Neural Profiling with fNIRS of Operator Performance in Teleoperated Human-like Social Robot Interactions

David Achanccaray and Javier Andreu-Perez and Hidenobu Sumioka

***Abstract*— Social robot teleoperation is a skill that must be acquired through practice with the social robot. Mobile neuroimaging and human-computer interface performance metrics permit the gathering of information from the operators' systemic and behavioral responses associated with their skill acquisition. Profiling the skill levels of social robot operators using this information can help improve training protocols. In this study, thirty-two participants performed real-world social robot teleoperation tasks. Brain function signals from the prefrontal cortex (PFC), and behavioral data from interactions with the system were collected using functional near-infrared spectroscopy (fNIRS). Participants were divided into two groups (high and low performance) based on an integrative metric of task efficiency, workload, and presence when operating the social robot. Significant differences were found in the operation time, width, and multiscale entropy of the hemoglobin oxygenation curve of the operator's PFC. Functional connectivity in the PFC also depicted differences in the low- and high-performance groups when connectivity networks were compared and in the leaf fraction metrics of the functional networks. These findings contribute to understanding the operator's progress during teleoperation training protocols and designing the interface to assist in enhancing task performance.**

***Index Terms*— Neural profile, fNIRS, teleoperation, functional connectivity, social interaction.**

I. Introduction

Social interactions shape human behavior. However, limitations, such as mobility, safety, or cultural context, might affect people's social development. Teleoperation of a humanoid robot could be an alternative to overcome these issues for a dyadic (robotic avatar-human) interaction. Nevertheless, the operator's performance is influenced by human factors, such as emotions, mental state, and workload [1]. These factors can be reflected in brain activity due to the involvement of cognitive and affective processes during social interactions [2]. A neural analysis to determine the operator's performance profile might provide an understanding of neural mechanisms related to teleoperation skills [3].

*This work was supported by JST Moonshot R&D Grant Number JPMJMS2011 (conducting the experiment) and JSPS Bilateral Collaborations Grant Number JPJSBP120245710 (data analysis).

David Achanccaray is with the Deep Interaction Lab, Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International, Kyoto 619-0228, Japan and the Center of Neuroergonomics, ISAE-SUPAERO, Université de Toulouse, Toulouse 31055, France. e-mail: david.achanccaray@isae-supaero.fr

Javier Andreu-Perez is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4-3SQ, UK. He also had a JSPS Invitational Fellowship for Research in Japan. e-mail: j.andreu-perez@essex.ac.uk

Hidenobu Sumioka is with the Deep Interaction Lab, Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International, Kyoto 619-0228, Japan. e-mail: sumioka@atr.jp

There are many techniques to record brain activity, such as electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). EEG is the most used non-invasive technique owing to its high temporal resolution and affordable cost. fNIRS is a promising non-invasive technique owing to its high spatial resolution. In addition, fNIRS is more robust compared to EEG for motor artifacts and makes it suitable for recording the operator's brain activity [3]. This technology measures the concentration changes of hemoglobin oxygenation (HbO) and hemoglobin deoxygenation (HbR) at multiple cortex locations [4]. Activated brain regions reach an increase in HbO and a decrease in HbR. This effect is known as the hemodynamic response [5].

Recently, technological advances in fNIRS have allowed the analysis of relationships between brain activity and cognitive processes in virtual and outside environments. These applications were focused on naturalistic environments (real-world situations) and social cognitive tasks. The prefrontal cortex (PFC) is known as the brain's cognitive center [3]. The PFC is a large brain region that supports cognitive processes related to voluntary and self-initiated behavior in response to situations with no predetermined limits. These processes are attention control, concentration, decision-making, social processing, and anxiety. They are affected by stress and workload. fNIRS fits effectively with studies about monitoring PFC activity due to its practical features [5], [6].

In this work, we decoded the fNIRS-based neural profile of the operator's performance during teleoperated social interactions by fNIRS data features and functional connectivity in PFC. These interactions were contextualized in the reception of a clinic. In addition, performance metrics, workload, and presence were assessed. This profiler can be useful in human-robot interaction to estimate the level of the operator's trust. It might also trigger some autonomous control and correct the operator if necessary. The Main Highlights of this work are the following:

- To the best of our knowledge, the first study of neuromonitoring of the hemodynamic response in PFC of the operators' performance in a social scenario using a humanoid robotic avatar.
- Significant differences and relationships in robot interaction metrics and PFC responses were found among high- and low-performing operators.

This article proceeds as follows: in section II, related work is outlined; in section III, the materials and methods are detailed; in section IV, the results of the analysis are elucidated; in section V, these results are discussed and contextualized

with the current literature; in section VI, the limitations are mentioned; and finally, section VII summarizes the findings and concludes the article.

II. Related Works

Several cognitive processes and mental state detection have been approached based on fNIRS in the PFC. The most common features extracted from fNIRS data were the mean, variance, skewness, kurtosis, slope, amplitude, and zero-crossing [2]. The power spectrum estimation was a neuromarker of deception (spontaneous) behavior during a poker game. This demanding task under the spontaneous behavior reached higher power than the controlled behavior [7]. In addition, mental stress was detected by the power spectrum under temporal demand and a modified Stroop color word task [8]. Another feature of the hemodynamic response is the laterality index, which was used to distinguish the short-term effect of meditation in attention control during the Stroop color word task [9]. Neural synchrony of a team of anesthesiologists was calculated based on HbO/HbR wavelet transform coherence to measure team engagement in a simulation of crisis event management training [10]. Other studies evaluated the task complexity based on HbO levels. Increases in the HbO level were related to the task complexity [11]. The HbO level was higher in the combined gait and cognitive tasks than in the walking task alone. The multiscale entropy of HbO was a complexity index in a visual memory task [12]. Participants watched scenes and were asked to choose if they wanted to remember or forget each scene. Then, they performed a scene recognition test to verify their choice. The HbO mean was also an indicator of recognizing 10 positive emotions clustered in 3 groups (playfulness, encouragement, and harmony) [13]. These emotion groups were recognized by using binary classifiers.

Cognitive processes require synchronized information flow within specialized functional networks in the brain. This mechanism underlying brain coordination is known as functional connectivity [14]. Cognitive tasks, such as a pattern recognition test, increased functional connectivity based on the hemodynamic response compared to the resting state. Network metrics distinguished the mental workload of both states [15]. In surgical tasks, HbO levels were attenuated in senior surgeons. Brain prefrontal-motor connectivity was more resilient in senior surgical residents under temporal demand [6]. Regarding teleoperation, the latency affected functional connectivity based on fNIRS [3]. The latency increased mental demand, anxiety, and complexity in operators during motion planning and control tasks; thus, functional connectivity increased within and between the PFC and motor cortex. This effect was mitigated with haptic feedback by reducing the average functional connectivity in all networks.

Performance and expertise prediction can carried out using fNIRS-based brain activity metrics in the PFC. The difference between HbO and HbR levels predicted the participant's performance in a variant go-stop task and different workload conditions, which incorporated 3 different types of trials (Go, Ignore, and Inhibit) [16]. This PFC marker reached better prediction results than the past performance. The PFC activity reveals relevant information related to task strategy and skill acquisition. In a gaming context (comparable to teleoperation), a reduced HbO level in the cortex indicated an attentional demand increase. This HbO level decrease was also a successful performance indicator on attending-demanding distraction tasks [17]. Another gaming work estimated the gamer's expertise based on spectroscopy time series and emotion prediction from facial expressions by training a multimodal supervised deep-learning model [18]. This framework might support the game design based on the gamer's neural response. Thus, fNIRS data features and network metrics provide a set of biomarkers to model a neural profile. They could estimate the operator's expertise or performance level in teleoperated social tasks.

III. Materials and Methods

In the following section, we will outline the participant cohort (III-A) used for this experiment, the hardware (III-B), and the experimental settings (III-C). This will be followed by the performance and physiological indices used (sections III-D and III-E) in the analysis (III-F).

A. Participants

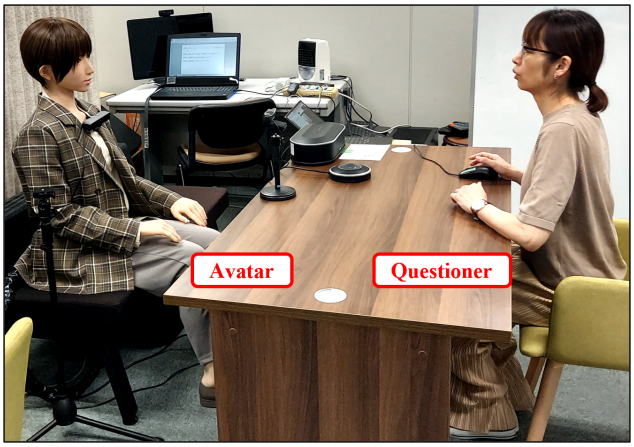
Thirty-two healthy subjects, 18 males and 14 females, participated in this experiment. Participants were between 19 and 36 years old (Mean = 23.16 and SD = 4.27) and did not belong to a specific working sector. All the participants were right-handed. They had no background in neurological diseases and no experience in teleoperation systems. Before the experiment, each participant signed an informed consent form. This document was approved by the Ethical Committee of the Advanced Telecommunications Research (ATR) Institute International (Japan). In addition, participants received an economic compensation of ¥/.4,000 (US$/.27.00 approximately).

B. Experimental Setup

All the participants carried out experiments using the humanoid robot "Erica" for teleoperated social tasks [19]. Participants manipulated the avatar "Erica" by a teleoperation interface in a local environment (Fig. 1a). An individual (questioner) interacted with each participant through the robotic avatar in a remote environment (Fig. 1b). Erica's teleoperation interface allows her to perform facial expressions of emotions, such as sad (negative) and happy (positive). This interface also allows her to execute hand (left and right) movements, such as palm up, pointing, palm side, and raising hand. In addition, the teleoperation interface has an auto-option to express emotions based on the embedded audio and image processing.

The brain activity was recorded by a fNIRS device WOT-HS (NeU Corporation). This device acquired the hemoglobin response from 34 optodes and was mounted in the participant's PFC.The space between optodes were 21 mm and 32

(a)



(b)

Fig. 1. Teleoperation of the robotic avatar "Erica" for performing social tasks. (a) The operator was manipulating the humanoid robotic avatar "Erica" through the teleoperation interface in the local environment. (b) The questioner was interacting with the robotic avatar in the remote environment.

mm. The wavelengths of LED[1] sources were 730 nm and 855 nm and the sampling rate was 10 Hz.

The teleoperation interface was shown on a 24-inch touchscreen, which eased the avatar Erica's manipulation. Another 24-inch screen was used in the remote environment, which showed the question list to the questioner. Each screen was connected to a PC. Both PCs had the following features: Windows 11 operating system, Intel Core i7-11700 CPU at up to 4.9 GHz, 32 GB RAM, and Nvidia GeForce RTX 3070 GPU. An interface for synchronizing the social task stages was developed in Python and superposed to Erica's teleoperation interface. This interface ran on both PCs, captured the mouse's events caused by the operator and questioner, and sent this data by the lab streaming layer (LSL[2]) system. Additionally, this interface was communicated by TCP to the open app LabRecorder[3]. LabRecorder received and saved the data streamed through the LSL system.

C. Experimental Procedure

The experiment consisted of performing communication tasks. This social task had verbal and non-verbal sections as a face-to-face interaction between two persons. Participants teleoperated a receptionist robot (avatar) to answer questions (verbal section) and carry out actions (non-verbal section). These actions were emotions and hand movements.

TABLE I. BEHAVIORAL RULES (NON-VERBAL) FOR THE OPERATOR BY THE TYPE OF QUESTIONS.

Type of Question	Weight	Emotion	Hand Movement
Location	2.0	Auto	Pointing
Health insurance validity	4.5	Positive/Negative	Palm side
Availability	2.5	Auto	Palm up
Complaint	3.0	Auto	Palm side
Greeting	1.0	Positive	Raise hand

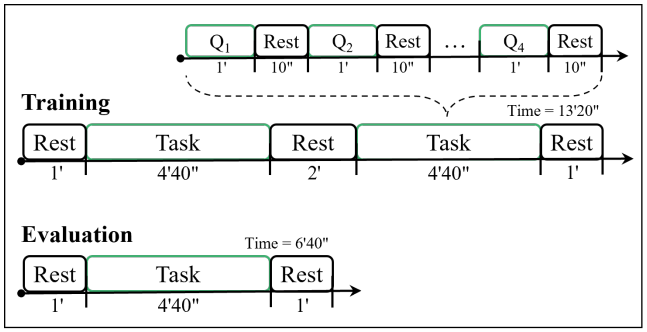


Fig. 2. Timeline of the training and evaluation sessions. The social task consisted of 4 questions (Q_1, ..., Q_4) asked by the questioner to the operator. Then, the operator answered each question and executed actions.

Each participant performed a training session and an evaluation session (Fig. 2). Before the training session, participants took a working memory test (digit span memory test). They had to memorize sequences of 3, 4, 5, 6, and 7 digits [20]. We used the Digit Span Tester[4] to evaluate the participants. Then, they memorized the behavioral rules (Table I) for 2 minutes and were evaluated with a test for 1 minute. The ATR lab staff that assisted during the experiments rated the complexity of questions between 1 and 5. These mean scores were assigned as question weights. Next, participants practiced the social task until they understood the procedure.

The training session consisted of two blocks of social tasks and three blocks of resting. Each social task block had four questions and a break between them. The questioner asked the participant these questions. Then, the participants responded to them verbally and non-verbally. Participants answered the questions using the information located next to the touchscreen (Fig. 1a); the questioner asked for doctors' availability, health insurance, department's location, complaints, and greetings in the context of a clinic's reception. The question lists were created randomly and shown on a screen to the questioner. After the training session, participants answered workload and presence tests. The NASA-TLX workload test [21], which rated six task load subscales (mental demand, physical demand, temporal demand, performance satisfaction, effort, frustration) between 1 and 10 and provided an overall workload score. The presence test was the adapted Slater-Usoh-Steed (SUS) questionnaire [22], [23] to our study. The SUS test consists of six questions ($q_i, i = 1, ..., 6$) rated on a scale of 1 to 7, and the mean score pro-

[1]LED: Light-emitting diode.
[2]LSL: https://github.com/sccn/labstreaminglayer
[3]LabRecorder: https://github.com/labstreaminglayer/App-LabRecorder
[4]Digit Span Tester: https://sourceforge.net/projects/digits-pantester

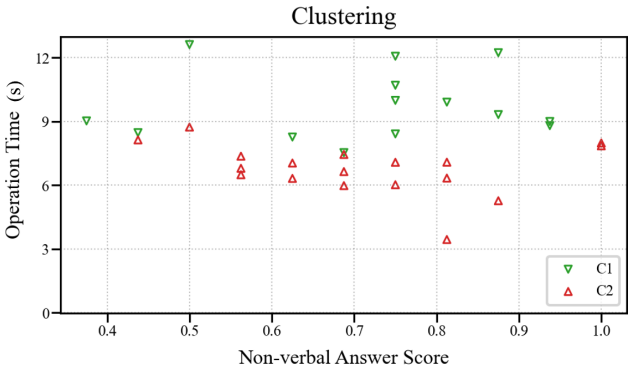


Fig. 3. GMM Clustering of performance metrics into C1 and C2 groups. These groups were associated with low and high performance, respectively.

vides an overall presence score. This test assessed the feeling of being in a virtual environment (q_{P1} = mean(q_1, q_4)), if it becomes the dominant reality (q_{P2} = mean(q_2, q_6)), and if it is remembered as a real place (q_{P3} = mean(q_3, q_5)).

The evaluation session had one block of social tasks and two blocks of resting. In addition, participants watched a black screen during the resting blocks of both sessions.

D. Performance Metrics

We considered 3 metrics for assessing the operator's performance. These metrics were the operation time, verbal answer score, and non-verbal answer score. The operation time (OT) is the total time of the operator's response. It was measured from the questioner ending the question until the operator finished the verbal and non-verbal answers. Then, the normalized operation time (OTn) was calculated as shown in Equation 1. The verbal answer score was 0 or 1 if the provided information by the operator was wrong or correct, respectively. The non-verbal answer score was 0 if the executed actions (emotion and hand movement) were wrong, 0.5 if an executed action was correct at least, and 1 if both executed actions were correct.

$$OTn_i = OT_i / W_i, \tag{1}$$

where, i is the number of question in a session, OTn_i is the normalized OT_i, and W_i is the question's weight according to Table I.

Next, a Gaussian mixture model (GMM) was trained for clustering using the training session data. This GMM separated the participants into 2 groups (Fig. 3), which we searched to associate with low- and high-performance. The GMM had 2 components and was trained with the mean OTn, mean verbal answer score, mean non-verbal answer score, working memory test score, workload test score, and presence test score of each participant. The Davies-Bouldin clustering index was 0.798. This index indicated that the clusters were well-separated.

E. Hemodynamic Response

The HbO and HbR signals for 34 channels (optodes) were computed from fNIRS data by using the modified Beer–Lambert Law [24]. Both signals were filtered to remove artifacts by a six-order Butterworth bandpass filter with cutoff frequencies of 0.05 Hz and 0.4 Hz. The signal-to-noise ratio (SNR) was evaluated in all signals; then, signals with SNR below 15 were discarded. Next, the HbO and HbR signals were segmented by finding the first positive peak during the operation time, in which the time limits of this peak defined the time interval for signal segments. Then, we computed several features of both signals. These features were the amplitude peak (maximum), mean, variance, skewness, kurtosis, width (or period), area under the curve (AUC), laterality index, and multiscale entropy. The mean of each feature for valid channels was used in the analysis. Next, the features of both GMM groups were compared statistically.

F. Analysis

Performance groups were statistically compared using performance metrics, test scores, fNIRS features, and connectivity network metrics. We applied the Shapiro–Wilk (S–W) test to verify the normal distribution of features. When the normal distribution was verified, a post hoc test Tukey's Honest Significant Difference was used for pairwise comparisons. The significance level was 5%. When the data deviated from a normal distribution, non-parametric statistical tests analyzed it. We adopted a non-parametric Mann-Whitney U (MWU) test to find the statistically significant differences between pairwise comparisons. The significance level was also 5%. In addition, the Spearman correlation with a significance level of 5% was applied to find associations between metrics. In this analysis, correlation (ρ) levels were defined, such as strong ($0.7 \leq |\rho| \leq 1$), moderate ($0.4 \leq |\rho| < 0.7$), and weak ($0.1 \leq |\rho| < 0.4$).

The functional connectivity analysis was carried out by calculating the adjacency matrix using the weighted phase lag index (WPLI) metric. WPLI is a more robust version of the phase lag index (PLI) metric to overcome field spread, noise, and sample-size bias issues. WPLI and PLI estimate the phase difference between signals for a specific frequency band [25]. In brain network analysis by graph theory, it is important to identify relevant connections. Usually, a threshold of connection weights is used. However, choosing an arbitrary threshold yields poor reproducibility and limited interpretation of results. A solution to the "thresholding problem" is the Minimum spanning tree (MST), which is a method that generates a subgraph based on the minimization of the sum of the weights [26]. The MST is unique, reliable, and reproducible. In addition, conventional network topology metrics can be applied to characterize the MST. We computed graph metrics in node-wise, edge-wise, and network-wise categories of low- and high-performance group networks [27]–[30]. All metrics are summarized in Table II.

IV. Results

The results are presented as follows: the statistical results from the questionnaires are detailed in section IV-A. In section IV-B, we outline the differences among various hemodynamic signal descriptors, concluding with an analysis of functional connectivity in the PFC (section IV-C).

TABLE II. SUMMARY OF METRICS.

Type	Metrics	Definition
Performance	Normalized operation time	Time to complete a response
	Verbal answer score	Right question's answer score
	Non-verbal answer score	Right behavioral action score
Workload Test	Mental demand, Physical demand, Temporal demand, Performance satisfaction, Effort, Frustration	Self-reported rate of NASA-TLX workload test's subscales
	Overall workload	Overall rate
Presence Test	Six questions	Self-reported rate of SUS presence test's questions
	Subscales: q_{P1}, q_{P2}, q_{P3}	Mean of score pairwise
	Overall presence	Mean rate
HbO Signal Features	Amplitude peak	Positive peak's maximum
	Mean, Variance, Kurtosis, Skewness	Statistical metrics
	Width	Duration of the positive peak
	AUC	Area under the positive peak
	Laterality index	Hemispheric dominance
	Multiscale entropy	Complexity of the signal
Node-Wise Functional Connectivity	Degree	Number of links of a node
	Betweenness centrality	Normalized number of shortest paths between two nodes
Edge-Wise Functional Connectivity	Network-based statistic	Identification of common connections or patterns between groups under different conditions
Network-Wise Functional Connectivity	Diameter	Normalized largest distance between two nodes
	Leaf fraction	Normalized leaf number, where the leaf number is the number of nodes with only one connection
	Tree hierarchy	Characterization of an optimal topology
	Kappa	Broadness measure of the degree distribution

A. Performance Metrics and Test Scores

The comparison of performance metrics between groups in the training session is shown in Fig. 4. Tukey's test revealed statistical differences between groups in the mean OTn (p<0.01). There were no statistical differences between groups in the evaluation session. Regarding the workload test, Tukey's test found statistical differences between groups in the mental demand subscale (p<0.05), which is shown in Fig. 5. There were no statistical differences between groups in the presence test.

The correlations between performance metrics, workload test scores, and presence test scores are shown in Table III.

B. fNIRS Features

The fNIRS features for both groups in the training session are shown in Fig. 6a and Fig. 6b. Tukey's test revealed statistical differences between groups in HbO-width (p<0.05) and HbO-multiscale entropy (p<0.05).

The correlations between performance metrics/test scores and fNIRS features are shown in Table IV. No statistically significant differences were found among any fNIRS features in the HbR chromophore. The correlations between fNIRS features are shown in Table V. Additionally, the GMM clustering of participants was used to combine performance metrics and fNIRS features as HbO-width vs. normalized operation time, which is shown in Fig. 7. This scatter plot showed the separability of groups and the potential of HbO-width as a performance indicator.

TABLE III. CORRELATIONS BETWEEN PERFORMANCE METRICS AND TEST SCORES.

Metrics	Metrics	ρ
Verbal answers	Effort	-0.43*
Non-verbal answers	Working memory	0.38*
Behavioral rules	Physical demand	0.35*
Working memory	Frustration	0.64**
Mental demand	Presence q_{P3}	0.37*
Overall workload	Mean presence	-0.39*

*p<0.05, **p<0.01.

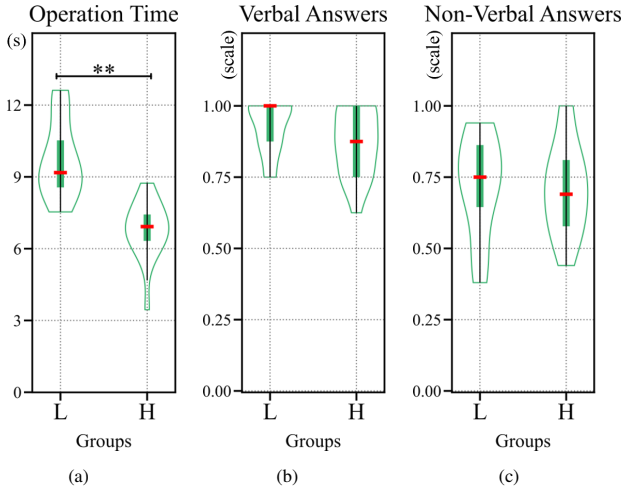


Fig. 4. Comparison of performance metrics from participants with low (L) and high (H) performance in the training session. *p<0.05, **p<0.01, Median: red line. (a) [Mean normalized operation time. (b) Mean verbal answer score (c) Mean non-Verbal answer score.

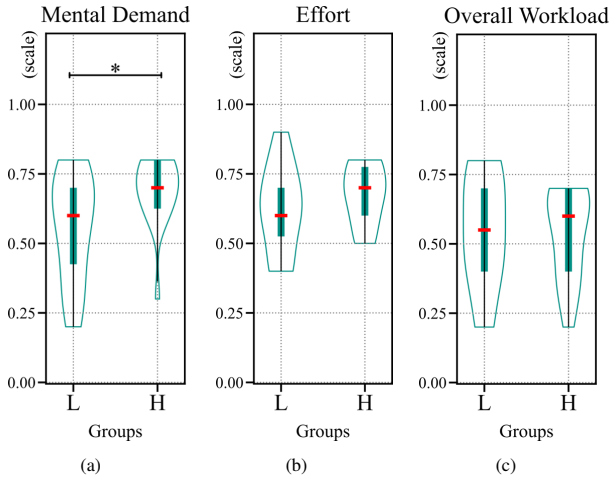


Fig. 5. Comparison of workload subscales and overall workload from participants with low (L) and high (H) performance. *p<0.05, **p<0.01, Median: red line. (a) Mental demand. (b) Effort. (c) Overall workload.

C. Functional Connectivity

The degree for low- and high-performance groups is shown in Fig. 8. The node size depended on the number of connections. HbO was used to compute functional connectivity, as statistical differences were found exclusively with this chromophore (sec. IV-B). The relevant brain areas had more connections during the task performance. These regions were concentrated in the PFC left region for the low-performance group, and in the PFC right region and lower left for the

TABLE IV. CORRELATIONS BETWEEN PERFORMANCE METRICS / TEST SCORES AND FNIRS FEATURES.

Metrics	Features	ρ
Normalized operation time	HbO-Mean	-0.35*
Normalized operation time	HbO-MSEntropy	0.39*
Behavioral rules	HbO-Mean	0.38*
Behavioral rules	HbO-Variance	0.39*
Effort	HbO-Width	-0.40*
Effort	HbO-MSEntropy	0.45**
Presence q_{P3}	HbO-LI	-0.39*

*$p<0.05$, **$p<0.01$.
LI: Lateralization index, MSEntropy: Multiscale entropy.

TABLE V. CORRELATIONS BETWEEN FNIRS FEATURES.

Features	Features	ρ
HbO-Amplitude peak	HbO-Mean	0.87**
HbO-Amplitude peak	HbO-Variance	0.90**
HbO-Amplitude peak	HbO-Width	0.37*
HbO-Amplitude peak	HbO-AUC	0.90**
HbO-Mean	HbO-Variance	0.77**
HbO-Mean	HbO-Width	0.45*
HbO-Mean	HbO-AUC	0.87**
HbO-Mean	HbO-MSEntropy	-0.45*
HbO-Variance	HbO-AUC	0.86**
HbO-Kurtosis	HbO-MSEntropy	-0.45*
HbO-Width	HbO-AUC	0.61**
HbO-Width	HbO-MSEntropy	-0.82**
HbO-AUC	HbO-MSEntropy	-0.53**

*$p<0.05$, **$p<0.01$. MSEntropy: Multiscale entropy.

high-performance group.

The betweenness centrality for low- and high-performance groups is shown in Fig. 9. The node size depended on the betweenness centrality value. In addition, the color map showed subnetworks in the PFC. The number of relevant nodes for global communication ("hub nodes") and subnetworks found in the PFC were greater for the low-performance group than for the high-performance group.

The network-based statistic found the connectivity patterns (dark blue squares) of the networks of low- and high-performance groups (Fig. 10). This figure also showed disconnections between groups in most PFC network edges. Color masks were superposed to highlight the intra- and inter-hemisphere relationships between PFC regions.

The network-wise metrics for both performance groups are shown in Fig. 6c. Tukey's test revealed statistical differences between groups in leaf fraction ($p<0.05$). The "hub nodes" had a more relevant role in network communication in the high-performance group than in the low-performance group.

V. DISCUSSION

In this work, we have demonstrated that performance metrics and physiological data can be helpful in discriminating between operators of a social robot while operating the robot in a real scenario. In terms of performance scores, the operation time of actuating the robot was significantly lower for the low-performance group versus the high-performance group. Operators who took longer to actuate the robot generally produced more incorrect answers. Participants were not informed in real-time that they made a mistake in order to avoid post-error behaviors such as cognitive tunneling effects. Studies have shown that errors can increase under high cognitive load or multitasking, especially when tasks are performed too slowly, leading to lapses in attention and

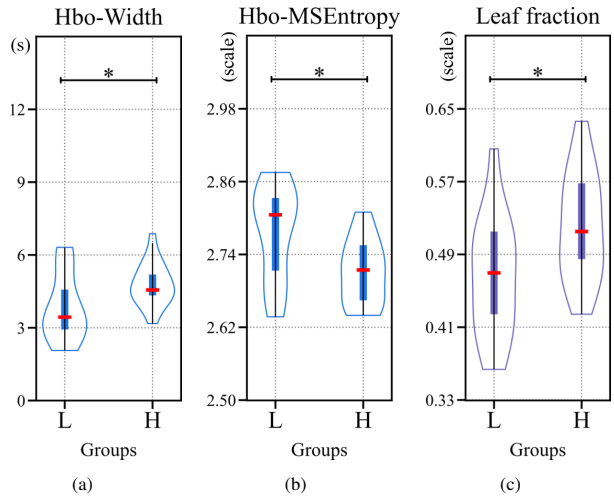


Fig. 6. Comparison of HbO features and MST metrics of functional connectivity from participants with low (L) and high (H) performance in the training session. *$p<0.05$, **$p<0.01$, Median: red line. (a) Hbo-Width. (b) Hbo-Multiscale entropy (MSEntropy). (c) MST Leaf fraction.

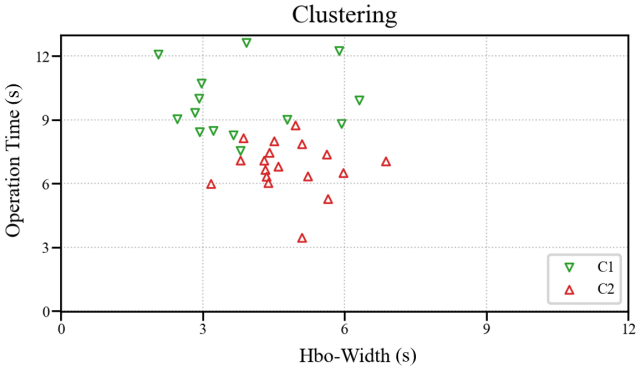


Fig. 7. Scatter plot of HbO-width vs Operation time in the training session.

performance [31]. In a dual-task experiment, Strobach et al. [31] showed that prolonged engagement with a task due to multitasking can increase errors due to divided attention and reduced focus on individual tasks. It is possible that for social robot operators, the uncertainty in the task can induce general anxiety to maintain engagement with the interlocutor [32], resulting in a higher response time and a more rushed response when in doubt.

Two metrics were found to be significant in discriminating both groups regarding the physiological prefrontal brain response. The first significant difference is the width of the HbO with the response, which is higher in the high-performance group than in the lower group. Research using fNIRS showed that during decision-making tasks, such as the Iowa Gambling Task and the Balloon Analog Risk Task [33], there were significant increases in HbO in the PFC. Also, the PFC showed more significant hemodynamic changes in free-choice decision-making. This may indicate that the PFC is prominently recruiting in an effort to make voluntary decisions and subsequent actions [34]. We noticed that the higher-performance group exhibited a prolonged HbO response and higher entropy, suggesting greater PFC engagement as expected for this type of task. In contrast, the lower-performance group did not engage the PFC as prominently.

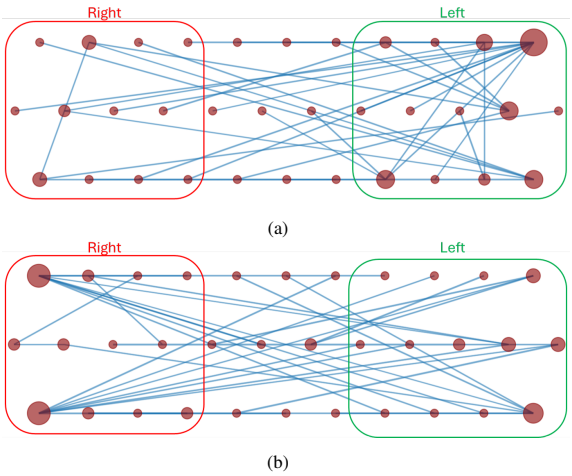


Fig. 8. Degree of functional of the connectivity network based on fNIRS. (a) Low-performance group. (b) High-performance group.

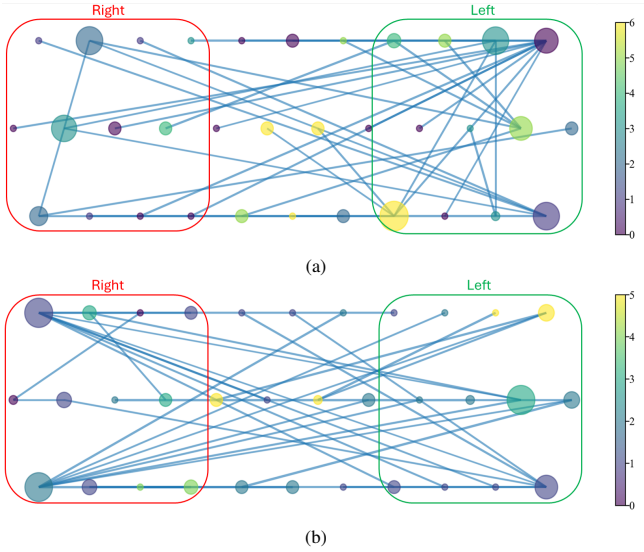


Fig. 9. Betweenness centrality of the functional connectivity network based on fNIRS. (a) Low-performance group. (b) High-performance group.

Lastly, multivariate connectivity metrics revealed differences in PFC functional networks, particularly in the leaf fraction metric. Although a lower leaf fraction has frequently been associated with higher cognitive performance [35], in decision-making scenarios, increased specialized connectivity, indicating a higher leaf fraction, has been observed to support executive functions required for making decisions, such as evaluating options and inhibiting impulsive responses [36]. This may explain the finding that higher-performing individuals exhibit a higher leaf fraction. Additionally, participants did not feel excessively loaded according to the NASA-TLX test (section IV-A), where a more integrated network might have been beneficial in contexts of higher cognitive load. Nevertheless, although this pattern was more prominent in the trial-based data, the PFC is likely to undergo network task adaptation, shifting between higher and lower leaf fraction states as needed, which has been suggested to be effective for decision-making [35]. For example, before deliberating on potential responses for the social robot, the operator's brain might shortly increase integration (lower leaf fraction) to process information more holistically. In

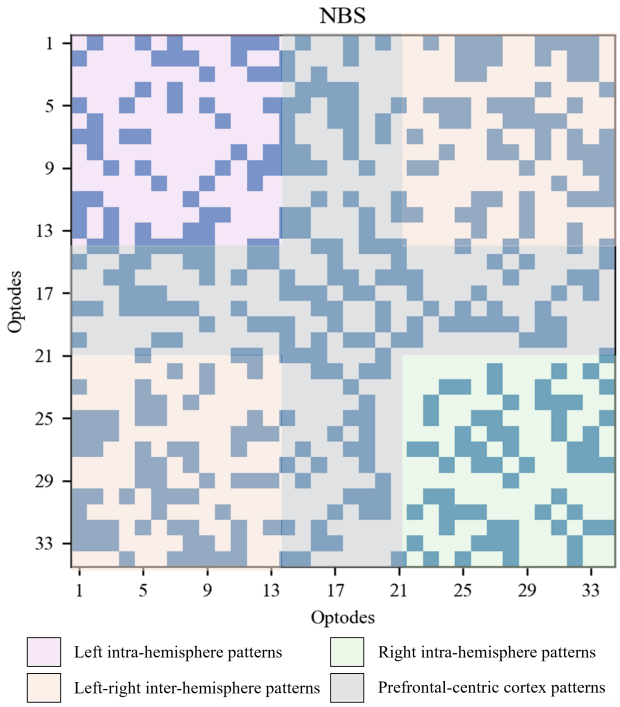


Fig. 10. Network-based statistic (NBS) of functional connectivity between performance groups based on fNIRS signals from 34 optodes. Intra- e Inter-hemisphere relationships between PFC regions are highlighted.

contrast, during the execution of a decision, more specialized connectivity (higher leaf fraction) may predominate to ensure the robot is controlled in the correct sequence. For this hypothesis, future work could focus on exploring these reconfiguration dynamics.

VI. Limitations

The study's limitation is that it only monitored the PFC area. While it was sufficient for the profiling objectives, investigating the interaction between different brain areas could have revealed other ways of profiling and led to a better understanding of how specific cognitive functions, such as motor skills and sensory information processing, influence the operator's profile. Future research will involve expanding the range of sensor modalities and their coverage. It will also include introducing interventions to enhance the learning abilities of robotic operators based on this profiling data.

VII. Conclusions

An experiment of operator control of a social robot in a real task was performed. The experiment consisted of a simulation of a clinical receptionist where a participant requested some information, and the operator needed to operate the robot with an indicated protocol that the operator needed to remember. The operator interaction with the robot and cognitive responses of the PFC captured by fNIRS were monitored. Participants were analytically clustered into two groups: low and high performance, based on the integration of their performance metrics and self-reported workload and presence rates. The results suggest univariate differences in HbO levels in signal width and multiscale entropy. Additionally, the multivariate channel-wise functional connectivity metric leaf fraction was found to be different

in low- and high-performing operators. These findings are valuable for understanding training protocols for operators and for enhancing human-computer interaction in human-robot docking and mutual or assisted operation. Specifically, the robot's ability to trust operator commands based on their estimated performance level can be improved, contributing to more effective and reliable control systems. Future research could focus on dynamic or real-time performance measures, as well as integrating neuromodulation techniques to enhance the operators' knowledge acquisition and training processes.

REFERENCES

[1] D. Rea, S. Seo, and J. Young, "Social robotics for nonsocial teleoperation: Leveraging social techniques to impact teleoperator performance and experience," *Current Robotics Reports*, vol. 1, 12 2020.

[2] N. Naseer and K.-S. Hong, "fnirs-based brain-computer interfaces: a review," *Frontiers in Human Neuroscience*, vol. 9, p. 3, 2015.

[3] Y. Ye, T. Zhou, Q. Zhu, W. Vann, and J. Du, "Brain functional connectivity under teleoperation latency: a fnirs study," 2023.

[4] X. Yin, B. Xu, C. Jiang, Y. Fu, Z. Wang, H. Li, and G. Shi, "A hybrid BCI based on EEG and fNIRS signals improves the performance of decoding motor imagery of both force and speed of hand clenching," *Journal of Neural Engineering*, vol. 12, p. 036004, apr 2015.

[5] P. Pinti, I. Tachtsidis, A. Hamilton, J. Hirsch, C. Aichelburg, S. Gilbert, and P. W. Burgess, "The present and future use of functional near-infrared spectroscopy (fnirs) for cognitive neuroscience," *Annals of the New York Academy of Sciences*, vol. 1464, no. 1, pp. 5–29, 2020.

[6] F. Deligianni, H. Singh, H. N. Modi, S. Jahani, M. Yucel, A. Darzi, D. R. Leff, and G. Z. Yang, "Expertise and task pressure in fnirs-based brain connectomes," 2020.

[7] J. Zhang, J. Zhang, H. Ren, Q. Liu, Z. Du, L. Wu, L. Sai, Z. Yuan, S. Mo, and X. Lin, "A look into the power of fnirs signals by using the welch power spectral estimate for deception detection," *Frontiers in Human Neuroscience*, vol. 14, 2021.

[8] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, "Mental stress analysis using the power spectrum of fnirs signals," in *2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 216–219, 2022.

[9] M. Izzetoglu, P. A. Shewokis, K. Tsai, P. Dantoin, K. Sparango, and K. Min, "Short-term effects of meditation on sustained attention as measured by fnirs," *Brain Sciences*, vol. 10, no. 9, 2020.

[10] J. Xu, J. M. Slagle, A. Banerjee, B. Bracken, and M. B. Weinger, "Use of a portable functional near-infrared spectroscopy (fnirs) system to examine team experience during crisis event management in clinical simulations," *Frontiers in Human Neuroscience*, vol. 13, 2019.

[11] A. Mirelman, I. Maidan, H. Bernad-Elazari, F. Nieuwhof, M. Reelick, N. Giladi, and J. M. Hausdorff, "Increased frontal brain activation during walking while dual tasking: an fnirs study in healthy young adults," *Journal of neuroengineering and rehabilitation*, vol. 11, pp. 1–7, 2014.

[12] T. Angsuwatanakul, J. O'Reilly, K. Ounjai, B. Kaewkamnerdpong, and K. Iramina, "Multiscale entropy as a new feature for eeg and fnirs analysis," *Entropy*, vol. 22, no. 2, 2020.

[13] X. Hu, C. Zhuang, F. Wang, Y.-J. Liu, C.-H. Im, and D. Zhang, "fnirs evidence for recognizably different positive emotions," *Frontiers in Human Neuroscience*, vol. 13, 2019.

[14] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers in Systems Neuroscience*, vol. 9, 2016.

[15] F. Racz, P. Mukli, Z. Nagy, and A. Eke, "Increased prefrontal cortex connectivity during cognitive challenge assessed by fnirs imaging," *Biomedical Optics Express*, vol. 8, pp. 3842–3855, 07 2017.

[16] H. Ayaz, A. Curtin, J. Mark, A. Kraft, and M. Ziegler, "Predicting future performance based on current brain activity: An fnirs and eeg study," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 3925–3930, 2019.

[17] M. J. Campbell, A. J. Toth, A. P. Moran, M. Kowal, and C. Exton, "Chapter 10 - esports: A new window on neurocognitive expertise?," in *Sport and the Brain: The Science of Preparing, Enduring and Winning, Part C* (S. Marcora and M. Sarkar, eds.), vol. 240 of *Progress in Brain Research*, pp. 161–174, Elsevier, 2018.

[18] A. R. Andreu-Perez, M. Kiani, J. Andreu-Perez, P. Reddy, J. Andreu-Abela, M. Pinto, and K. Izzetoglu, "Single-trial recognition of video gamer's expertise from brain haemodynamic and facial emotion responses," *Brain Sciences*, vol. 11, no. 1, 2021.

[19] D. F. Glas, T. Minato, C. T. Ishi, T. Kawahara, and H. Ishiguro, "Erica: The erato intelligent conversational android," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 22–29, 2016.

[20] A. Baddeley, "Short-term memory," in *Memory*, pp. 41–69, Routledge, 2020.

[21] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload* (P. A. Hancock and N. Meshkati, eds.), vol. 52 of *Advances in Psychology*, pp. 139–183, North-Holland, 1988.

[22] M. Slater, A. Steed, J. McCarthy, and F. Maringelli, "The influence of body movement on subjective presence in virtual environments," *Human Factors*, vol. 40, no. 3, pp. 469–477, 1998. PMID: 9849105.

[23] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks, "Walking > walking-in-place > flying, in virtual environments," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, (USA), p. 359–364, ACM Press/Addison-Wesley Publishing Co., 1999.

[24] A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends in neurosciences*, vol. 20, no. 10, pp. 435–442, 1997.

[25] M. Vinck, R. Oostenveld, M. van Wingerden, F. Battaglia, and C. M. Pennartz, "An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias," *NeuroImage*, vol. 55, no. 4, pp. 1548–1565, 2011.

[26] N. Blomsma, B. de Rooy, F. Gerritse, R. van der Spek, P. Tewarie, A. Hillebrand, W. M. Otte, C. J. Stam, and E. van Dellen, "Minimum spanning tree analysis of brain networks: A systematic review of network size effects, sensitivity for neuropsychiatric pathology, and disorder specificity," *Network Neuroscience*, vol. 6, pp. 301–319, 06 2022.

[27] A. Mheich, F. Wendling, and M. Hassan, "Brain network similarity: methods and applications," *Network Neuroscience*, vol. 4, pp. 507–527, 07 2020.

[28] J. Gonzalez-Astudillo, T. Cattai, G. Bassignana, M.-C. Corsi, and F. D. V. Fallani, "Network-based brain–computer interfaces: principles and applications," *Journal of Neural Engineering*, vol. 18, p. 011001, jan 2021.

[29] A. Zalesky, A. Fornito, and E. T. Bullmore, "Network-based statistic: Identifying differences in brain networks," *NeuroImage*, vol. 53, no. 4, pp. 1197–1207, 2010.

[30] C. Phunruangsakao, D. Achanccaray, S. Bhattacharyya, S.-I. Izumi, and M. Hayashibe, "Effects of visual-electrotactile stimulation feedback on brain functional connectivity during motor imagery practice," *Scientific Reports*, vol. 13, no. 1, p. 17752, 2023.

[31] T. Strobach, A. Schütz, and T. Schubert, "On the importance of task 1 and error performance measures in prp dual-task studies," *Frontiers in psychology*, vol. 6, p. 123458, 2015.

[32] L.-A. Tuscan, J. D. Herbert, E. M. Forman, A. S. Juarascio, M. Izzetoglu, and M. Schultheis, "Exploring frontal asymmetry using functional near-infrared spectroscopy: a preliminary study of the effects of social anxiety during interaction and performance tasks," *Brain imaging and behavior*, vol. 7, pp. 140–153, 2013.

[33] M. Cazzell, L. Li, Z.-J. Lin, S. J. Patel, and H. Liu, "Comparison of neural correlates of risk decision making between genders: an exploratory fnirs study of the balloon analogue risk task (bart)," *Neuroimage*, vol. 62, no. 3, pp. 1896–1911, 2012.

[34] C. S. Soon, M. Brass, H.-J. Heinze, and J.-D. Haynes, "Unconscious determinants of free decisions in the human brain," *Nature neuroscience*, vol. 11, no. 5, pp. 543–545, 2008.

[35] Q. Chenot, E. Lepron, X. De Boissezon, and S. Scannella, "Functional connectivity within the fronto-parietal network predicts complex task performance: A fnirs study," *Frontiers in Neuroergonomics*, vol. 2, p. 718176, 2021.

[36] G. Fraga-González, D. J. Smit, M. J. Van der Molen, J. Tijms, C. J. Stam, E. J. d. Geus, and M. W. Van der Molen, "Graph analysis of eeg functional connectivity networks during a letter-speech sound binding task in adult dyslexics," *Frontiers in psychology*, vol. 12, p. 767839, 2021.