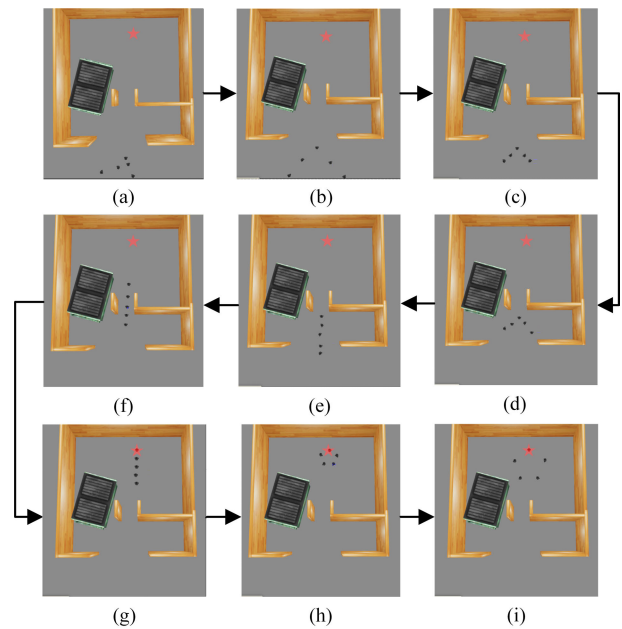


Fig. 1. The process for the MRS control task.

interaction test-bed. Following the dual-task paradigm, it consists of a primary task (MRS control) and a secondary task (system monitoring) illustrated below. The inputting interfaces for conducting primary and secondary tasks will be introduced as well.

A. Dual-Task Paradigm

1) *Primary Task*: The MRS control task is designed to simulate the application of employing MRS to enter a disaster scene, so as to gain essential information for supporting follow-up firefighting operation. It involves manipulating the MRS from the initial position to the target position, passing through a gate and a narrow passage on the half way. As depicted in Fig. 1, there are nine steps for the MRS control task, where the red pentagram denotes the target position. (1) In the beginning of each MRS control trial, five mobile robots appear in the same initial position as shown in Fig. 1(a); (2) A command for achieving a triangle formation should be issued, and the converged MRS formation is shown in Fig. 1(b); (3) A command for achieving a dense formation should be issued for avoiding collisions with the gate, and the resulting dense MRS is shown in Fig. 1(c); (4) The subject drives the MRS forward to pass the gate, by inputting velocity commands with the hand-controller (Fig. 1(d)) using the dominant hand; (5) A command for changing into the vertical line shape is issued before passing through the narrow passage ahead, Fig. 1(e) shows the MRS with the vertical line formation; (6) The subject drives the MRS forward to pass the narrow passage, using the hand-controller to input the velocity command (Fig. 1(f)); (7) The MRS is further driven to the target position by the subject with the hand-controller (Fig. 1(g)); (8) A command for achieving a pentagon formation should be issued, and the result is shown in Fig. 1(h); (9) A command for achieving a sparse formation

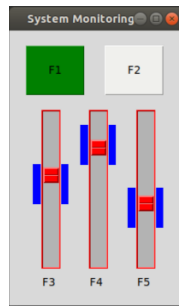


Fig. 2. The GUI for the system monitoring task.

is then issued, and the task ends (Fig. 1(i)). In particular, the velocity command and the formation command are required to be inputted asynchronously to ensure a safe passing through the gate and the narrow passage, and five different formation commands should be issued throughout the MRS control task without any cue.

2) *Secondary Task*: The system monitoring task is designed to simulate the practical scenario where the MRS operator has to consistently bind his/her attention on system health and function issues such as the fuel status, payload status, datalink status and etc. We implemented the off-the-shelf System Monitoring Task (SMT) from the NASA's Multi Attribute Test Battery (MATB) [38]. The GUI program for SMT was developed using Tkinter toolbox with Python, and it was displayed overlying the Gazebo simulator (Fig. 2). In this task, the subject is required to press a key on a keyboard with the non-dominant hand in response to an abnormal event. The abnormal events and responses are defined as follows: (1) The alarm lamp on the left turns from green to white (pressing F1); (2) The alarm lamp on the right turns from white to red (pressing F2); (3) The value for the item being monitored falls out of the normal range, i.e., the red bar is out of the blue zone in each slider (pressing F3, F4 or F5 for the corresponding slider). Concerning the rationale of the adopted SMT, three remarks are given below. First, the SMT is simple enough that it can be executed using the non-dominant hand simultaneously with the primary task without suppressing it. Second, both the primary and secondary tasks require participant's visual-motor integration and thus draw upon the same cognitive resources. Third, the SMT is suitable to assess cognitive load since key-pressing movements are less precise and reaction times become slower with increasing cognitive load [39].

B. Input Interfaces for Primary and Secondary Tasks

Fig. 3 provides an overview of the interfaces utilized in the dual-task paradigm. A multimodal input (MMI) is designated for the primary task (i.e., MRS control), where the operator's dominant hand manipulates the haptic hand-controller (Geomagic Touch by 3D Systems, Inc.) to produce the *velocity control inputs* for MRS, and he/she utilizes a hybrid gaze-BCI for providing *formation control inputs* for MRS. For the secondary task, to react to abnormal event visual alarms appearing on the system monitoring GUI, the operator uses the non-dominant hand to press corresponding keys (*system monitoring inputs*) on the keyboard.

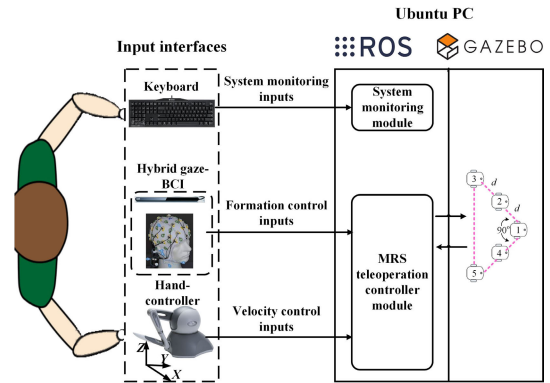


Fig. 3. Overview of the interfaces for primary and secondary tasks.

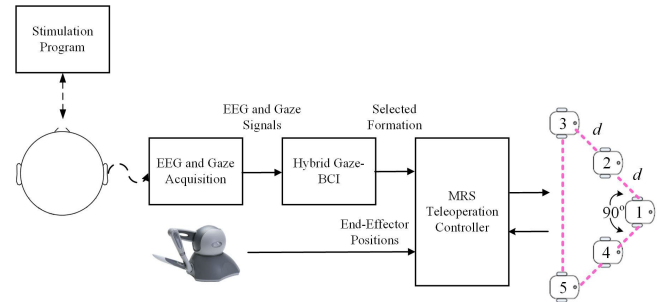


Fig. 4. The overall MRS control architecture.

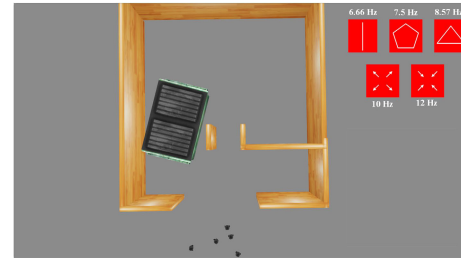


Fig. 5. The formation command items and associated flickering frequencies.

IV. METHODS

In this section, the details for MRS teleoperation control with hand-controller and hybrid gaze-BCI will be provided. Fig. 4 depicts the overall MRS control architecture, consisting of stimulation program, EEG and gaze acquisition, hybrid gaze-BCI and MRS teleoperation controller that are illustrated below.

A. Stimulation Program

In our study, we defined altogether $K = 5$ commands for MRS formation control, including 3 commands for formation shapes (vertical line, pentagon, triangle) and 2 commands for formation densities (dense, sparse). Such five commands were displayed with image icons overlying the Gazebo simulator (see Fig. 5), displaying on the LCD screen (24.5 inch, 1920×1080 pixels, refresh rate of 60 Hz).

Regarding the icon size for gaze-based interaction, a too large command icon would occlude the display of the scene and the MRS, while a too small one would probably require great user efforts in placing gaze points precisely within a small region. In this study, the icon size was designed

following the rule suggested in [11]. Specifically, assuming that the captured gaze points were normally distributed in the x and y directions around a mean with offset $o_{x/y}$ (gaze tracking accuracy) from the target center and a standard deviation $\sigma_{x/y}$ (gaze tracking precision), the target width and height were derived as: $T_{w/h} = 2(o_{x/y} + 2\sigma_{x/y})$, where $\sigma_{x/y}$ was multiplied with 2 in order to make about 95% of values lie in two standard deviations of the mean for normally distributed data, i.e., 95% of gaze points were supposed to fall inside the target region. According to [40], the eye-tracker adopted in our study (eyeTracker 5, Tobii, Stockholm, Sweden) strays on average 35 pixels ($o_{x/y}$) from the target and has a standard deviation of 18 pixels ($\sigma_{x/y}$). Ultimately, the size for each icon was designed to be 140×140 pixels with a 50-pixel gap between nearby ones.

To evoke SSVEP, each command icon was associated with a visual stimulation flickering at different frequencies (6.66 Hz, 7.5 Hz, 8.57 Hz, 10 Hz, 12 Hz) suggested by [41]. The flickering stimulus program was developed using OpenGL.

B. EEG and Gaze Acquisition

EEG Signals were acquired with a 10-20 montage active electrode cap (ActiCap, BrainAmp, BrainProducts, Munich, Germany) and an EEG amplifier (SynAmps II, Neuroscan, Compumedics, Victoria, Australia), with a sampling rate of 1000 Hz. Signals collected from 9 electrodes (O1, O2, OZ, PO5, PO3, POZ, PO4, PO6 and PZ) of interest above the visual cortex were referenced to the left mastoid and the ground electrode was placed on the forehead. All the electrodes impedance was kept below 10 k Ω during the experiments. The EEG amplifier applied a band-pass filter between 0.15 and 200 Hz as well as a 50 Hz notch filter on the signals, before sending them to the recording PC. The event triggers generated by the stimulation program on host PC were sent to the Neuroscan amplifier with the parallel port. To enable further processing on host PC, the EEG data and the event triggers in recording PC were transmitted to the host PC with LAN cable (Ethernet) TCP/IP in real-time (transmission delay <1ms) and then down-sampled to 250 Hz.

The Tobii eye-tracker 5 was attached to the bottom of the monitor of host PC, allowing moderate head movements. The eye-tracker provided the 2D gaze positions on the screen of host PC via USB at at 33 Hz. For facilitating the subsequent synchronization with EEG signals whose sampling rate was 250 Hz, the gaze data were then resampled to 30 Hz. Subjects were seated at 65 cm from the screen. Before the experiments, the eye-tracker's in-built 6-point calibration procedure was applied for each subject, lasting about 30 s.

C. Hybrid Gaze-BCI

This section illustrates the scheme for the collaborative gaze and BCI input, featuring simultaneous processing and fusion. Firstly, for each input modality, a probabilistic model is built to estimate the probability for each target that the user is trying to select. Next, these two evidences are fused at the decision level for inferring the target command.

1) *Evidence of Selected Command by BCI*: To facilitate the practical usage of hybrid gaze-BCI in MRS operation, we have realized a plug-and-play SSVEP-based BCI which does not require calibrations in advance for each user. In specific, we apply the Filter bank canonical correlation analysis (FBCCA) [42] to recognize the frequency of SSVEP.

Let $\mathbf{X} \in \mathbb{R}^{N_c \times N_t}$ denote an L s-long EEG epoch beginning with the stimulus onset, where N_c is the number of channels and N_t represents the number of instances within an epoch. The FBCCA approach involves three steps. The first step is to decompose the original EEG signals \mathbf{X} into sub-band components $\mathbf{X}_{SB_n} \in \mathbb{R}^{N_c \times N_t}$ ($n = 1, 2, \dots, N$), by applying N filters with different pass-bands. Previous study [42] has validated that sub-bands containing multiple harmonic bands with a high cut-off frequency at the upperbound frequency of SSVEP is an effective design for FBCCA. Thus in this work, $N = 5$ band-pass filters are utilized, whose bands are 6–80 Hz, 12–80 Hz, 18–80 Hz, 24–80 Hz and 30–80 Hz, respectively. The second step of FBCCA is to calculate the canonical correlation values ($\rho_k^n, k = 1, 2, \dots, K$) between the n th SSVEP sub-band components \mathbf{X}_{SB_n} and the k th reference signal of sine-cosine waves. The final step of FBCCA is to compute the evidence for the k th target stimulus T_k with the weighted square sum of the N correlation values corresponding to N sub-band components:

$$P_{BCI}(T_k) = \frac{\sum_{n=1}^N w_{SB}(n)(\rho_k^n)^2}{\sum_{k=1}^K \sum_{n=1}^N w_{SB}(n)(\rho_k^n)^2}, \quad (1)$$

where weights for the sub-band components are $w_{SB}(n) = n^{-1.25} + 0.25$, suggested by [42]. Considering the SSVEP-BCI modality only, the target which maximizes $P_{BCI}(T_k)$ is selected as its output.

2) *Evidence of Selected Command by Gaze*: Each gaze sample $\mathbf{g}_t = [g_t^x, g_t^y]^T \in \mathbb{R}^2$ is concatenated to form an L s-long gaze epoch $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{M_t}] \in \mathbb{R}^{2 \times M_t}$ beginning with the same time instance as the SSVEP flickering stimulus onset. Note that t is the index of gaze samples within an epoch and M_t is the total number of gaze samples of an epoch. We employ the Kalman filter to track the gaze position (g_t) and gaze velocity (\dot{g}_t) for denoising the eye tracking signal, as well as to build a probabilistic distribution of the user's focus on screen at t . The process and measurement equations of the Kalman filter are:

$$\xi_{t+1} = \begin{bmatrix} g_{t+1}^x \\ g_{t+1}^y \\ \dot{g}_{t+1}^x \\ \dot{g}_{t+1}^y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1/30 & 0 \\ 0 & 1 & 0 & 1/30 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xi_t + \omega_t \quad (2)$$

$$\mathbf{g}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \xi_t + \eta_t, \quad (3)$$

where ξ_t is the state vector, ω_t and η_t are assumed to be white, mutually independent Gaussian noise processes, the posterior state mean and covariance are $\hat{\xi}_{t|t}$, $\mathbf{Q}_{t|t} = \mathbb{E}\{[\xi_t - \hat{\xi}_{t|t}][\xi_t - \hat{\xi}_{t|t}]^H\}$, respectively. It is rational to assume that the user's gaze does not deviate far from the target, and thus the evidence, i.e., the posterior probability of the k th target given

the gaze epoch \mathbf{G} could be calculated by:

$$P_{gaze}(T_k) = \mathbb{E} \left\{ \int_{\mathcal{A}_k} \mathcal{N}(\xi_t; \hat{\xi}_{t|t}, \mathbf{Q}_{t|t}) d\xi_t \right\}, \quad (4)$$

where \mathcal{A}_k represents the zone on screen where target k appears, and the mean (\mathbb{E}) is over t . As for the gaze modality alone, the target which maximizes $P_{gaze}(T_k)$ is selected as the output.

3) *Gaze and BCI Fused Command Selection*: The synchronized EEG and gaze epochs of length L s are then fused at the decision level by applying the Dempster's rule of combination.

$$P_{fusion}(T_k) = \frac{1}{M} P_{BCI}(T_k) * P_{gaze}(T_k), \quad (5)$$

where $M = \sum_{k=1}^K P_{BCI}(T_k) * P_{gaze}(T_k)$ is the normalization constant. The target that maximizes $P_{fusion}(T_k)$ is the consensus decision from the EEG and gaze input modalities.

D. MRS Teleoperation Controller

This section presents the teleoperation architecture with a distributed multi-robot control algorithm. The hand-controller and the hybrid gaze-BCI reside on the master side, simulated multiple mobile robots are slaves.

1) *Kinematic Model of Mobile Robot*: We consider 5 simulated two-wheeled differentially driven mobile robots whose coordinates are denoted by $\mathbf{P}_i = [x_i, y_i]^T \in \mathbb{R}^2$, $i = 1, 2, \dots, 5$, where (x_i, y_i) is the coordinate of i -th robot in the world coordinate frame $X_W O_W Y_W$. $\mathbf{v}_i = [v_{xi}, v_{yi}]^T \in \mathbb{R}^2$ and $\tilde{\mathbf{v}}_i = [\tilde{v}_{xi}, \tilde{v}_{yi}]^T \in \mathbb{R}^2$ represent the velocity vectors of the robot in the world frame $X_W O_W Y_W$ and the robot frame $X_R O_R Y_R$, respectively. Assuming no wheel slipping occurs, the kinematic model of the nonholonomic wheeled robot can be written as:

$$\dot{\mathbf{P}}_i^R = [\cos \theta_i, \sin \theta_i, 0]^T v_i + [0, 0, 1]^T w_i \quad (6)$$

where $\mathbf{P}_i^R = [\mathbf{P}_i^T, \theta_i]^T$, θ_i denotes robot's orientation angle with respect to the x -axis ($-\pi < \theta_i \leq \pi$), v_i and w_i denote the wheel's linear velocity and its angular velocity, respectively. v_i, w_i could be derived from \mathbf{v}_i and $\tilde{\mathbf{v}}_i$ as follows:

$$\begin{cases} v_i = \sqrt{(v_{xi})^2 + (v_{yi})^2} \\ \tilde{\mathbf{v}}_i = \mathbf{R}(\theta_i) \mathbf{v}_i \\ w_i = \text{atan2}(\tilde{v}_{yi}, \tilde{v}_{xi}) \\ \dot{\theta}_i = w_i \end{cases} \quad (7)$$

where $\mathbf{R}(\theta_i)$ is the rotation matrix from $X_W O_W Y_W$ to $X_R O_R Y_R$, as shown in Fig. 6.

2) *Formation Library*: The multiple robots may need to form appropriate formations according to the situations. For each formation shape, the desired relative distances between robots are predefined. For example, the matrix storing the relative distances between robots with the triangle formation (Fig. 3)

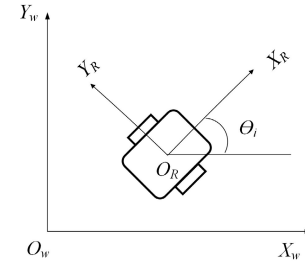


Fig. 6. The world frame and robot frame.

is given by:

$$\mathbf{D} = [d_{ij}]_{5 \times 5} = \begin{bmatrix} 0 & d & 2d & d & 2d \\ d & 0 & d & \sqrt{2}d & \sqrt{5}d \\ 2d & d & 0 & \sqrt{5}d & 2\sqrt{2}d \\ d & \sqrt{2}d & \sqrt{5}d & 0 & d \\ 2d & \sqrt{5}d & 2\sqrt{2}d & d & 0 \end{bmatrix}, \quad (8)$$

where d_{ij} denotes the desired relative distance between robot i and j . d is the minimal distance between robots, determining the density of the formation (two positive values for d are used in our study to represent a dense formation and a sparse formation, respectively). The participant could use the hybrid gaze-BCI to select the intended formation command icon appearing on the screen.

3) *Distributed Multi-Robot Control*: We have adopted the distributed MRS control algorithm presented in [43] to implement the MRS control. The distributed multi-robot control consists of the following three possible control items: (1) a velocity control of the robot using a hand-controller device $\mathbf{v}_i^h = [v_{xi}^h, v_{yi}^h]^T \in \mathbb{R}^2$, (2) a formation control to form the desired formation $\mathbf{v}_i^f = [v_{xi}^f, v_{yi}^f]^T \in \mathbb{R}^2$, (3) a collision avoidance control to avoid obstacles $\mathbf{v}_i^o = [v_{xi}^o, v_{yi}^o]^T \in \mathbb{R}^2$. Subsequently, we implement the following distributed control on each robot, for the i th robot

$$\dot{\mathbf{P}}_i = \mathbf{v}_i^h + \mathbf{v}_i^f + \mathbf{v}_i^o. \quad (9)$$

a) *Velocity control*: \mathbf{v}_i^h . The desired velocity input of the robot is given by $v_{xi}^h = K_{hx} I(q_x, q_{x0})$, $v_{yi}^h = K_{hy} I(q_y, q_{y0})$, where q_x and q_y denote the end-effector positions of the hand-controller along x axis and y axis (Fig. 3), respectively. $I(q, q_0)$ is a function defined by:

$$I(q, q_0) = \begin{cases} 1 & q > q_0, \\ -1 & q < -q_0, \\ 0 & \text{otherwise}, \end{cases} \quad (10)$$

where q_0 is a positive threshold constant. K_{hx} and K_{hy} are constant speed values.

b) *Formation control*: \mathbf{v}_i^f . It implies the control item to avoid collisions among robots and to preserve the desired relative distance d_{ij} between robot i and robot j , for achieving a desired formation, as defined by:

$$\mathbf{v}_i^f = - \sum_{j=1}^N \frac{\partial \phi_{ij}^f (\|\mathbf{P}_i - \mathbf{P}_j\|^2)^T}{\partial \mathbf{P}_i}, \quad (11)$$

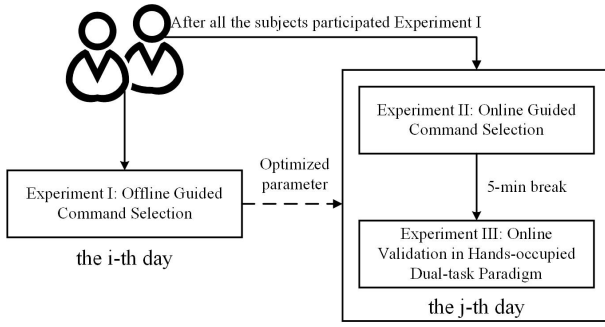


Fig. 7. Overview of the three experiments.

where ϕ_{ij}^f is an artificial potential function to generate an attractive action if $\|\mathbf{P}_i - \mathbf{P}_j\| > d_{ij}$, a repulsive action if $\|\mathbf{P}_i - \mathbf{P}_j\| < d_{ij}$ and a null action if $\|\mathbf{P}_i - \mathbf{P}_j\| = d_{ij}$ as in [43].

c) *Collision avoidance control*: \mathbf{v}_i^o . Such a control item is again defined by an artificial potential field that prevent MRS to collide with obstacles under a distance threshold $d_0 \in \mathbb{R}^+$:

$$\mathbf{v}_i^o = - \sum_{j \in \mathcal{O}_i} \frac{\partial \phi_{ij}^o(\|\mathbf{P}_i - \mathbf{P}_j^o\|)^\top}{\partial \mathbf{P}_i}, \quad (12)$$

where \mathcal{O}_i represents the set of obstacles of the i th robot with \mathbf{P}_j^o being the position of the j th obstacle in \mathcal{O}_i . The artificial potential field function ϕ_{ij}^o as in [43] produces a repulsive action if $\|\mathbf{P}_i - \mathbf{P}_j^o\| < d_0$ and a null action otherwise.

In summary, each robot is controlled by the distributed MRS control (equation (9)), and it corresponds to the summation of the three control items. Then the linear velocity and the angular velocity of each wheeled robot could be derived according to equation (7), for driving the robot via equation (6) in Gazebo.

V. EXPERIMENTS

A. Subjects

Eight healthy subjects (2 females, 6 males, aged 22-26 years, all right-handed, with normal or corrected-to-normal vision) were recruited. All of them were naive to SSVEP experiments, two of them had eye tracking experience. The study was approved by the Southeast University Ethics Committee (2019ZDSYLL001-P01) and carried out in accordance with Declaration of Helsinki. Fig. 7 provides an overview of the experiments. Each subject participated in Experiment I, II and III (illustrated in following sub-sections) after another subject. After all the eight subjects had took part in Experiment I, the subject came back to our lab again on another day to conduct Experiment II and III in sequence with a 5-min break between them. Experiment I and II focused on the guided command selection task with MRS remained still, while Experiment III involved the concurrent MRS control and monitoring tasks. All the subjects had a good rest (e.g., a whole-night sleep or a nap after lunch) before the experiments.

B. Experiment I: Offline Guided Command Selection

The purpose of Experiment I was to fine-tune the epoch length of synchronized EEG and gaze for the hybrid gaze-BCI in a guided command selection task.

1) *Experimental Procedure*: Experiment I was separated into 30 blocks. Each block consisted of 5 trials, which contained the five formation control commands presented in a random order, resulting in altogether 150 trials. Each trial began with a visual cue indicating a target command (i.e., highlighted with white square frame) and the cue lasted for 1 s. Subjects were asked to shift their gaze to the target as soon as possible within the cue duration. Upon the cue offset, all the stimuli flickered simultaneously for 3 s. Subjects were asked to avoid eye blinks during the stimulation period. Following the stimulus offset, subjects had a short break for 2 s. To avoid visual fatigue, there was a 10-second rest between two consecutive blocks.

2) *Performance Metrics and Statistical Analysis*: The classification accuracy (ACCRY) and information transfer rate (ITR) were utilized for the performance evaluation. The ITR in bits/min was defined as Wolpaw et al. [44]:

$$ITR = (\log_2 K + p \log_2 p + (1 - p) \log_2 \left[\frac{1 - p}{K - 1} \right]) * \frac{60}{T}, \quad (13)$$

where K is the number of targets ($K = 5$ in our study), p is the classification accuracy and T is the duration per selection (including the epoch length for identifying the target and the 1-sec cue duration for gaze shifting). In experiment I, these performance metrics were estimated for different epoch lengths (from 0.3 s to 2.9 s with a step of 0.2 s), so as to determine the optimal epoch length for hybrid gaze-BCI. The Friedman test was conducted to assess the effect of gaze-based, BCI-based and hybrid gaze-BCI based approaches on ACCRY and ITR, post-hoc analysis was conducted with the Bonferroni correction.

C. Experiment II: Online Guided Command Selection

The experiment II was to evaluate the optimized hybrid gaze-BCI for online guided command selection.

1) *Experimental Procedure*: Experiment II consisted of five blocks, each block included also 5 trials corresponding to five commands in random order. The trial timing in Experiment I and II were identical except for the length of the stimulation flickering duration (determining the epoch length for synchronized EEG and gaze), which was optimized according to the results of Experiment I on all the subjects.

2) *Performance Metrics and Statistical Analysis*: The same performance metrics (ACCRY and ITR) were adopted in Experiment II as those of Experiment I, but calculated with optimized constant epoch length for all the subjects. The statistical analysis means and procedure were also the same as in Experiment I.

D. Experiment III: Online Validation in Hands-Occupied Dual-Task Paradigm

The purpose of Experiment III was to validate the effectiveness of the proposed hybrid gaze-BCI supplementing to the manual hand-controller with a dual-task paradigm, where both hands of the operator were occupied.

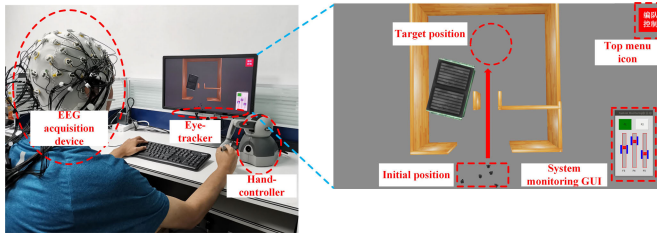


Fig. 8. The “Hand-CTR + HGBCI” experimental setup for dual-task paradigm experiment.

1) *Experimental Conditions and Procedure: Experimental conditions.* The primary task (MRS control) is conducted with the following two conditions:

- (1) *Hand-CTR + HGBCI*, where a subject inputs the velocity commands for MRS with the hand-controller (Hand-CTR) by the dominant hand and issues the formation commands with the supplementary hands-free hybrid gaze-BCI (HGBCI).
- (2) *Hand-CTR*, where a subject issues both the formation and velocity commands for MRS with the hand-controller only, using the dominant hand.

In the meanwhile, for both the two conditions, the secondary task (system monitoring) is handled with keyboard by the non-dominant hand. Such two concurrent tasks leave both hands busy.

2) *Experimental Setup for “Hand-CTR + HGBCI” Condition:* Fig. 8 shows the experimental setup. When inputting the velocity with the hand-controller, it is confined to the end-effector’s XOY frame (see Fig. 3). The target formation is selected with the hybrid gaze-BCI, when it has detected the user intended formation command icon.

Moreover, since it is not necessary to display the formation command menu throughout the task, we designed an appearing and flickering scheme for the formation command menu. Specifically, whenever the subjects intended to issue the formation commands, they had to gaze the top menu icon in the upper right corner (see Fig. 8) for at least 1 s, in order to trigger the display of the formation command menu (it is analogous to click the “start” menu residing on the bottom left corner on a Windows PC with a mouse). The top menu icon then disappeared, followed by the intermediate appearance and flickering of the formation command menu (Fig. 5), the subject was then required to promptly orient the visual attention onto the desired command icon and keep focusing on it. At the same time, the gaze and the EEG data were acquired simultaneously. Once the epoch length of the synchronized gaze and EEG data reached a predefined value, the subject’s intended command was inferred and then issued. Meanwhile, the formation command menu disappeared and the top menu icon reappeared in the upper right corner.

3) *Experimental Setup for “Hand-CTR” Condition:* In such an experimental setup, the inputting for the velocity control was also achieved by the dominant hand with the hand-controller. To input the formation control commands, a widely adopted approach [45], [46] was implemented. Specifically, as shown in Fig. 9, the formation command was constrained to be inputted with the dominant hand from the YOZ frame

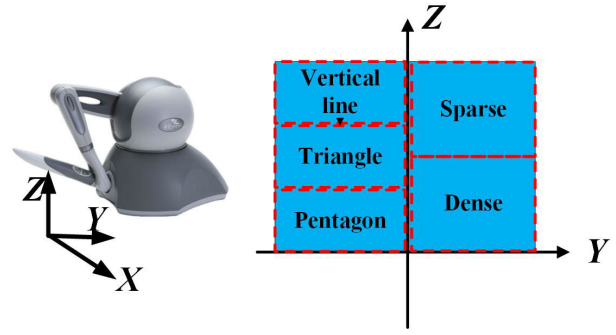


Fig. 9. The scheme for selecting formation commands in the “Hand-CTR” condition.

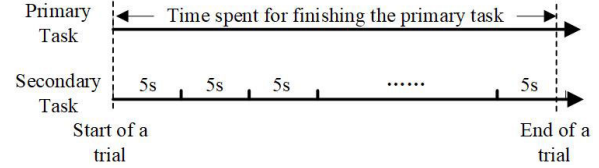


Fig. 10. The trial timing diagram for the dual-task paradigm.

TABLE I
OVERVIEW OF MEASUREMENTS FOR EXPERIMENT III

	Construct	Measured by
Objective Primary Task Metrics		(1) accuracy for inputting formation commands (ACC_{in})
		(2) time spent for finishing task ($TIME_{fi}$)
Objective Cognitive Load		(1) response accuracy (ACC_{resp})
		(2) response time ($TIME_{resp}$)
Subjective Workload		workload rating score (NASA-TLX)

of the hand-controller, where five non-overlapping zones were divided and a 1.5 s-long stay of the end-effector in each zone would trigger a formation command.

4) *Experimental Procedure:* Fig. 10 shows the trial timing diagram for the dual-task paradigm. The concurrent system monitoring and MRS control tasks began at the same time for each trial. There was no time limit for accomplishing the MRS control task in each trial, but there was a random abnormal event every 5 s for the system monitoring task. A trial ended when the MRS control task was finished. Each subject carried out 30 trials under the two conditions (15 trials per condition) in a counterbalanced order, and had a short rest between two trials. Before the formal experiments, the subjects took a practice session to get familiar with the interface usage and the task workflow (lasting about 5 minutes). Since there was no time limit for accomplishing the MRS control task and the formation command could be re-issued when the subject found that the command issued was not the desired one, all the subjects finished the MRS control task successfully.

5) *Performance Metrics and Statistical Analysis:* Table I provides an overview of all measurements for assessing the two conditions, which we explain in detail below.

Objective Primary Task Metrics. The performance of the MRS control task was assessed with the following two metrics. (1) The accuracy for inputting the formation commands ($ACC_{in} = \frac{75}{75+N_{err}}$), where $75 = 15 \text{ trials} \times 5 \text{ formation commands/trial}$ required to issue, and N_{err} represents total number of formation commands re-issued in all the trials.

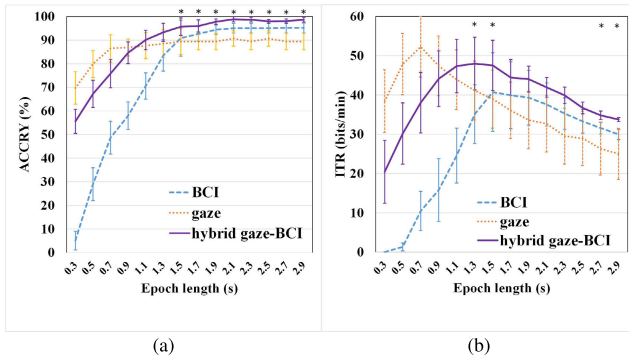


Fig. 11. Offline guided command selection experimental results. (a) ACCRY and (b) ITR for gaze only, BCI only and fusion based hybrid gaze-BCI approaches using different epoch lengths. When the performance of hybrid gaze-BCI is better than the other two approaches and the post-hoc analysis shows the performance difference is of statistical significance ($p < 0.05$), such a difference is marked by “**”.

This metric was defined following [32]. The true online classification accuracy for hybrid gaze-BCI may be higher than this estimate because subjects possibly did not follow the optimal strategy for completing the task. (2) The time spent for finishing the task ($TIME_{fi}$) in a trial. The initial and ending positions were fixed for the MRS control task, and the movement speeds of MRS with converged formations were the same across trials. Thereby, the completion time was dominated by the time spent on obtaining the required MRS formations.

Objective Cognitive Load. The objective measurement for cognitive load was the performance in the secondary task. It was assessed by the following two metrics that are well-identified in previous work, similar to physiological activity measurements [31]. (1) The response accuracy (ACC_{resp}): the percentage of correctly pressed keys in response to the random abnormal events out of all the trials. (2) The response time ($TIME_{resp}$): the time taken from the abnormal event onset to the key press.

Subjective Workload. At the end of experiments, subjects were asked to report their subjective mental workload rating scale score (0 ~ 100) with the NASA Task Load Index (NASA-TLX) tool, including mental demand, physical demand, temporal demand, effort, performance and frustration level (the average score was reported for such six scale factors).

The Wilcoxon signed rank test was utilized for the statistical analysis of the difference in measurements above between two conditions.

VI. RESULTS

A. Results for Offline Guided Command Selection

The offline performance of approaches based on BCI only, gaze only and hybrid gaze-BCI are reported, respectively. Fig. 11 depicts the performance as a function of EEG and gaze epoch length (from 0.3 s to 2.9 s with a step of 0.2 s) across trials and subjects.

From Fig. 11(a), it can be observed that the gaze alone results in higher ACCRY than other two approaches with short epochs (< 0.9 s), which is explained as follows. First, for the menu layout in our study, gaze could reach a specified area in

a very short time (generally less than 1 s, the length of cue duration), i.e., the subject’s gaze generally had successfully switched to the target before the SSVEP flickering stimulus onset (the beginning of the gaze epoch to be processed latter). Second, since it takes a certain time for the brain to activate the SSVEP response to the flickering stimuli, the BCI fails to provide satisfactory ACCRY with such short epochs. As a consequence, the hybrid gaze-BCI based on fusion is inferior to the gaze alone based approaches in this case. With epochs longer than 0.9 s, ACCRY of gaze-based approaches barely increases as shown in Fig. 11(a), since gaze points are unable keep still but oscillate around or off the target. By contrast, as the epoch length increases, ACCRY of BCI continues to rise until the length reaches to 2.1 s (ACCRY: around 95%) and then stops increasing with longer epoch lengths. Moreover, the evidence fusion of BCI and gaze input modalities indeed improves over the single modality alone after the epoch length reaches to 1.1 s, as it has always obtained higher ACCRY than the other two since then. Moreover, when the epoch length is greater than 1.5 s, the post-hoc analysis shows that the ACCRY differences between hybrid gaze-BCI and gaze/BCI are all statistically different ($p < 0.05$).

From Fig. 11(b), it can be seen that although the gaze-based approach could reaches to the highest ITR among these three ones with 0.7-s long epochs, its ACCRY is less than 90% and the ITR has kept decreasing with longer epochs. On the contrary, with a longer epoch (1.3 – 1.5 s), the fusion based hybrid gaze-BCI has attained the highest ITR (1.3 s: 47.98 ± 6.70 , 1.5s: 47.52 ± 6.41 bits/min) among the three ones. The post-hoc analysis with Bonferroni correction demonstrates that epoch length 1.3 s and 1.5 s lead to statistically different ITR between hybrid-BCI and gaze/BCI alone ($p < 0.05$).

Based on the results for the offline experiment, to obtain a high ACCRY and a high ITR simultaneously, the epoch lengths in the latter online guided command selection experiment and dual-task paradigm one were both set to 1.5 s.

B. Results for Online Guided Command Selection

Friedman test showed that different approaches have effects on both ACCRY and ITR, and post-hoc analysis with Bonferroni correction reveals that differences in ACCRY and ITR between hybrid gaze-BCI (ACCRY: $96.00 \pm 2.57\%$, ITR: 47.35 ± 6.82 bits/min) and gaze only technique (ACCRY: $89.41 \pm 3.51\%$, ITR: 38.95 ± 7.41 bits/min), as well as those between hybrid gaze-BCI and BCI only one (ACCRY: $90.83 \pm 3.04\%$, ITR: 40.72 ± 8.53 bits/min), are statistically significant (hybrid vs. gaze: $p < 0.05$, hybrid vs. BCI: $p < 0.05$). Such results indicate that the hybrid gaze-BCI with 1.5-second long gaze and EEG epochs is more accurate and efficient to infer the target than the gaze alone and the BCI alone, leaving reliable and efficient selection of commands feasible.

C. Results for Hands-Occupied Dual-Task Paradigm

The measurements across subjects and trials under two conditions are shown in Fig. 12.

As for the performance of the MRS control task demonstrated in Fig. 12 (a) and Fig. 12 (b), both ACC_{in} and $TIME_{fi}$ under the “Hand-CTR + HGBCI” setup outperform

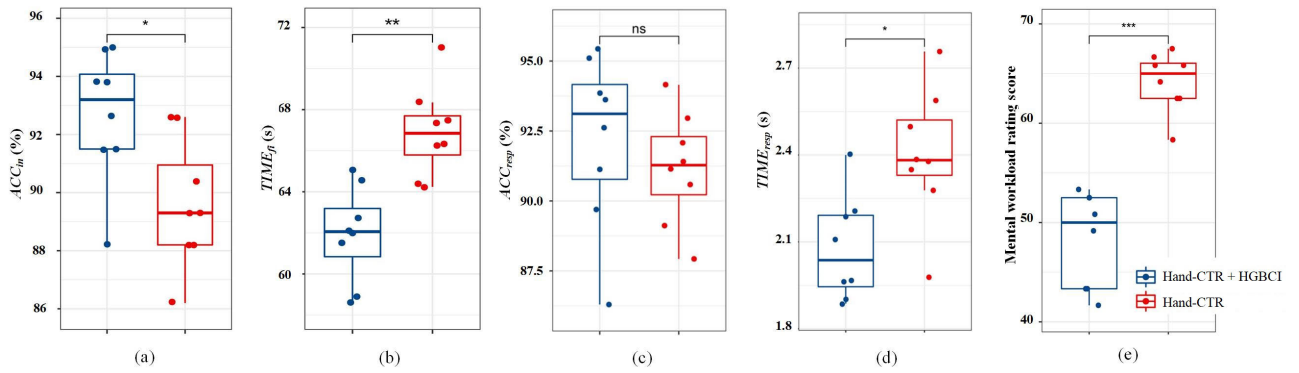


Fig. 12. Hands-occupied dual-task paradigm experimental results under two conditions. (a) the accuracy of inputting the formation commands (b) the time spent on finishing the MRS control task (c) the accuracy of pressed keys responding to abnormal events in the system monitoring task (d) the response time in the system monitoring task (e) the workload ratings for carrying out the dual tasks. The performance differences with statistical significance $p < 0.001$, $p < 0.01$ and $p < 0.05$ are marked by “***”, “**” and “*”, respectively, while non-significant difference is marked by “ns”.

those under the “Hand-CTR” one ($ACC_{in}: 92.66 \pm 2.26\%$ vs. $89.60 \pm 2.21\%$, $TIME_{fi}: 61.93 \pm 2.33$ s vs. 66.93 ± 2.20 s). Moreover, both performance differences between such two conditions are statistically significant ($ACC_{in}: p = 0.0313$ and $TIME_{fi} : p = 0.0019$). These results indicate that the inputting manner for formation commands with the hand-controller is inferior to that with the hybrid gaze-BCI, which is explained as follows. It may be difficult for the subject to directly move the end-effector of the hand-controller into the precise zone for issuing a desired formation command, the “trial and error” strategy was likely to be adopted. Namely, the subject randomly moved the end-effector to a zone to trigger a formation command, and then made a decision on the correctness of such a command with the visual feedback of the on-going MRS formation transforms; once the command was found to be the undesired one, the subject continued to move the end-effector into another zone, and such process repeated until the desired formation was obtained. By contrast, under the “Hand-CTR + HGBCI” setup, the “trial and error” strategy was avoided by displaying the candidate formation command icons. In other words, the selection of formation command through the hybrid gaze-BCI may be much more intuitive, accurate and easier than that through the hand-controller.

Regarding the performance of the system monitoring task shown in Fig. 12 (c) and Fig. 12 (d), both conditions lead to satisfying ACC_{resp} (average $ACC_{resp} > 90\%$), and there is no statistical significance between the two experimental setups for ACC_{resp} ($p = 0.38$). Nevertheless, the $TIME_{resp}$ of “Hand-CTR + HGBCI” is superior to that of “Hand-CTR” with statistical significance (2.08 ± 0.03 vs. 2.40 ± 0.05 , $p = 0.01$). The longer response time under “Hand-CTR + HGBCI” may be attributed to the following reason. The only difference between the two conditions was in the formation command selection phase. When inputting the formation command via the hand-controller, the subject was prone to pay close attentions on the end-effector of the hand-controller so as to ensure it was moved to the correct zone for the MRS control task. Consequently, it is inevitable to divide more visual resources for the system monitoring task on the host PC display with “Hand-CTR” than with “HGBCI” that naturally requires the subject to keep visual attentions on the display. In other

words, the maintained amount of available visual resources by “HGBCI” may have allowed subjects to react faster for the secondary task than by “Hand-CTR”. The phenomenon of reaction times becoming slower with increasing cognitive load has been described by a large body of research (see [39] for an overview).

As can be seen from Fig. 12 (e), the subjective workload rating scores for completing the two tasks in parallel are significantly lower under “Hand-CTR + HGBCI” (48.33 ± 4.80) than those under “Hand-CTR” (64.17 ± 3.00) for all the subjects ($p = 0.00091$). Such results are possibly due to the following reasons. First, it generally took less time to finish the MRS control task under “Hand-CTR + HGBCI” than under “Hand-CTR”. Consequently, “Hand-CTR + HGBCI” resulted in a less heavier mental demand. Second, under the “Hand-CTR” setup, the subject had to move the arm up and down to place the end-effector of the hand-controller into the pre-specified zones in YOZ frame for issuing the formation commands, leading to physical demands to a certain extent. On the contrary, under “Hand-CTR + HGBCI”, the formation commands was inputted by the hand-free hybrid gaze-BCI, without physical activities of limbs. Third, as being reported above, under “Hand-CTR”, the system monitoring task suffered from loss in attentions when inputting the formation commands with the hand-controller, which slowed down the reactions in response to abnormal monitoring events, leaving a frantic task pace. In addition, other advantages of the hybrid gaze-BCI for inputting formation commands with occupied both hands, such as being intuitive, easy-to-use and less prone to error, also may contribute to the less frustration and higher satisfaction on the overall performance reported by the subjects under “Hand-CTR + HGBCI” than “Hand-CTR”.

VII. DISCUSSIONS

A. Multimodal Input Design for MRS Control

The haptic hand-controller is a commonly-adopted device for MRS teleoperation [43], [45], [46], with which the operator could input manipulation commands for realizing supervisory control of MRS and receive haptic feedback on the MRS status for supporting situational awareness. However, such

a manual unimodal input (UMI) device would be inadequate for all the control functions required in our inspection application under the hands-occupied dual-tasking scene. In specific, it is elegantly appropriate for velocity control by mapping end-effector positions to continuous MRS's velocity, but less natural for formation control by mapping the positions to MRS's discrete formation shapes. To create a more operator-friendly interface for MRS control in our considered scenario, this study has presented an MMI design, which outperformed UMI hand-controller according to the experimental results. The highlights for the design are detailed below.

First, the velocity control function is still designated to the hand-controller that excels at inputting continuous velocity commands for MRS, while we propose to realize the formation control function with a hybrid gaze-BCI. It intuitively maps the operator's naturally attended spatial-frequency stimuli, decoded collaboratively from gaze and SSVEP signals, to discrete formation shape commands. Under the hands-occupied dual-tasking scene, such a hands-free input modality is more appropriate than other hands-on alternatives [6]. Moreover, this mapping provides advantages in terms of intuitiveness and ease of interaction, with respect to the previous less natural mapping using the hand-controller.

Second, the proposed design of simultaneous processing and evidence fusion for gaze and SSVEP contributes to the effectiveness of the hybrid gaze-BCI as a modality in the MMI. As have been verified in section VI-B, the hybrid gaze-BCI is found to be able to provide both higher ACCRY and higher ITR than gaze alone and BCI alone for online command selection, suggesting that it enables reliable and efficient selection of on-screen commands to control MRS. Besides, compared to the inputting accuracy of the hybrid gaze-BCI ($96.00 \pm 2.57\%$) in the online single-task paradigm, its inputting accuracy in the dual-task paradigm ($ACC_{in} = 92.66 \pm 2.26\%$) has not deteriorated severely, indicating that the hybrid gaze-BCI built in the single-task paradigm maintained the applicability in the dual-task one.

B. Multimodal Input in Dual-Task Paradigm

Although the feasibility of hands-free MRS control with EEG or gaze as inputs has been confirmed in previous studies [7], [33], [35], all these existing works have only examined their performance in a single-task paradigm (the MRS control task only) and without the involvement of hands. By contrast, we have especially investigated the inputs in a dual-task paradigm that closely simulates a real-world MRS teleoperation scenario, where a healthy operator carries out the MRS control task and system monitoring task concurrently. In particular, it is remarkably to point out that such a dual-task paradigm is more challenging than the single-task one, since the concurrent two tasks compete for cognitive resources, interfering with each other. Moreover, these two tasks always keep the subject's both hands busy. Toward enhancing the operator-MRS interaction in real-world applications, the present study is to our knowledge, the first to investigate the superiority of MMI built by extending the hand-controller with hands-free hybrid gaze-BCI over the unimodal hand-controller under dual-tasking.

In particular, the superiority of the MMI is disclosed from the following three aspects. 1) *The performance of the MRS control task*: subjects complete the primary task (MRS control) with significantly better performance with the MMI than the hand-controller alone (see Fig. 12 (a) and Fig. 12 (b)). 2) *The objective cognitive load*: the significantly shorter reaction time for the secondary task (system monitoring) implies that the MMI leads to lower cognitive load for subjects than the hand-controller alone (see Fig. 12 (d)). 3) *The subjective workload*: all the subjects have reported lower perceived workload with the MMI than those with the hand-controller only (see Fig. 12 (e)).

C. Limitations and Future Work

Featuring an effective hands-free input channel alternative to previous unnatural hands-on one, this study has presented a novel MMI for MRS control under dual-tasking. Results show that it improves the MRS control performance and reduces the operator's cognitive load over UMI. Nevertheless, the proposed method does not exclude the possibility that some type of well-designed, optimized manual or other hands-free interface could enable better performance than the hybrid gaze-BCI. Moreover, there are also several limitations of current work. We have designated the hybrid gaze-BCI to input formation shape commands only, other control functions that are also inappropriate to be realized with hand-controller, could be used to assess superiority of the new MMI to prior UMI. For example, the hybrid gaze-BCI could be further designed to select one of the robots and issue discrete commands to change its camera's view angle for better inspection of environment, etc. Moreover, the uniform epoch length was used across subjects, effective and efficient methods for finding the optimal epoch length for each subject will be explored in future work. In addition, this study did not use physical robots controlled remotely in real time and the haptic feedback of MRS statuses was not implemented, both of which may limit the generalizability of the results found here. Experiments involving interactions with physical MRS will be carried out in future work to assess the adequacy and efficiency of the interface developed.

VIII. CONCLUSION

For the concurrent MRS control and system monitoring tasks where both hands are busy, to overcome the limitation of previous less natural hands-on input for discrete formation commands via the hand-controller, this work has presented a hands-free alternative, i.e., the hybrid gaze-BCI. Moreover, we have assessed the effectiveness of the MMI consisting of the proposed hybrid gaze-BCI and the manual hand-controller in a hands-occupied dual-task paradigm with simulated MRS. Experimental results demonstrate that the proposed sensor-processing solutions for the hybrid gaze-BCI enable more reliable and efficient target selection than gaze/BCI alone. Furthermore, results show that the MMI leads to improved MRS control performance, reduced user cognitive load and perceived workload over the unimodal hand-controller. As such, the findings from this study reinforce the great potential of the hands-free hybrid gaze-BCI that extends traditional

manual MRS control input devices, for augmenting the hands-occupied healthy operator under dual-tasking.

REFERENCES

- [1] A. Kolling, P. Walker, N. Chakraborty, K. Sycara, and M. Lewis, "Human interaction with robot swarms: A survey," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 1, pp. 9–26, Feb. 2015.
- [2] M. Dorigo, G. Theraulaz, and V. Trianni, "Reflections on the future of swarm robotics," *Sci. Robot.*, vol. 5, no. 49, Dec. 2020.
- [3] G.-J. M. Kruijff et al., "Experience in system design for human-robot teaming in urban search and rescue," in *Field and Service Robotics*. Springer, 2014, pp. 111–125.
- [4] V. Villani, L. Sabatini, G. Riggio, C. Secchi, M. Minelli, and C. Fantuzzi, "A natural infrastructure-less human-robot interaction system," *IEEE Robot. Autom. Lett.*, vol. 2, no. 3, pp. 1640–1647, Jul. 2017.
- [5] S. Oviatt, A. Arthur, and J. Cohen, "Quiet interfaces that help students think," in *Proc. 19th Annu. ACM Symp. User Interface Softw. Technol.*, 2006, pp. 191–200.
- [6] H. M. Khalid, "Modelling and analysis of CAD expert behaviour in using manual input devices," *Int. J. Ind. Ergonom.*, vol. 27, no. 2, pp. 79–92, Feb. 2001.
- [7] A. V. Isachenko, D. G. Zhao, E. V. Melnichuk, I. A. Dubynin, B. M. Velichkovsky, and S. L. Shishkin, "The pursuing gaze beats mouse in non-pop-out target selection," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 3518–3523.
- [8] D. G. Zhao, N. D. Karikov, E. V. Melnichuk, B. M. Velichkovsky, and S. L. Shishkin, "Voice as a mouse click: Usability and effectiveness of simplified hands-free gaze-voice selection," *Appl. Sci.*, vol. 10, no. 24, p. 8791, Dec. 2020.
- [9] D. G. Zhao et al., "A passive bci for monitoring the intentionality of the gaze-based moving object selection," *J. Neural Eng.*, vol. 18, no. 2, Mar. 2021, Art. no. 026001.
- [10] J. Pi, P. A. Koljonen, Y. Hu, and B. E. Shi, "Dynamic Bayesian adjustment of dwell time for faster eye typing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2315–2324, Oct. 2020.
- [11] A. M. Feit et al., "Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2017, pp. 1118–1130.
- [12] M. Xu, J. Han, Y. Wang, T.-P. Jung, and D. Ming, "Implementing over 100 command codes for a high-speed hybrid brain-computer interface using concurrent P300 and SSVEP features," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 11, pp. 3073–3082, Nov. 2020.
- [13] Y. Zhang et al., "Two-stage frequency recognition method based on correlated component analysis for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1314–1323, Jun. 2018.
- [14] A. Frisoli et al., "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Trans. Syst., Man, Cybern. C. Appl. Rev.*, vol. 42, no. 6, pp. 1169–1179, Nov. 2012.
- [15] D. P. McMullen et al., "Demonstration of a semi-autonomous hybrid brain-machine interface using human intracranial eeg, eye tracking, and computer vision to control a robotic upper limb prosthetic," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 784–796, Dec. 2014.
- [16] A. Évain, F. Argelaguet, G. Casiez, N. Roussel, and A. Lécuyer, "Design and evaluation of fusion approach for combining brain and gaze inputs for target selection," *Frontiers Neurosci.*, vol. 10, p. 454, Oct. 2016.
- [17] H. Zeng et al., "Semi-autonomous robotic arm reaching with hybrid gaze-brain machine interface," *Frontiers Neurorobotics*, vol. 13, p. 111, Jan. 2020.
- [18] C. Brennan et al., "Performance of a steady-state visual evoked potential and eye gaze hybrid brain-computer interface on participants with and without a brain injury," *IEEE Trans. Human-Mach. Syst.*, vol. 50, no. 4, pp. 277–286, Apr. 2020.
- [19] Y. Tan et al., "An autonomous hybrid brain-computer interface system combined with eye-tracking in virtual environment," *J. Neurosci. Methods*, vol. 368, Feb. 2021, Art. no. 109442.
- [20] S. Soekadar et al., "Hybrid EEG/EOG-based brain/neural hand exoskeleton restores fully independent daily living activities after quadriplegia," *Sci. Robot.*, vol. 1, no. 1, Dec. 2016, Art. no. eaag3296.
- [21] S. Crea et al., "Feasibility and safety of shared EEG/EOG and vision-guided autonomous whole-arm exoskeleton control to perform activities of daily living," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, 2018.
- [22] M. Xu, F. He, T.-P. Jung, X. Gu, and D. Ming, "Current challenges for the practical application of electroencephalography-based brain-computer interfaces," *Engineering*, vol. 7, no. 12, pp. 1710–1712, Dec. 2021.
- [23] W. M. Perlstein, M. A. Cole, M. Larson, K. Kelly, P. Seignourel, and A. Keil, "Steady-state visual evoked potentials reveal frontally-mediated working memory activity in humans," *Neurosci. Lett.*, vol. 342, no. 3, pp. 191–195, May 2003.
- [24] E. C. Lalor et al., "Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 19, pp. 1–9, 2005.
- [25] Z. Işcan and V. V. Nikulin, "Steady state visual evoked potential (SSVEP) based brain-computer interface (BCI) performance under different perturbations," *PLoS ONE*, vol. 13, no. 1, Jan. 2018, Art. no. e0191673.
- [26] Y.-P. Lin, Y. Wang, and T.-P. Jung, "A mobile SSVEP-based brain-computer interface for freely moving humans: The robustness of canonical correlation analysis to motion artifacts," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 1350–1353.
- [27] S. Kanoga, M. Nakanishi, A. Murai, M. Tada, and A. Kanemura, "Robustness analysis of decoding SSVEPs in humans with head movements using a moving visual flicker," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016009.
- [28] S. Li and X. Zhang, "Implicit intention communication in human-robot interaction through visual behavior studies," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 4, pp. 437–448, Aug. 2017.
- [29] B. Park and R. Brünken, "The rhythm method: A new method for measuring cognitive load—An experimental dual-task study," *Appl. Cognit. Psychol.*, vol. 29, no. 2, pp. 232–243, Mar. 2015.
- [30] R. Brunken, J. L. Plass, and D. Leutner, "Direct measurement of cognitive load in multimedia learning," *Educ. Psychol.*, vol. 38, no. 1, pp. 53–61, Jan. 2003.
- [31] C. D. Wickens, "Processing resources and attention," in *Multiple-Task Performance*. Boca Raton, FL, USA: CRC Press, 2020, pp. 3–34.
- [32] Y. Ke, P. Liu, X. An, X. Song, and D. Ming, "An online SSVEP-BCI system in an optical see-through augmented reality environment," *J. Neural Eng.*, vol. 17, no. 1, Feb. 2020, Art. no. 016066.
- [33] S. Zhao, Z. Li, R. Cui, Y. Kang, F. Sun, and R. Song, "Brain-machine interfacing-based teleoperation of multiple coordinated mobile robots," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 5161–5170, Jun. 2016.
- [34] G. K. Karavas, D. T. Larsson, and P. Artemiadis, "A hybrid BMI for control of robotic swarms: Preliminary results," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 5065–5075.
- [35] J.-H. Jeong, D.-H. Lee, H.-J. Ahn, and S.-W. Lee, "Towards brain-computer interfaces for drone swarm control," in *Proc. 8th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2020, pp. 1–4.
- [36] C. I. Penalosa and S. Nishio, "BMI control of a third arm for multi-tasking," *Sci. Robot.*, vol. 3, no. 20, Jul. 2018, Art. no. eaat1228.
- [37] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, and R. P. N. Rao, "Cortical activity during motor execution, motor imagery, and imagery-based online feedback," *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 4430–4435, Mar. 2010.
- [38] Y. Santiago-Espada, R. Myer, K. Latorella, and J. Comstock, "The multi-attribute task battery II (MATB-II) software for human performance and workload research: A user's guide (NASA/TM-2011-217164)," Nat. Aeronaut. Space Admin., Langley Res. Center, Hampton, VA, USA, Tech. Rep. L-20031, 2011.
- [39] G. M. Grimes and J. S. Valacich, "Mind over mouse: The effect of cognitive load on mouse movement behavior," in *Proc. Int. Conf. Inf. Syst. Fort Worth, TX, USA: Association for Information Systems*, 2015, pp. 1–13.
- [40] A. Housholder, J. Reaban, A. Peregrino, G. Votta, and T. K. Mohd, "Evaluating accuracy of the Tobii eye tracker 5," in *Proc. Int. Conf. Intell. Hum. Comput. Interact.* Cham, Switzerland: Springer, 2021, pp. 379–390.
- [41] V. P. Oikonomou et al., "Comparative evaluation of state-of-the-art algorithms for SSVEP-based BCIs," 2016, *arXiv:1602.00904*.
- [42] 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.
- [43] C. Ju and H. I. Son, "A distributed swarm control for an agricultural multiple unmanned aerial vehicle system," *Proc. Inst. Mech. Eng., I, J. Syst. Control Eng.*, vol. 233, no. 10, pp. 1298–1308, 2019.
- [44] J. R. Wolpaw, "Brain-computer interfaces (BCIs) for communication and control," in *Proc. 9th Int. ACM SIGACCESS Conf. Comput. Accessibility*, 2007, pp. 1–2.
- [45] C. Masone, P. R. Giordano, H. H. Bulthoff, and A. Franchi, "Semi-autonomous trajectory generation for mobile robots with integral haptic shared control," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 6468–6475.
- [46] Y. Zhang, G. Song, Z. Wei, H. Sun, and Y. Zhang, "Bilateral teleoperation of a group of mobile robots for cooperative tasks," *Intell. Service Robot.*, vol. 9, no. 4, pp. 311–321, Oct. 2016.