Neural Correlation of EEG and Eye Movement in Natural Grasping Intention Estimation

Chengyu Lin, Chengjie Zhang, Jialu Xu, Renjie Liu, Yuquan Leng, and Chenglong Fu, *Member, IEEE*

***Abstract*— Decoding the user's natural grasp intent enhances the application of wearable robots, improving the daily lives of individuals with disabilities. Electroencephalogram (EEG) and eye movements are two natural representations when users generate grasp intent in their minds, with current studies decoding human intent by fusing EEG and eye movement signals. However, the neural correlation between these two signals remains unclear. Thus, this paper aims to explore the consistency between EEG and eye movement in natural grasping intention estimation. Specifically, six grasp intent pairs are decoded by combining feature vectors and utilizing the optimal classifier. Extensive experimental results indicate that the coupling between the EEG and eye movements intent patterns remains intact when the user generates a natural grasp intent, and concurrently, the EEG pattern is consistent with the eye movements pattern across the task pairs. Moreover, the findings reveal a solid connection between EEG and eye movements even when taking into account cortical EEG (originating from the visual cortex or motor cortex) and the presence of a suboptimal classifier. Overall, this work uncovers the coupling correlation between EEG and eye movements and provides a reference for intention estimation.**

***Index Terms*— EEG, eye movements, intent, feature extraction, recognition.**

I. Introduction

BRAIN-COMPUTER interface (BCI) creates a communication link between the human brain and external devices [1], [2], aiming to enhance intent perception in many applications. Decoding natural grasping intentions empowers the control of wearable robots, such as prostheses and exoskeletons for individuals with disabilities [3], [4], [5], subsequently boosting BCI naturalness and enhancing their daily lives.

The Electroencephalogram (EEG) and eye movements are two natural representations occurring when the user generates the grasp intent in their mind [6], [7], [8]. Specifically, the intent towards grasping the target object is mainly expressed through the mind's thought and the visual perception of the eye. For example, when people want to eat something, they tend to look at it and have grasping thoughts in their brains. Therefore, the users' intentions can be inferred from EEG and eye movements, which have already been successfully used to control prosthetic limbs or assistive robots [9], [10], [11], [12].

The EEG and eye movement features have been used to identify different intention forms due to their respective strengths and characteristics [13], [14]. The rich spatial information of eye movements is often related to the target position, e.g., visual search and behavior analysis [15], [16]. Moreover, the brain has a distinct EEG for different events [17] and is used to classify different event-related states, such as the typical steady-state visual evoked potential (SSVEP) [18] and motor imagery (MI) of the body [19], [20]. Therefore, the degree of freedom of BCI is increased by combining the EEG-generated command with the eye movement spatial information. For instance, Kim et al. [21] utilized EEG and eye movements to control a drone's flight, where the EEG controlled its ascent and descent, and the eye movements controlled the flight directions (forward, backward, left, and right). Meena et al. [22] built the BCI system called Gaze-MI, which combined data from the EEG and eye movements into eight control commands to make interaction much easier.

Meanwhile, the same intention can be recognized by eye movements and EEG due to the coupling between the brain and eye in neural connections [23]. It should be noted that EEG can be decoded into positional information similar to eye movement signals [24]. For instance, using a BCI setup, Wodlinger et al. [25] demonstrated a continuous translation and orientation control of a prosthetic limb by a person with tetraplegia. Similarly, eye movements are directly related to cognitive intentions and can be decoded into commands similar to EEG [26], [27]. Therefore, the same intent can be decoded by fusing EEG and eye movements [28], [29].

At present, eye movements and EEG intentions have been decoded in rigorous experimental conditions and applied to control robots. However, the neural correlation between

Manuscript received 30 January 2023; revised 2 July 2023 and 21 September 2023; accepted 19 October 2023. Date of publication 26 October 2023; date of current version 7 November 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U1913205 and Grant 52175272, in part by the Stable Support Plan Program of Shenzhen Natural Science Fund under Grant 20200925174640002, in part by the Science, Technology, and Innovation Commission of Shenzhen Municipality under Grant ZDSYS20200811143601004 and Grant JCYJ20220530114809021, and in part by the Special Funds for the Cultivation of Guangdong College Students-Scientific and Technological Innovation ("Climbing Program" Special Funds) under Grant pdjh2022a0453. *(Corresponding author: Chenglong Fu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Sustech Medical Ethics Committee under Approval No. 20210009.

The authors are with the Shenzhen Key Laboratory of Biomimetic Robotics and Intelligent Systems and the Guangdong Provincial Key Laboratory of Human-Augmentation and Rehabilitation Robotics in Universities, Department of Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: lincy2019@mail.sustech.edu.cn; fucl@sustech.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3327907

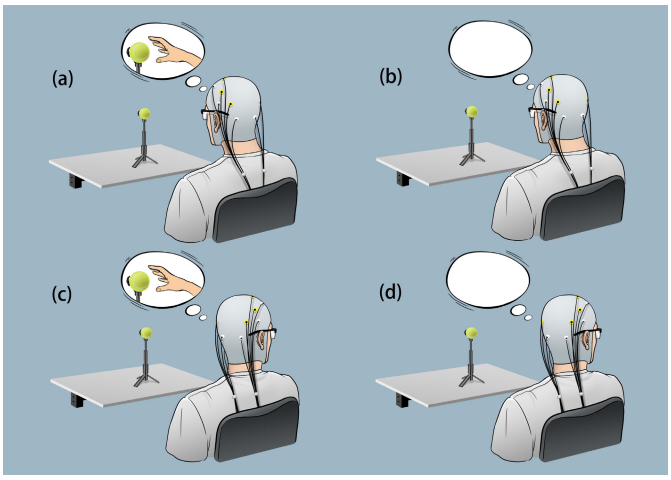

Fig. 1. Four typical grasping-related intents. (a) Gaze at a Target with Grasp Intent (Gaze Grasp), (b) Gaze at a Target without Grasp Intent (Gaze No-Grasp), (c) Not Gaze at a Target with Grasp Intent (No-Gaze Grasp), and (d) Not Gaze at a Target without Grasp Intent (No-Gaze No-Grasp).

EEG and eye movement remains unknown. Motivated by this research gap, this paper studies the consistency of EEG and eye movement by decoding the natural grasping intent based on feature fusion. Thus, this work provides a valuable reference for intention estimation.

The structure of this paper unfolds as such: Section II delineates the specifics of the experiments, along with data extraction and synchronization procedures. Section III elaborates on the methodologies adopted for estimating intent via EEG and eye movement. The ensuing Section IV showcases the corresponding experimental outcomes. Discussions of these results, complemented by insights into potential avenues for future research, are covered in Section V. The paper culminates with conclusions in Section VI.

II. EEG and Eye Movements Experiments

A. Research Goal

To decipher the inherent intent behind grasping through feature fusion and delve into the consistency between EEG and eye movements, we designed an experiment centered on natural grasp intentions. In this experiment, we decoded the intent across six potential task pairs, drawing insights from both EEG and eye movement data.

B. Research Design

The grasping intent towards the target object is primarily expressed through the brain's cognition and the eye's visual perception. Therefore, the natural grasping intent can be summarized as the following four typical grasping-related intents illustrated in Fig. 1.

a) Gaze at a Target with Grasp Intent (Gaze Grasp). The user typically obtains the environmental information first and generates the grasp intent through the brain's processing, such as grabbing objects that suddenly appear.

b) Gaze at a Target without Grasp Intent (Gaze No-Grasp). The user only looks at the object but does not have the intention to grasp it, such as during a visual search.

c) Not Gaze at a Target with Grasp Intent (No-Gaze Grasp). The user first generates the grasping intent in mind and then obtains the information about the object through their vision. For example, the intention of drinking water occurs, and the user picks up the cup through visual perception.

d) Not Gaze at a Target without Grasp Intent (No-Gaze No-Grasp). The user neither looks at the target object nor generates a grasping intent, such as doing something unrelated to the target object.

These four typical grasping-related intentions represent the possible intentions for the target objects and can be extended to visually guided intention recognition tasks. Therefore, this paper decodes these four intents using EEG and eye movement features and explores the neural correlation between these features.

C. Procedure

The tasks were performed without visual guidance and stimulation to explore the natural intent expression of EEG and eye movements, minimizing any impact on the experimental results. As a result, the subjects were guided by voice prompts. The experiment was conducted in a quiet room, with participants seated 50 cm away from the target object (tennis ball). They performed the four typical grasping-related intents presented in Section II-B in a randomly alternated order. By gazing at the object and generating imagery grasping intent in their brain, the subjects carried out the task without physically moving their bodies.

Each participant performed 4 tasks randomly in each session, with each task conducted 10 times. Therefore, in 5 sessions, each participant performed 200 trials in total. After the onset of the voice cue, the subjects randomly performed an imaginary grasping task for 4 s, as illustrated in Fig. 2. In the case of the Gazing at Target condition, the subjects were instructed to gaze at the tennis ball on the table, while in the case of the Not Gazing at Target condition, the subjects chose freely where to gaze, either on the wall or the table. In the Grasp Intent conditions, the subjects were instructed to generate the intention to grasp the tennis ball using motor imagery.

D. Participants and Apparatus

Nine healthy participants (six males and three females, 21×25 years old) were recruited to participate in the experiment. They have normal or corrected to normal vision. All experiments were approved by the Sustech Medical Ethics Committee (approval number: 20210009, date: 2021/3/2). The participants were informed about the study procedure and signed the consent form before the experiment. They were asked to wear EEG and eye-tracking devices and make necessary adjustments until the data could be reliably recorded.

An actiCHamp amplifier and actiCAP active electrodes (Brain Products, Germany) encompassing the entire head were used to record the EEG signals (see Fig. 3). The impedances

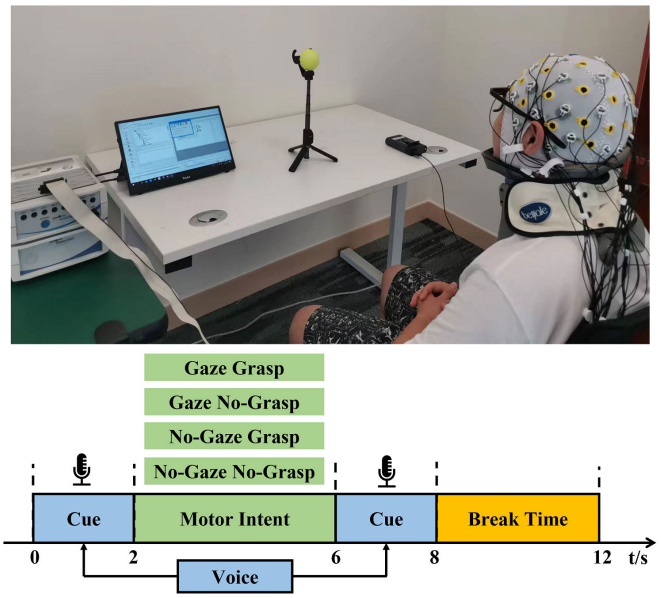

Fig. 2. Diagram of the trial and task setup: Each trial commenced with a 2-second auditory cue, followed by the execution of one of the four grasping-related tasks. Subsequently, the user was prompted by a voice to rest for 4 seconds.

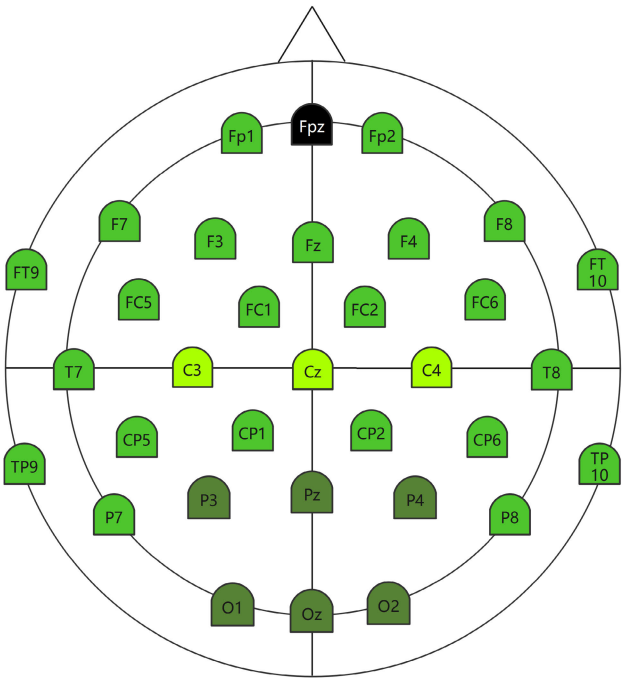

Fig. 3. Electrode positions corresponding to the 32 electrode locations symmetrically placed to cover the entire head. The vision-related EEG is distributed in dark green areas, and the motor-related EEG is distributed in light green areas.

of the reference EEG electrodes remained below 15 Kohms. Utilizing the Tobii glasses 2 (Tobii, Canada), the eye movements were recorded at a 40 Hz sampling rate.

E. Data Extraction and Synchronization

To capture grasping-related intentions through eye movement and EEG data, a data flow framework has been designed for the extraction and synchronization of these

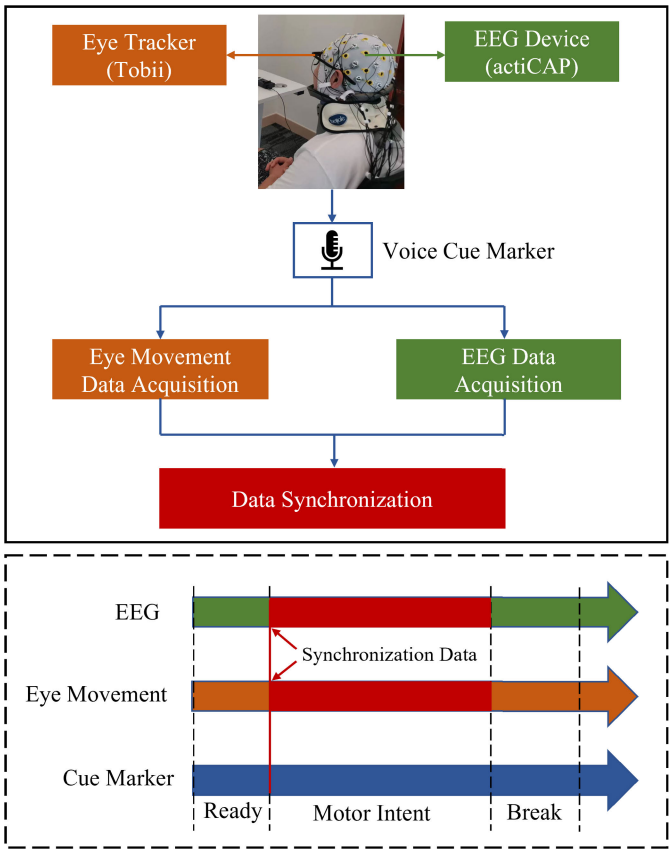

Fig. 4. The data extraction and synchronization flow framework of EEG and eye movement. The eye movement, EEG, task flow, and synchronization data are presented in orange, green, blue, and red, respectively.

metrics (Fig. 4). Specifically, the eye trackers and EEG devices collected the eye movements and EEG, and the data were labeled by synchronizing the markers of the task flow. Through segmenting the task process, eye movement and EEG data corresponding to 200 imaginative grasping tasks, each with a duration of 4 seconds, were accurately extracted.

III. Intent Estimation Method

A. EEG Data Preprocessing and Feature Extraction

Raw EEG data include loud noises, such as blinking and electromyography (EMG). Therefore, prior to feature extraction, BrainVison (BrainProduct, USA) was used for off-line preprocessing of EEG data to reduce noise and improve signal-to-noise ratio. In this study, the recorded EEG signals were band-pass filtered between 0.5×40 Hz using a 4th-order Butterworth filter. Additionally, independent component analysis was used to remove electrooculogram (EOG) or EMG.

The main objective of the feature extraction phase is to extract significant features from the data that can effectively characterize EEG fragments. Therefore, different EEG features were extracted and combined to improve classification performance. Below is a detailed explanation of each extracted feature.

1) Standard deviation (SD). This feature helps compare the different dispersion of the various EEG data samples from

their means and reflects the EEG signal fluctuation in distinct intent. SD is defined as follows:

$$SD = \sqrt{\frac{1}{N}\sum_{N}^{n=1}(X(n) - \mu_X)^2}, \tag{1}$$

where μ_X is the mean value of the signal.

2) Spectral entropy (SE). SE depicts the uniformity of EEG power spectrum distribution and can reflect the nonlinear characteristics of EEG. It can be calculated using power spectral density (PSD):

$$SE = -\sum_{f_n}^{f=0} PSD(f)\log_2(PSD(f)), \tag{2}$$

where f_n denotes the half of the sampling frequency and PSD denotes the normalized power spectral density.

3) Fractal dimension (FD). The EEG signal is treated as a geometric figure by the fractal dimension, and its correlation and evolutionary characteristics are assessed by quantifying the fractional space occupied. FD provides a method to estimate the complexity of nonlinear EEG signals. The resulting calculations are as follows:

$$FD = \frac{\log_{10}^{n}}{\log_{10} n + \log_{10}(n/(n + 0.4 \times N))}, \tag{3}$$

where f_N is the number of sign changes in the signal's derivative and f_n is the length of the sequence, which is calculated by subtracting the digital signal's effect on consecutive series.

4) Hjorth parameters were developed to calculate the complexity of EEG signals in the time domain to reflect event-related changes in the brain. Hjorth consists of three parameters: activity, mobility, and complexity, which are described below:

$$Activity = \mathrm{Var}(y(t)) \tag{4}$$

$$Mobility = \sqrt{\frac{\mathrm{Var}\left(\frac{dy(t)}{dt}\right)}{\mathrm{Var}(y(t))}} \tag{5}$$

$$Complexity = \frac{\mathrm{Mobility}\ \left(\frac{dy(t)}{dt}\right)}{\mathrm{Mobility}\ (y(t))}, \tag{6}$$

where $y(t)$ is the signal and the power, mean frequency, and standard deviation of the power spectrum, respectively, are represented by the activity, mobility, and complexity parameters.

5) Wavelet feature (WF). Wavelet transform captures the time-frequency (TF) domain of EEG signals. In wavelet decomposition, TF transformation is performed for each time point and each frequency resolution of the original signal. In this paper, wavelet decomposition is used to decompose signals in the theta (3-8 Hz), alpha (9-12 Hz), beta (13-30 Hz), and gamma (31-40 Hz) bands as follows:

$$E_j = \sum_{N}^{k=1}\left(D_j(k)^2\right), \tag{7}$$

where k is the number of wavelet coefficients and D_j are the detail coefficients of the jth level of wavelet decomposition that correspond to the EEG band j.

TABLE I
CONDITIONS AND TASK PAIRS

Grasp intent	Gaze Grasp vs. No-Gaze Grasp
No-Grasp Intent	Gaze No-Grasp vs. No-Gaze No-Grasp
Grasp intent vs. No-Grasp intent	Gaze Grasp vs. Gaze No-Grasp
	No-Gaze Grasp vs. No-Gaze No-Grasp
	Gaze No-Grasp vs. No-Gaze Grasp
	Gaze Grasp vs. No-Gaze No-Grasp

Each feature was computed using task data lasting 4 seconds from a non-overlapping window. Subsequently, a discerning feature set, encompassing [SD, SE, FD, Hjorth, WF], was standardized to serve as the feature vector for EEG decoding. Combining various feature types, such as those from the time domain, frequency domain, and time-frequency domain, inherently increases data dimensionality. This enhanced dimensionality typically leads to improved classification accuracy [30], [31], [32]. Consequently, event-related natural grasping intentions can be decoded with high precision.

B. Eye Movements Data Preprocessing and Feature Extraction

The eye movements reflect the user's intent and consist of the fixation pattern and pupil size. Changes in pupil diameter are closely related to visual concentration processing and the user's cognitive psychology [33], [34]. The average pupil diameter represents the average cognitive level of the users, while a diameter variance can represent the change in the users' intents and reflect the characteristics of the cognitive level over time. Therefore, the pupil's mean diameter (MD) and amplitude change (AC) were selected as the pupil features.

The fixation pattern helps the user gather visual information about the environment by keeping the visual gaze in one position. Therefore, it can be used to analyze the visual cognitive processes and reflect intent about the target object. Rapid eye movements, called saccades (S), can shift focus from one area to another, which helps a user better understand environmental information. Each feature was derived from 4-second task data taken from a non-overlapping window. The [MD, AC, S] set was then standardized and used as the feature vector for decoding eye movements to discern natural grasping intentions.

C. Evaluation Metrics

Upon synchronization and extraction, eye movement and EEG data were randomly divided into training and testing datasets at a 4:3 ratio. We devised six distinct binary classifiers with the goal of investigating the performance of all six possible task pairings, as illustrated in Table I. The reported accuracy was calculated using the following formula:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%, \tag{8}$$

where TP, TN, FP, and FN represent the total sample of the true positive, true negative, false positive, and false negative, respectively.

TABLE II
CLASSIFICATION PERFORMANCE COMPARISON OF VARIOUS CLASSIFIERS, INCLUDING RANDOM FOREST CLASSIFIER (RFC), K-NEAREST NEIGHBORS (KNN), SUPPORT VECTOR MACHINE (SVM), LINEAR DISCRIMINANT ANALYSIS (LDA), AND LOGISTIC REGRESSION (LR), EVALUATED FOR BOTH EYE MOVEMENT AND EEG DECODING

S_i	RFC	KNN	SVM	LDA	LR	Mean
EEG Classification Accuracy - (%) mean						
S1	80.8	76.8	82.6	80.1	82.0	80.5
S2	86.7	82.5	92.2	89.5	88.9	88.0
S3	84.8	83.9	89.3	86.6	89.3	86.8
S4	85.7	78.0	86.1	86.8	87.7	84.9
S5	70.5	63.0	72.8	69.9	74.9	70.2
S6	79.2	71.9	83.2	80.6	83.3	79.6
S7	92.7	75.4	86.6	86.4	87.9	85.8
S8	83.3	79.9	87.1	85.0	88.2	84.7
S9	85.8	78.7	85.9	83.6	85.5	83.9
Mean	83.3	76.7	85.1	83.2	85.3	/
Eye Movement Classification Accuracy - (%) mean						
S1	84.5	82.4	79.5	83.3	83.5	82.6
S2	85.0	83.2	82.8	81.7	84.3	83.4
S3	87.3	80.9	81.8	81.8	82.5	82.9
S4	87.5	86.0	86.8	84.7	87.3	86.5
S5	77.5	75.5	77.7	74.1	75.0	76.0
S6	84.3	85.5	85.1	83.2	85.4	84.7
S7	88.8	84.1	85.1	86.4	86.8	86.2
S8	78.2	72.0	68.7	68.0	74.2	72.2
S9	84.7	75.5	77.8	74.7	76.4	77.8
Mean	84.2	80.5	80.6	79.8	81.7	/

In this research, the Paired t-test was utilized to discern notable differences in classification efficacy between EEG and eye movements. The Pearson correlation coefficient was leveraged to evaluate the interrelation of eye movement and EEG over the six designated task pairs. Moreover, during the investigation of coherence between eye movements, visual cortex EEG, and motor cortex EEG, a one-way ANOVA was employed to underline the pronounced disparities among them.

IV. RESULTS

A. Classifier Performance in Intent Decoding

The classification performance of various classifiers, including Random Forest Classifier (RFC), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Logistic Regression (LR), was evaluated for both eye movement and EEG decoding. The average performance of several classifiers across these six task pairs was compared in the Table II. Therefore, LR ($C_{EEG} = 85\%$) and RFC ($C_{EYE} = 84\%$) were identified as the optimal classifiers for EEG and eye movement, respectively. The other classifiers are referred to as suboptimal classifiers. Here, C_{EEG} represents the classification accuracy of EEG and C_{EYE} represents the classification accuracy of eye movements. Regardless of decoding for eye movements or EEG, suboptimal classifiers also have a relatively high classification accuracy, decreasing by less than 10% compared to the optimal classifier.

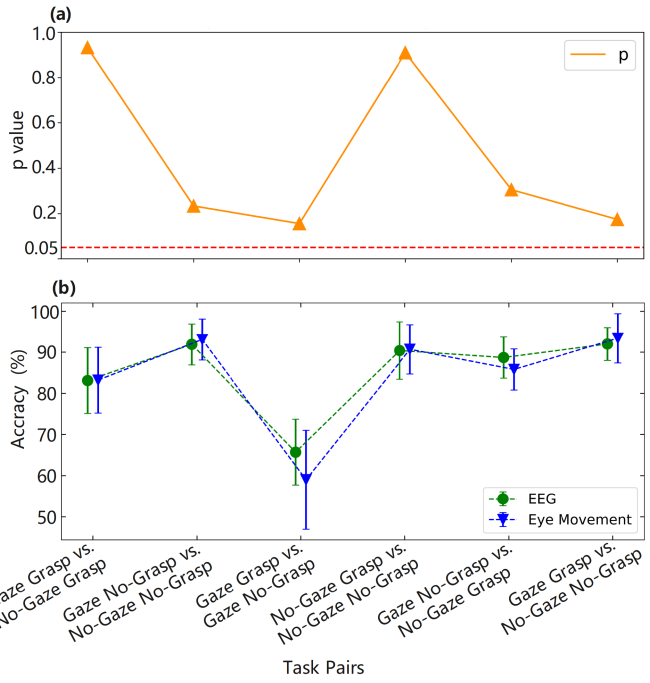

Fig. 5. Performance of the EEG and eye movements in the natural grasp intent estimation. a) P value of the Paired t-test between the EEG and eye movements for all 6 task pairs and b) Grand-average classification of EEG and eye movements accuracy for all 6 task pairs.

B. Coupling Correlation Between EEG and Eye Movements

The classification performance of EEG and eye movements for each task pair was assessed using Paired t-test analysis, as illustrated in Fig. 5. The experimental outcomes demonstrated that both EEG and eye movements achieve a remarkably consistent accuracy for each task pair during the natural grasp intent estimation. Moreover, no statistically significant differences were observed in eye movement and EEG performance among the nine subjects for all possible tasks ($P = 0.332$, $t = -0.979$).

The experimental results demonstrated that the Pearson correlation coefficient between eye movement and EEG for all task pairs across the nine subjects was 0.81 ($P < 0.01$), suggesting that EEG and eye movement exhibit an inherent coupling in task recognition. We also report Pearson correlation coefficients for eye movement and EEG for each subject, as illustrated in Fig. 6. Most subjects had Pearson correlation coefficients greater than 0.8, except for s2 (0.75) and s3 (0.77). Taken together, these findings indicate that the coupling between EEG and eye movement intent patterns remains consistent and strong.

The coupling of eye movement and EEG remains consistent across the six task pairs, even when the decoding performance of one signal decreases. Table III displays the consistency performance of the EEG under the suboptimal classifier (KNN) and the eye movement under the optimal classifier (RFC) for the six task pairs of nine subjects. Here,

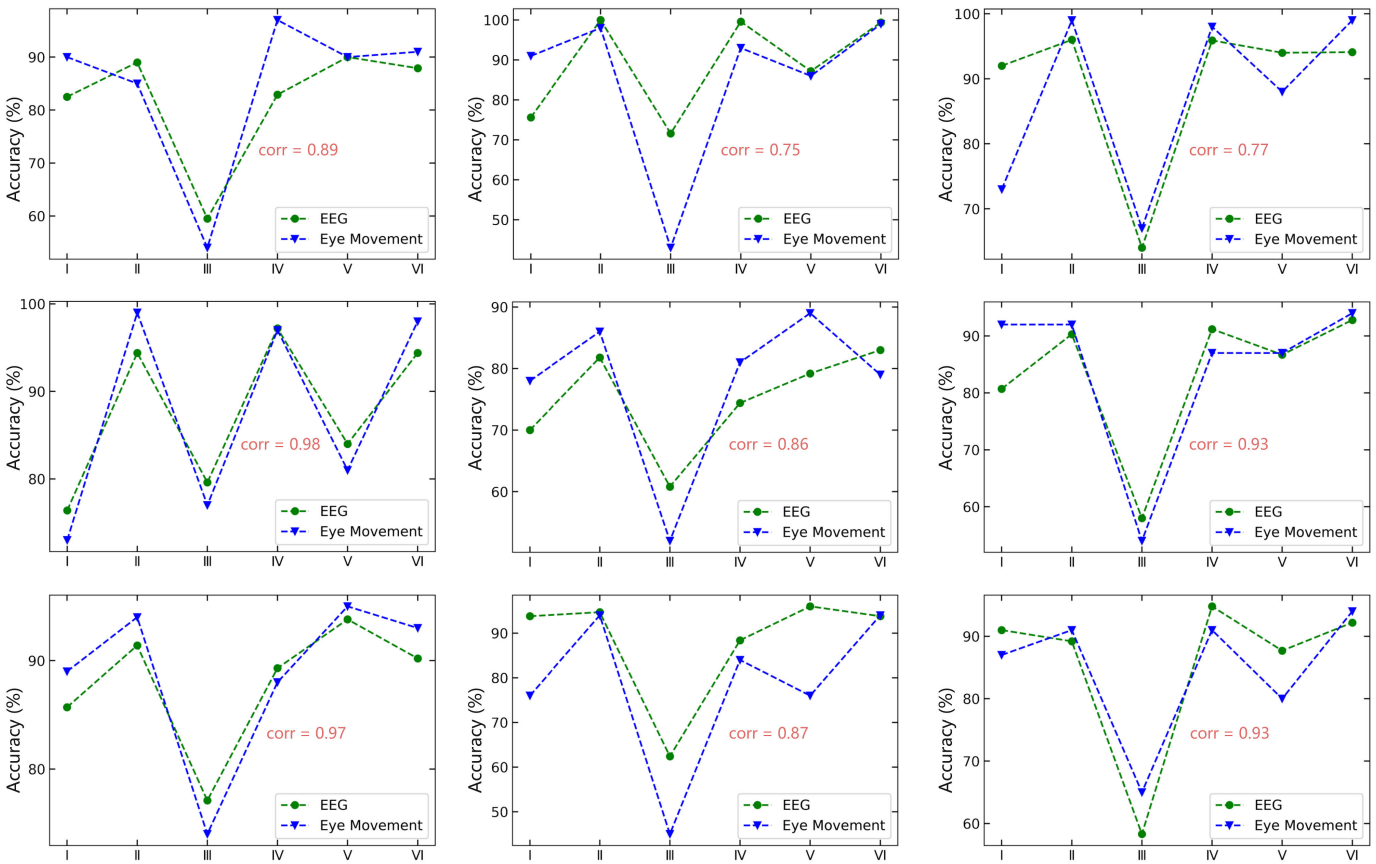

Fig. 6. Classification accuracy of EEG and eye movements across six task pairs for nine subjects, with corresponding Pearson correlation coefficients (corr) between eye movements and EEG. Green represents EEG, and blue represents eye movement. The six tasks in the Roman numerals of the x-axis are: I: Gaze Grasp vs. No-Gaze Grasp; II: Gaze No-Grasp vs. No-Gaze No-Grasp; III: Gaze Grasp vs. Gaze No-Grasp; IV: No-Gaze Grasp vs. No-Gaze No-Grasp; V: Gaze No-Grasp vs. No-Gaze Grasp; VI: Gaze Grasp vs. No-Gaze No-Grasp.

TABLE III

THE CONSISTENCY PERFORMANCE OF THE EEG UNDER THE SUBOPTIMAL CLASSIFIER (KNN) AND THE EYE MOVEMENT UNDER THE OPTIMAL CLASSIFIER (RFC) FOR THE SIX TASK PAIRS OF NINE SUBJECTS. HERE, Si_j REPRESENTS THE EYE MOVEMENT OR EEG OF A SUBJECT, WHERE i DENOTES THE SUBJECT INDEX (RANGING FROM 1 TO 9), AND j DENOTES THE SIGNAL TYPE, EITHER G (EEG) OR Y (EYE). CORR REPRESENTS THE PEARSON CORRELATION COEFFICIENT BETWEEN EYE MOVEMENT AND EEG FOR SIX TASK PAIRS. SIX TASKS IN ROMAN NUMERALS ARE: I: GAZE GRASP VS. NO-GAZE GRASP; II: GAZE NO-GRASP VS. NO-GAZE NO-GRASP; III: GAZE GRASP VS. GAZE NO-GRASP; IV: NO-GAZE GRASP VS. NO-GAZE NO-GRASP; V: GAZE NO-GRASP VS. NO-GAZE GRASP; VI: GAZE GRASP VS. NO-GAZE NO-GRASP

Si_j	$S1_g$	$S1_y$	$S2_g$	$S2_y$	$S3_g$	$S3_y$	$S4_g$	$S4_y$	$S5_g$	$S5_y$	$S6_g$	$S6_y$	$S7_g$	$S7_y$	$S8_g$	$S8_y$	$S9_g$	$S9_y$
I	80.0	90.2	65.6	98.1	75.4	73.2	61.6	73.3	56.8	78.3	66.3	92.3	84.3	89.1	85.5	76.4	83.0	87.3
II	87.4	85.3	99.4	98.3	95.7	99.4	89.2	99.5	69.8	86.4	81.5	92.7	74.0	94.6	86.3	94.8	86.6	91.6
III	43.5	54.3	52.4	43.5	62.9	67.2	68.0	77.2	42.0	52.9	47.0	54.8	56.2	74.5	52.1	45.5	48.7	65.7
IV	84.7	97.4	99.2	93.6	92.9	98.6	88.1	97.7	72.8	81.3	86.2	87.5	80.7	88.8	83.5	84.3	91.2	91.7
V	85.5	90.6	78.8	86.3	80.6	88.3	70.4	81.8	62.4	89.4	62.7	87.4	83.3	95.5	88.3	76.6	76.0	80.4
VI	79.7	91.6	99.8	99.3	96.0	99.2	90.8	98.2	74.4	79.7	87.7	94.3	73.9	93.2	83.6	94.8	86.5	94.7
Corr	0.94		0.75		0.97		0.99		0.79		0.79		0.78		0.86		0.97	

Si_j represents the eye movement or EEG of a subject, where i denotes the subject index (ranging from 1 to 9), and j denotes the signal type, either g (EEG) or y (eye). The KNN classifier performance of EEG is 8.6% lower compared to that of REG, as illustrated in Table II. The average Pearson correlation coefficient between eye movement and EEG for all task pairs across the nine subjects was 0.80 ($P < 0.01$). A majority of the subjects exhibited strong coupling between EEG and eye movement in the decoding of natural grasp intent.

In the section, the classification accuracies for the six tasks, using EEG channel inputs from the visual cortex or motor cortex (see Fig.3) with those using eye movement is analysed by ANOVA, as shown in Fig. 7. Experimental findings reveal that when recognizing using partial EEG data (from the visual cortex or motor cortex) in comparison to whole-channel EEG, the outcomes are slightly less robust than with eye movements. Nonetheless, for the majority of task pairs, the performances of eye movements and EEG from different regions align consistently (P $>$ 0.05). It's evident

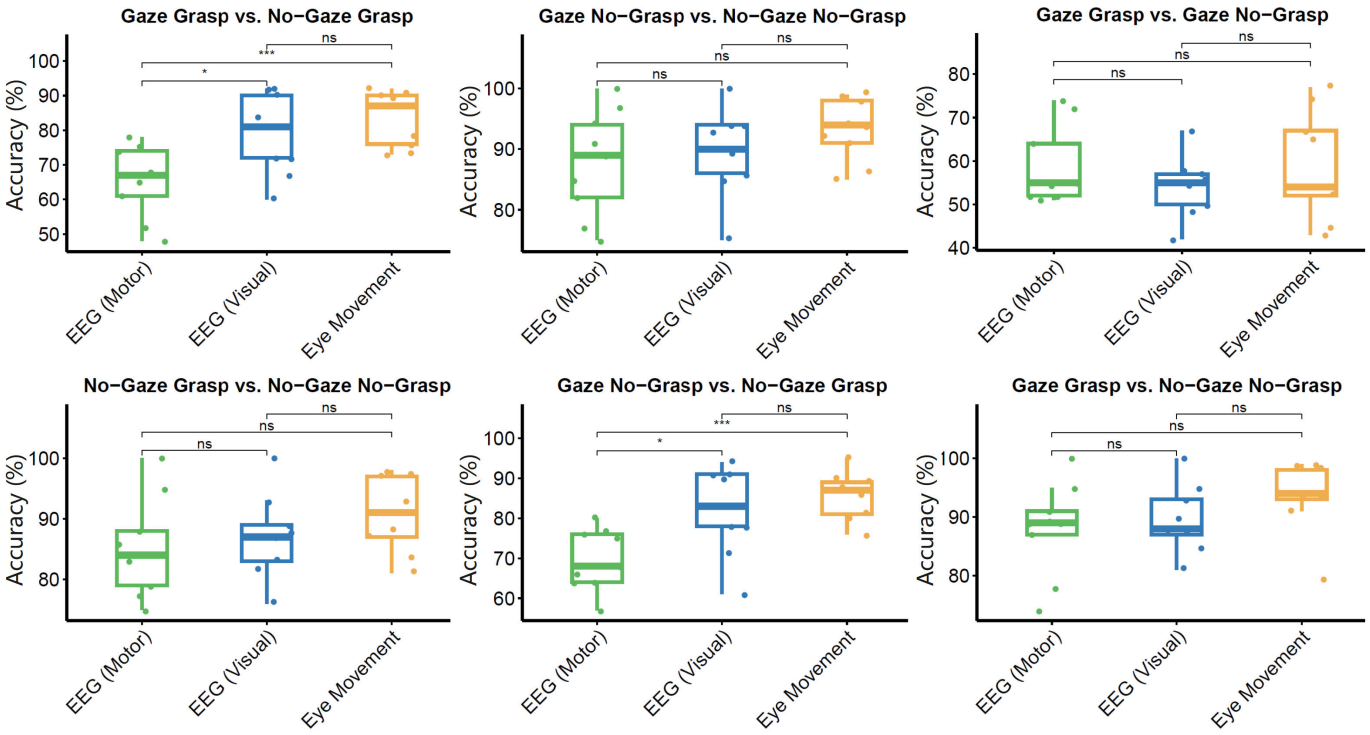

Fig. 7. Classification accuracies for the six tasks, comparing the performance using EEG channel inputs from the visual cortex and motor cortex with those using eye movement-related features. The boxplots display the distribution of classification accuracies for six task pairs. Asterisks denote significant effects at *P $<$ 0.05, **P $<$ 0.01, and ***P $<$ 0.001. "ns" stands for "not significant," indicating there's no statistical difference.

that visual-related intentions manifest differently in EEG and eye movements. This distinction is most pronounced in the Gaze Grasp vs. No-Gaze Grasp and Gaze No-Grasp vs. No-Gaze Grasp comparisons, where EEG from the visual region demonstrates greater consistency than from the motor region.

C. Gaze and Grasp Intent Decoding Performance

Gaze intent detection comprises the Gaze Grasp vs. No-Gaze Grasp task pair and the Gaze No-Grasp vs. No-Gaze No-Grasp task pair. Although eye movement serves as an evident detection signal, EEG can also offer the same judgment in this condition. Specifically, no significant difference was observed between EEG and eye movements in gaze intent detection ($P = 0.93$ for Gaze Grasp vs. No-Gaze Grasp and $P = 0.23$ for Gaze No-Grasp vs. No-Gaze No-Grasp, respectively). For the grasp intent conditions, the classification accuracy of Gazing at Target vs. Not Gazing at Target reached 83.1% (EEG) and 83.2% (eye movements) across all subjects. For the no-grasp intent conditions, the classification accuracy of Gazing at Target vs. Not Gazing at a Target for all subjects reached 91.9% (EEG) and 93.1% (eye movements), which is approximately 10% higher than Gaze Grasp vs. No-Gaze Grasp condition. Overall, all subjects achieved an impressive classification accuracy in Gaze intent detection. Moreover, eye movement and EEG have statistically significant differences between the two Gaze intent task pairs (Fig. 8).

The grasp intent detection is analyzed for the four Grasp Intent vs. No-Grasp Intent task pairs (see Table I). Considering the Gaze Grasp task pairs, i.e., Gaze Grasp vs. Gaze No-Grasp and Gaze Grasp vs. No-Gaze No-Grasp, both eye movement and EEG show statistically significant differences between the two Gaze Grasp task pairs (Fig. 8). For the Gaze Grasp vs. Gaze No-Grasp task pair, the average decoding ability is relatively poor ($C_{EEG} = 65.7\%$ and $C_{EYE} = 59\%$, and all subjects reached a significantly low accuracy in the six task pairs, as shown in Fig. 6. For the Gaze Grasp vs. No-Gaze No-Grasp task pair, the classification accuracy across all subjects reached 92% (EEG) and 93.4% (eye movements), which is approximately 30% higher than Gaze Grasp vs. Gaze No-Grasp. In this condition, gazing at the object significantly affects the detection of grasp intent for both EEG and eye movement. Regarding the No-Gaze Grasp task pairs, i.e., No-Gaze Grasp vs. Gaze No-Grasp and No-Gaze Grasp vs. No-Gaze No-Grasp, no statistically significant differences are observed between the two Gaze Grasp task pairs. Both EEG and eye movement achieve excellent performance in these conditions.

V. Discussion

In this research, we thoroughly examined the coupling correlation between EEG and eye movement decoding in the context of gaze and grasp intent detection. Our findings reveal a robust neural association between these two signals, paving the way for novel insights into intent recognition during natural grasp tasks. This discovery holds promising implications for neuroprosthetics, rehabilitation, and human-machine interaction.

The observed consistency in performance between EEG and eye movement decoding across all subjects demonstrates that both modalities offer reliable information for accurate intent recognition. Furthermore, a Pearson correlation coefficient of 0.819 for all task pairs underscores the inherent coupling

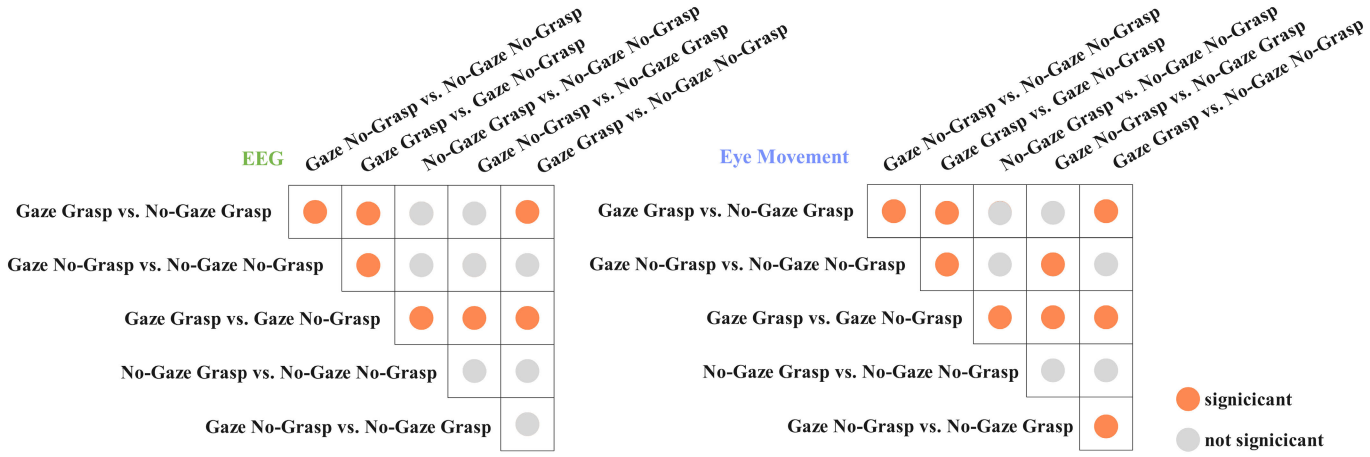

Fig. 8. Statistically significant differences between the task pairs' accuracy are assessed after multiple comparisons correction using the False Discovery Rate (FDR) method, with an $\alpha = 0.05$. The performance of EEG is presented in the left table, while the eye movements' performance is shown in the right table. The pairs with statistically significant differences are highlighted in orange circles.

between these two modalities, implying that they may share common underlying neural processes during gaze and grasp intent tasks. Our study also emphasizes the potential of employing suboptimal classifiers, such as KNN, to achieve strong coupling between EEG and eye movement decoding in natural grasp intent tasks. The results suggest that even with a lower-performing classifier, the majority of subjects still displayed robust coupling between the two modalities. This finding indicates that the intrinsic connection between EEG and eye movement signals is resilient enough to withstand variations in classification algorithms, which could hold significant implications for the development of real-world intent recognition systems.

Several limitations exist in the current work. First, the consistency between EEG and eye movements is reflected in the decoding ability of human intent. Consequently, the decoding capabilities depend on hardware and classification algorithms. As sensor technology and algorithms evolve, the accuracy of eye movement or EEG classification will improve. This may potentially lead to changes in the consistency between eye movement and EEG during natural grasping. However, our experiment demonstrates a certain degree of robustness in eye movement and EEG, even when using suboptimal classifiers. Second, the experimental results may be influenced by various factors. Although speech stimulation reduced visually-related interference, it still affected EEG signals to some extent. The target's shape, color, and other characteristics may impact the user's eye movements or EEG, such as grasp position characteristics (e.g., the handle of a water cup) because there are differences in gaze patterns and brain responses to different objects. These factors may affect the decoding performance of natural grasp intent.

In future research, we plan to explore several avenues to build upon the current findings and address unanswered questions. We aim to investigate and visualize the EEG and eye movement representations at the feature layer to better understand the relationship between the internal connection and feature fusion performance. Additionally, we will conduct a comparative analysis of the performance of feature layer fusion and decision fusion techniques to determine the optimal approach for enhancing the theoretical reference of EEG and eye movement fusion. By pursuing these research directions, we hope to advance our understanding of the interplay between EEG and eye movement signals and unlock their full potential in developing more accurate and reliable brain-computer interfaces.

VI. Conclusion

In conclusion, this study has made significant strides in uncovering the coupling correlation between EEG and eye movement signals during natural grasp intent estimation. By decoding six grasp intent pairs using a combination of feature vectors and an optimal classifier, our findings have demonstrated that the consistency between EEG and eye movement patterns is maintained when a user generates a natural grasp intent. Furthermorethe results indicate a robust connection between EEG and eye movements even when considering cortical EEG (from the visual cortex or motor cortex) and a suboptimal classifier. These insights have important implications for the development of wearable robotic applications, with the potential to improve the daily lives of individuals with disabilities. By enhancing our understanding of the neural correlation between EEG and eye movement signals, our work provides a valuable reference for future studies on intention estimation and lays the foundation for more accurate and reliable brain-computer interfaces.

References

[1] F. Lotte et al., "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[2] 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.

[3] B. J. B. Lee, A. Williams, and P. Ben-Tzvi, "Intelligent object grasping with sensor fusion for rehabilitation and assistive applications," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 8, pp. 1556–1565, Aug. 2018.

[4] P. D. Marasco et al., "Neurorobotic fusion of prosthetic touch, kinesthesia, and movement in bionic upper limbs promotes intrinsic brain behaviors," *Sci. Robot.*, vol. 6, no. 58, Sep. 2021, Art. no. eabf3368.

[5] C. Zhang, C. Lin, Y. Leng, Z. Fu, Y. Cheng, and C. Fu, "An effective head-based HRI for 6D robotic grasping using mixed reality," *IEEE Robot. Autom. Lett.*, vol. 8, no. 5, pp. 2796–2803, May 2023.

[6] T. Aflalo et al., "Decoding motor imagery from the posterior parietal cortex of a tetraplegic human," *Science*, vol. 348, no. 6237, pp. 906–910, May 2015.

[7] S. R. 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.

[8] J. Zou and Q. Zhang, "EyeSay: Brain visual dynamics decoding with deep learning & edge computing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2217–2224, 2022.

[9] 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, Jul. 2014.

[10] J. Pereira, P. Ofner, A. Schwarz, A. I. Sburlea, and G. R. Müller-Putz, "EEG neural correlates of goal-directed movement intention," *NeuroImage*, vol. 149, pp. 129–140, Apr. 2017.

[11] K. Zhang et al., "Foot placement prediction for assistive walking by fusing sequential 3D gaze and environmental context," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2509–2516, Apr. 2021.

[12] 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, Aug. 2020.

[13] M. Kim, B. H. Kim, and S. Jo, "Quantitative evaluation of a low-cost noninvasive hybrid interface based on EEG and eye movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 2, pp. 159–168, Mar. 2015.

[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] S. Li, X. Zhang, and J. D. Webb, "3-D-gaze-based robotic grasping through mimicking human visuomotor function for people with motion impairments," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2824–2835, Dec. 2017.

[16] Y. L. Cio, M. Raison, C. L. Ménard, and S. Achiche, "Proof of concept of an assistive robotic arm control using artificial stereovision and eye-tracking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 12, pp. 2344–2352, Dec. 2019.

[17] W. Gao et al., "Learning invariant patterns based on a convolutional neural network and big electroencephalography data for subject-independent P300 brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1047–1057, 2021.

[18] G. R. M′uller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an SSVEP-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 361–364, Jan. 2008.

[19] J.-H. Kim, F. Bießmann, and S.-W. Lee, "Decoding three-dimensional trajectory of executed and imagined arm movements from electroencephalogram signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 5, pp. 867–876, Sep. 2015.

[20] B. J. Edelman, B. Baxter, and B. He, "EEG source imaging enhances the decoding of complex right-hand motor imagery tasks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 4–14, Jan. 2016.

[21] B. H. Kim, M. Kim, and S. Jo, "Quadcopter flight control using a low-cost hybrid interface with EEG-based classification and eye tracking," *Comput. Biol. Med.*, vol. 51, pp. 82–92, Aug. 2014.

[22] Y. K. Meena, H. Cecotti, K. Wong-Lin, and G. Prasad, "Towards increasing the number of commands in a hybrid brain-computer interface with combination of gaze and motor imagery," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 506–509.

[23] E. Heremans, W. Helsen, and P. Feys, "The eyes as a mirror of our thoughts: Quantification of motor imagery of goal-directed movements through eye movement registration," *Behavioural Brain Res.*, vol. 187, no. 2, pp. 60–351, 2008.

[24] J. L. Collinger et al., "High-performance neuroprosthetic control by an individual with tetraplegia," *Lancet*, vol. 381, no. 9866, pp. 557–564, Feb. 2013.

[25] B. Wodlinger, J. E. Downey, E. C. Tyler-Kabara, A. B. Schwartz, M. L. Boninger, and J. L. Collinger, "Ten-dimensional anthropomorphic arm control in a human brain-machine interface: Difficulties, solutions, and limitations," *J. Neural Eng.*, vol. 12, no. 1, Feb. 2015, Art. no. 016011.

[26] B. Yang, J. Huang, X. Chen, X. Li, and Y. Hasegawa, "Natural grasp intention recognition based on gaze in human–robot interaction," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 4, pp. 2059–2070, Apr. 2023.

[27] F. Koochaki and L. Najafizadeh, "A data-driven framework for intention prediction via eye movement with applications to assistive systems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 974–984, 2021.

[28] S. Cheng, J. Wang, L. Zhang, and Q. Wei, "Motion imagery-BCI based on EEG and eye movement data fusion," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2783–2793, Dec. 2020.

[29] X. Ge et al., "Improving intention detection in single-trial classification through fusion of EEG and eye-tracker data," *IEEE Trans. Human-Mach. Syst.*, vol. 53, no. 1, pp. 132–141, Feb. 2023.

[30] F. Hou et al., "EEG-based emotion recognition for hearing impaired and normal individuals with residual feature pyramids network based on time–frequency–spatial features," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[31] Y. Li, L. Guo, Y. Liu, J. Liu, and F. Meng, "A temporal-spectral-based squeeze-and- excitation feature fusion network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1534–1545, 2021.

[32] W.-Y. Hsu and Y.-W. Cheng, "EEG-channel-temporal-spectral-attention correlation for motor imagery EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1659–1669, 2023.

[33] E. H. Hess and J. M. Polt, "Pupil size in relation to mental activity during simple problem-solving," *Science*, vol. 143, no. 3611, pp. 1190–1192, Mar. 1964.

[34] 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.