Using Physiological Information to Classify Task Difficulty in Human-Swarm Interaction

Joseph P. Distefano[1], Hemanth Manjunatha[1], Souma Chowdhury[1],
Karthik Dantu[2], David Doermann[2], and Ehsan T. Esfahani[1*]

Abstract—Human-swarm interaction has recently gained attention due to its plethora of new applications in disaster relief, surveillance, rescue, and exploration. However, if the task difficulty increases, the performance of the human operator decreases, thereby decreasing the overall efficacy of the human-swarm team. Thus, it is critical to identify the task difficulty and adaptively allocate the task to the human operator to maintain optimal performance. In this direction, we study the classification of task difficulty in a human-swarm interaction experiment performing a target search mission. The human may control platoons of unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) to search a partially observable environment during the target search mission. The mission complexity is increased by introducing adversarial teams that humans may only see when the environment is explored. While the human is completing the mission, their brain activity is recorded using an electroencephalogram (EEG), which is used to classify the task difficulty. We have used two different approaches for classification: A feature-based approach using coherence values as input and a deep learning-based approach using raw EEG as input. Both approaches can classify the task difficulty well above the chance. The results showed the importance of the occipital lobe (O1 and O2) coherence feature with the other brain regions. Moreover, we also study individual differences (expert vs. novice) in the classification results. The analysis revealed that the temporal lobe in experts (T4 and T3) is predominant for task difficulty classification compared with novices.

I. INTRODUCTION

Human-swarm interaction (HSI) has gained a lot of attention in recent years due to its many applications, including exploration, search, rescue, and surveillance [1]. The combination of mobile swarm robotics and human-in-the-loop systems generates an opportunity to leverage the best of both worlds [2]. However, increased task difficulty decreases the efficiency of the human operator, which in turn decreases the overall efficiency of the human-swarm team. Thus it is required to maintain an optimal level of task difficulty and intelligently adapt the task allocation between human operators and robot swarms. Hence identifying the operator's perceived task difficulty [3] occupies a central role in efficient HSI.

However, the task difficulty cannot be predicted in real-time as relevant task information is not known prior to the interaction. Frequently during human-swarm interaction, the state of the swarms or swarm mission does not provide adequate information to completely quantify the nature of the task. In this direction, physiological information such as brain activity and eye movement can augment the swarm states and provide secondary information about the task and state of the operator. Thus providing a means to infer task difficulty. Specifically, by understanding the various features in the brain activity that contribute to predicting task difficulty, we allow an opportunity to predict when the operator may need assistance or is not getting enough feedback. This analysis could also potentially be used to provide insight on game balance or the training of operators [4]. Moreover, due to individual differences in human brain cognition, the perceived task difficulty depends on individual differences [5]. These individual differences can be identified in a skilled marksman vs. a novice marksman. For instance, the human's visuomotor performance and efficiency have been shown to affect cortico-cortical communications in rifle shooting [6]. Thus, identifying different features that could be utilized for expert vs. novice operators could provide a better classification of task difficulty. In this regard, we explore the use of brain activity (measured through electroencephalography) to classify the task difficulty of a human operator interacting with a robot swarm.

To study task difficulty prediction using EEG in HSI, we have built a simulation platform that allows for human supervisory control of swarm robotics while completing missions with increasing difficulty levels. In this study, task difficulty is represented through the game difficulty levels. The simulation platform has a graphical user interface that allows the users to interact with the swarms. While the human interacts with the swarm, an electroencephalogram (EEG) is used to record the human's brain activity. We consider two approaches to classify the task difficulty level using EEG: 1) Features-based approach: Here, a set of well-known features (e.g., coherence values) are extracted and used to classify the task difficulty using a support vector machine (SVM). 2) Deep learning approach: Here, raw EEG epochs are used as input to a convolution neural network (CNN), where the output is the task difficulty level. When human participants are involved, the individual differences in skill level affect the perceived task difficulty. Hence, we consider individual differences by categorizing the participants as experts or novices and separately classifying task difficulty in each category. The results show that both feature-based SVM (71.60%) and CNN (83.80%) approaches can classify the task difficulty well above the chance, with CNN performing better than the SVM classifier.

In terms of experts and novices, the classification accuracy is higher among experts when compared to novices.

II. METHOD AND MATERIALS

A. Human Subject Study Framework

The framework for the human-swarm interaction simulation platform consisted of three separate modules (Fig. 1). The three modules are the gaming framework, the physiological data recording, and the lab streaming layer. These modules exchange information and offer different functionalities to create synchronous data recording during a human supervisory control swarm mission. These three modules run parallel using an open-source Ray framework [7].

The gaming framework includes two separate components that work in parallel throughout the game. The first component is the simulation environment which was developed using the open-source library pybullet [8]. The simulation environment is run through a real-time 3D physics engine. The second component is the graphical user interface that displays the information from the simulation environment. The interactive GUI presents all information needed from the simulation for the user to complete the mission.

Comprised in the simulation environment is the rendered map, which incorporates roads and buildings. In the simulation environment, the robotic swarms can execute three primitives with a high level of autonomy: formation control, path planning, and task allocation. The user is only required to input the UxV target location. In the environment, the human is supervising three UGV and three UAV platoons. More information on the design of the simulation environment can be found in [9]. This simulation environment was also used for learning swarm tactics over complex environments in [10]. The lab streaming layer module is the terminal location for all data being recorded. This data includes all the environment, EEG, game, and user data. Every instance of information is inputted into the lab streaming layer, labeled with a time stamp, and stored for post-processing.
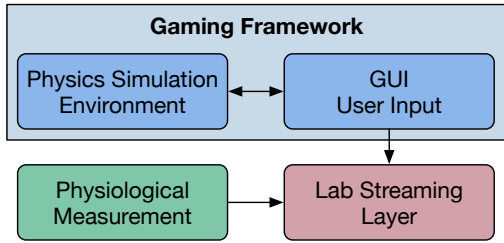

Fig. 1: The framework consists of three components: Gaming environment, physiological measurement, and lab streaming layer for time synchronization.

B. Graphical User Interface

The simulation environment was portrayed as a 2D interactive graphical user interface using the open-source pygame library. The design of the GUI was based on popular real-time strategy games such as starcraft. The GUI consists of a partially observable environment map, platoon panels to
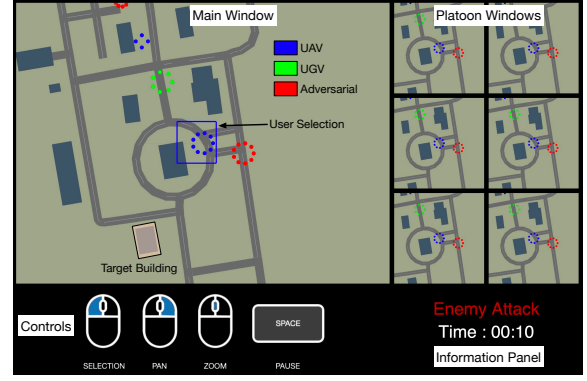

Fig. 2: The GUI presents all information to the user needed to complete the mission. It allows for the control of UxV platoons through the main window. Each platoon has a panel window. The user feedback panel provides information about enemy encounters and the remaining time in the game.

display an up-close overview of the platoons, a user feedback panel that notifies the user of game information, and a user controls panel (Fig. 2). The GUI allows for user control allowing complete freedom to allocate the platoons anywhere on the map. More information on GUI components can be found in [9].

C. Physiological Measurement

Participant's brain activity is recorded through a non-invasive B-Alert X24 EEG headset from Advanced Brain Monitoring© at 20 channels and 2 reference electrodes with a sampling rate of 256 Hz. The sensor locations are O1, O2, P4, POz, P3, Pz, Cz, C3, C4, Fz, F3, F4, T6, T4, F8, Fp1, Fp2, F7, T5, and T3 (Fig. 3).

D. Experiments

Students who had experience playing real-time strategy games were recruited from the University at Buffalo. The international review board approved the human subject study (IRB # STUDY00003659) before experimentation. All participants
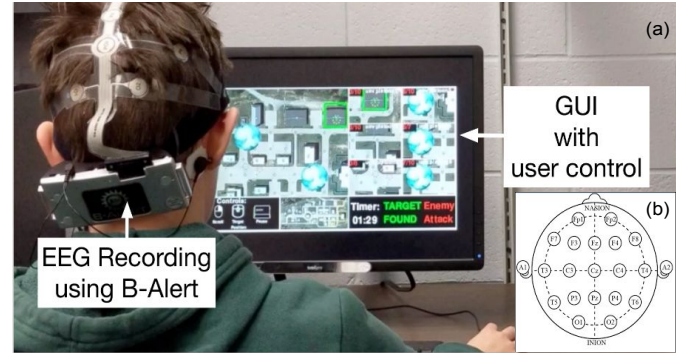

Fig. 3: (a) Experimental setup for monitoring the human physiological states while supervising a group of UAVs and UGVs in target search task. (b) International EEG 10-20 electrode placements.

had normal or correct vision. The participants were asked to play the game before to show proficiency in controls. After gaining adequate practice time, the participants underwent individual difference assessments.

After taking the individual difference and EEG baselines, the participants played three randomized games of different difficulties. The game's difficulties vary depending on the different complexities in the environment. The easiest task included a fully observable environment with no adversarial teams. The second level included a static adversarial team that blocked the paths and target buildings. Finally, the hardest level included a dynamic adversarial team whose position changed over time. The adversarial teams were initially hidden and made visible when the user's platoon was within the predetermined range in both static and dynamic levels. If the user's platoons and adversarial teams collided, a battle would result in the loss of platoons in a stochastic manner. If a platoon is defeated (zero members), it will disappear from the map, and the user may no longer use it.
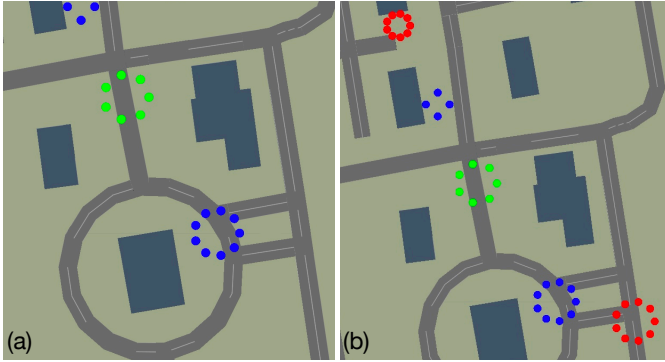

Fig. 4: This figure shows the game being played both with complexities and without complexities. In a) you can see the human has no adversarial units where as in b) adversarial units (red) are blocking the paths.

E. Individual difference

When working with human participants, individual differences are inevitable. Furthermore, individual differences affect the cognitive ability of humans and their perception of task difficulty. Thus it is desirable to capture this information and use it in analysis. We have used visual search (VS) and multi-object traction (MOT) to capture the individual differences [11]. These differences allow us to categorize the participants as experts and novices.

VS and MOT are EEG baseline procedures that assess individual skill levels and categorize the participant as an expert or a novice. In MOT, a participant should track a set of pre-marked circles. These circles are then moved around a space, and the user must select the marked circle at the end of the trial (Fig. 5(a)). VS is tested by flashing a screen with multiple shifted 'T's and asking the user to identify the direction of the correct T within a 2-second window (Fig. 5(b)). The user's scores depend on their ability to identify the circles and find the direction of the T. If the participant
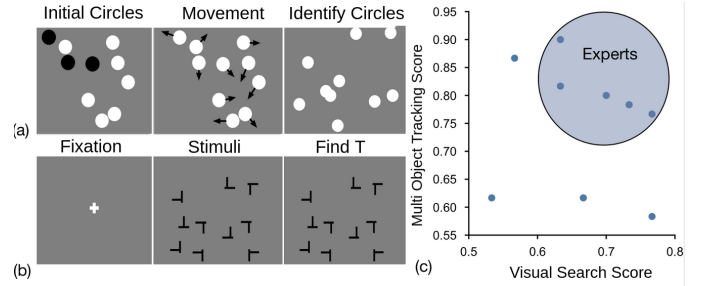

Fig. 5: (a) Multi-object tracking experiment. Target objects are highlighted briefly and the participants must track them during random movement and identify them when they stop. (b) Visual search experiment. Participants are required to identify the direction of the target shape ('T') among the detractors by pressing the arrow key corresponding to the direction. (c) Identifying experts and novices using MOT and VS scores.

scored well on both tests, they are considered an expert. Out of 9 participants, 5 were labeled as experts, and the remaining were labeled as novices.

III. DATA ANALYSIS

The data recorded during the human subject study underwent post-processing and further analysis. The details of the data analysis techniques are elaborated in the section below.

A. Classification

Classifying task difficulties is posed as a three-class classification problem. The baseline mission data with no adversarial team is the least difficult level. The difficulty level is increased by introducing static and dynamic adversarial teams. Before classification, the class labels were balanced. We explored two classification methods, support vector machine (SVM) classifier with EEG features and Convolutional neural network (CNN) approach with raw EEG.

We have considered three types of data splitting schemes. Due to the individual differences in EEG research, it is important to explore different training data sets to obtain accurate feature selection results and compare the results obtained from different training sets. 1) Subject independent: pool all the participant's data, balance, and randomly split 66% for training and 33% for testing. 2) Subject dependent: Out of 10 participants, training is done in 9 participants, and testing is done on the remaining participants (leave-one-out). The process is repeated until every participant is considered for testing. 3) Individual difference-based: The participants are split into experts and novices depending on the individual difference score (see Section II-E). Then the classification is carried separately for experts and novices. The discussed data splitting scheme is shown in Fig. 6. Note that the data is split after the feature selection step, and the scheme is the same for SVM and CNN.

B. Feature Extraction

The EEG signals were bandpass filtered (0.1-70 Hz) to remove the electrical noise of the headset and the slow drift
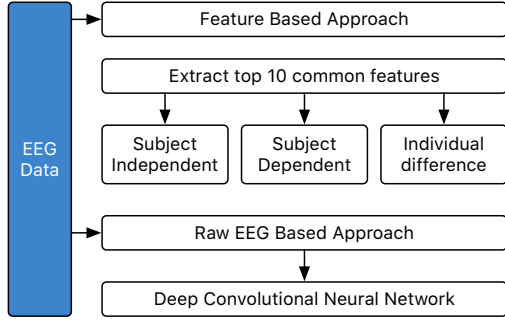
Fig. 6: Data splitting scheme for classification.

caused by the electrodes moving on the scalp. EEG signals were epoched into 2-second windows, and epochs with large artifacts were dropped. Increasing the window length would increase the noise, and decreasing the length would reduce the data points available for learning. Hence, we heuristically choose a window length of 2 seconds. The eye blink and drastic body movement artifacts were also removed using independent component analysis with the Picard algorithm in MNE-Python [12].

Once a cleaned epoched EEG signal was obtained, the features were extracted. The first set of features extracted were from the b-alert live software (Advanced Brain Monitoring©. Carlsberg, CA, USA). These features include mental workload, engagement, and distraction. Additional details on feature extraction and baselines used can be found in [13].

The second set of features were coherence values from the brain activity frequency bands. Coherence measures the similarity of two signals in a predefined frequency band for a given period. In this paper, we have used 6 frequency bands; delta [4-8 Hz], low alpha [8-10 Hz], high alpha [11-13 Hz], low beta [14-22 Hz], high beta [23-35 Hz] and gamma [36-44 Hz]. The coherence values (as given by equation 1) were calculated for each electrode pair and frequency band.

$$coh = \frac{|E[Sxy]|}{\sqrt{E[Sxx] * E[Syy]}} \quad (1)$$

where S_{xx}, S_{yy} are power spectral densities of the channels, and S_{xy} is the cross power spectral density of the channel being used.

C. Feature Selection

Since the number of features is larger, the classification algorithm can over-fit by learning noise. To avoid over-fitting, we perform feature selection. The present study recorded EEG from 20 channels using a 10-20 international system and considered six frequency bands for feature extraction. This results in 1200 coherence features. To reduce the number of features even further, we explored the existing literature on task difficulty. We found that the frontal, parietal, motor, and occipital regions, especially during visual-motor tasks, are predominantly correlated with task difficulty [14]. Thus we selected T4, T3, O1, P3, Pz, F3, Fz, F4, C4, P4, C3, Cz, and O2 electrodes to use.

A recursive feature elimination (RFE) was performed to obtain the best set of features. The RFE takes the initial set of features used as the input for the machine learning algorithm and ranks the features. The features deemed significant stay in the data set, and the features that have small effects on the classification are removed. This is repeated until a user-defined number of features are identified. RFE is stochastic in nature. As a result, different features can be selected in a different trial. To reduce the effect of randomness, we ran RFE 10 times and selected the most common features. For RFE, we have used linear SVM with RBF kernel.
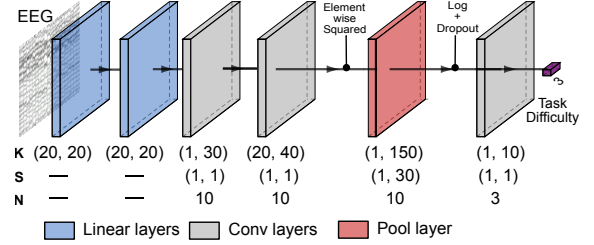
D. Convolutional neural network

Fig. 7: Architecture of the deep learning model. K: Kernel size, S: Stride, and N: Number of filters.

The architecture of the convolution neural network (CNN) used for classification, as shown in Fig. 7 is inspired by [15] and [16]. The first two linear layer operation is given by

$$x' = (x - W_{sh}\bar{x}) \odot (W_{sc}\tilde{x}_{sh}) \quad (2)$$

Where x is the input epoch, $\bar{x}$ is the mean value calculated across each electrode and $\tilde{x}_{sh}$ is the standard deviation of $(x - W_{sh}\bar{x})$. Note W_{sh} and W_{sc} are the learnt weights of the linear layers and the $\odot$ represents element-wise multiplication (Hadamard product). Note that if W_{sh} is identity and W_{sc} is the inverse covariance matrix (precision matrix), then the equation 2 reduces to a standard z-score. The linear layers are succeeded by two convolution operations; kernel size 1×30 and 20×40 (stride of 1). The number of filters is kept as 10 for both convolution layers. After two convolution layers, the output is element-wise squared, followed by average pooling. The kernel for average is 1×150, and stride is 1×30. After pooling, the log function is applied to the output, followed by a dropout. The dropout percentage is kept at 20% to avoid over-fitting. Finally, a convolution operation with kernel size 1×10 (stride=1) is applied to reduce the feature size to 3. Finally, a softmax function is used to convert the output to class probability.

IV. RESULTS AND DISCUSSION

In this section, we provide the classification results and discuss the importance of selected features. Note the discussion of selected features is in the context of the SVM classifier. Table I shows the classification results using two different methods and different data splitting schemes (see Section III-A). For subject-independent classification, deep learning performed

better than the feature-based SVM classifier. However, for subject-dependent classification, SVM performance was better than the deep learning approach. Nonetheless, both classifier's performance is poor for subject-dependent classification. The main reason for poor performance can be attributed to individual differences among participants. Interestingly, when the classification is performed separately between expert and novice, the classification rate is higher for experts than novices. This trend holds both for the SVM and CNN approaches. A probable explanation might be that experts make more efficient use of neural resources leading to different brain dynamics when compared with novices [5].

Fig. 8: The electrodes that have significant coherence are displayed on the EEG 10-20 system.

TABLE I: Classification results

Classification Method	Subject Independent	Subject Dependent	Expert	Novice
CNN	83.80±1.42	42.19±2.88	90.11±4.87	86.13±5.63
SVM	71.60±2.49	50.1±1.23	77.98±1.65	72.14±1.05

The significant features extracted from the feature selection are listed in Table II. Among the selected features, the Pz-O2 and O1-P4 coherence values show the inner connectivity in the parietal and occipital lobes of the brain. The F3-C3 and F3-O2 coherence values portray the connectivity between the prefrontal cortex with the brain's motor cortex and occipital lobe. These features signify which sections of the brain interact differently during different task difficulties.

TABLE II: Significant Features

Coherence Electrodes	Frequency Band
Pz-O2	8-10 Hz (Low Alpha)
F3-C3	14-22 and 23-35 Hz (Low and high beta)
F3-O2	14-22 and 23-35 Hz (Low and high beta)
O1-P4	14-22 and 35-44 Hz (Low beta and gamma)

TABLE III: Expert vs Novice Significant Features

Expert Features	Novice Features
T4-O2 (High beta 23-35 Hz)	O1-C4 (Low Beta 14-22 Hz)
T3-O1 (High beta 23-35 Hz)	Cz-O2 (Low Beta 14-22 Hz)
Fz-F4 (Gamma 36-44 Hz)	Pz-C4 (Low Beta 14-22 Hz)

The expert features show a strong correlation between the temporal lobe and the occipital lobe. These are unique features that did not show up in the previous feature selection. The novice features all contained electrodes in the motor cortex. Fig. 8 shows the visual representation of the significant features on the EEG 10-20 electrode locations. As you can see, all participant's features connect the prefrontal cortex and the occipital lobe. However, the expert's connectivity features are unique because they connect the occipital lobe to the temporal side lobes.

In terms of most selected features, the Pz-O2 coherence is representative of the occipital and parietal lobe connectivity. In previous research, the low alpha band has reports
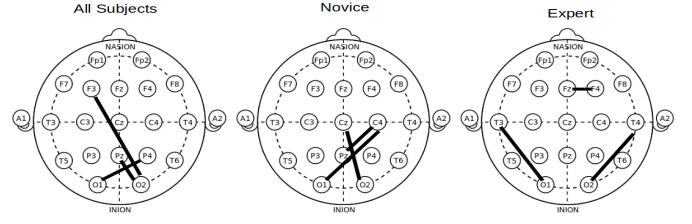
of synchronization and variance when performing different tasks. [17], [18]. It was also found that the Pz electrode spectral density drastically changes due to task load in human-computer interaction with an event-related potential [19]. The beta frequency band's F3-C3 and F3-O2 coherence values represent the prefrontal cortex connectivity with the motor cortex and the occipital lobe. The occipital lobe activation indicates the visual demand of humans, and the frontal lobe signifies engagement and data processing [20]. In this human swarm interaction mission, both the visual demand and the engagement would increase due to the added adversarial complexities. On similar lines, the O1-P4 electrode can also be accredited to the increase in visual demand.

The selected features between experts and novices are significantly different. The expert features are unique due to the involvement of the temporal lobe. Previous research shows that electrodes Fz, T4, and T3 power spectral densities strongly correlate with psychomotor efficiency. This has been shown both in rifle shooting by experts and novices and through a physical human-robot interaction task difficulty analysis [21, 6]. Therefore, these three electrodes are included in the expert features but are lacked in the novice features. Although the Fz, T4, and T3 coherence features are not selected, the connection with the occipital lobe could be from the visual intensity of the human swarm interaction task.

V. CONCLUSION

This paper presents an analysis of brain activity during a human swarm interaction study to quantify task-difficulty in a target search mission. The participants played three separate games to find the target location with increasing task difficulty due to adversarial complexities in a partially observable map. While the participants provided supervisory commands, their brain activity was measured, which is used to quantify the task difficulty level.

Task difficulty quantification was posed as a three class classification problem using EEG data as the input. The three-class problem corresponds to the three different levels of difficulty in the mission: no adversarial team, static adversarial team, and dynamic adversarial team. We explored two classification methods: support vector machine (SVM) classifier with EEG features and deep learning approach with raw EEG. The EEG features consisted of coherence values extracted from 2-second epochs from T4, T3, O1, P3, Pz, F3, Fz, F4, C4, P4, C3, Cz, and O2 sites and in the alpha, beta, and gamma frequency

bands. For the deep learning approach, the raw epoch was directly used as an input. Irrespective of the classification method, the EEG signals were filtered and cleaned with ICA. Both feature-based and deep learning approaches were able to classify task difficulty well above the chance (33%). Most of the prominent features in the classification showed a connection between the prefrontal cortex with the motor cortex and occipital lobes.

Due to the individual differences in the skill level, the participants were separated into experts and novices based on the multi-object tracking and visual search tasks scores. Among the experts, coherence values at the temporal lobe (T3 and T4) were most significant, indicating that the experts have more psychomotor efficiency than novices. However, most of the selected coherence features involved the occipital lobe regardless of expert or novice. In future work, we wish to explore the relationship between task difficulty classification and the cognitive workload of the participants. We also hope to explore more complex features including the human brain connectivity graphs.

REFERENCES

[1] Andreas Kolling et al. "Human swarm interaction: An experimental study of two types of interaction with foraging swarms". In: *Journal of Human-Robot Interaction* 2.2 (2013).

[2] Aya Hussein and Hussein Abbass. "Mixed initiative systems for human-swarm interaction: Opportunities and challenges". In: *2018 2nd Annual Systems Modelling Conference (SMC)*. IEEE. 2018, pp. 1–8.

[3] Jason R Tregellas, Deana B Davalos, and Donald C Rojas. "Effect of task difficulty on the functional anatomy of temporal processing". In: *Neuroimage* 32.1 (2006), pp. 307–315.

[4] Kate C Ewing, Stephen H Fairclough, and Kiel Gilleade. "Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a biocybernetic loop". In: *Frontiers in human neuroscience* 10 (2016), p. 223.

[5] Shanshan Dong et al. "Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty". In: *Brain research* 1616 (2015), pp. 146–156.

[6] Sean P Deeny et al. "Electroencephalographic coherence during visuomotor performance: a comparison of cortico-cortical communication in experts and novices". In: *Journal of motor behavior* 41.2 (2009), pp. 106–116.

[7] Philipp Moritz et al. "Ray: A distributed framework for emerging AI applications". In: *13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 18)*. 2018, pp. 561–577.

[8] Erwin Coumans and Yunfei Bai. *PyBullet, a Python module for physics simulation for games, robotics and machine learning*. http://pybullet.org. 2016-2019.

[9] Hemanth Manjunatha et al. "Using Physiological Measurements to Analyze the Tactical Decisions in Human Swarm Teams". In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2020, pp. 256–261.

[10] Amir Behjat et al. "Learning Robot Swarm Tactics over Complex Adversarial Environments". In: *2021 IEEE International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*. IEEE. 2021.

[11] Amirhossein H Memar and Ehsan T Esfahani. "Physiological Measures for Human Performance Analysis in Human-Robot Teamwork: Case of Tele-Exploration". In: *IEEE Access* 6 (2018), pp. 3694–3705.

[12] Pierre Ablin, Jean-François Cardoso, and Alexandre Gramfort. "Faster independent component analysis by preconditioning with Hessian approximations". In: *IEEE Transactions on Signal Processing* 66.15 (2018), pp. 4040–4049.

[13] Chris Berka et al. "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks". In: *Aviation, space, and environmental medicine* 78.5 (2007), B231–B244.

[14] Amirhossein H Memar and Ehsan T Esfahani. "Objective assessment of human workload in physical human-robot cooperation using brain monitoring". In: *ACM Transactions on Human-Robot Interaction (THRI)* 9.2 (2019), pp. 1–21.

[15] Robin Tibor Schirrmeister et al. "Deep learning with convolutional neural networks for EEG decoding and visualization". In: *Human brain mapping* 38.11 (2017), pp. 5391–5420.

[16] Nikolaos Passalis et al. "Deep Adaptive Input Normalization for Price Forecasting using Limit Order Book Data". In: *arXiv preprint arXiv:1902.07892* (2019).

[17] II Korobeinikova and NA Karatygin. "Spectral-coherent characteristics of EEG alpha rhythm in different efficiency of visual spatial task performance in humans". In: *Bulletin of experimental biology and medicine* 153.5 (2012), pp. 601–604.

[18] Alan Gevins and Michael E Smith. "Neurophysiological measures of cognitive workload during human-computer interaction". In: *Theoretical Issues in Ergonomics Science* 4.1-2 (2003), pp. 113–131.

[19] Michael E Smith et al. "Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction". In: *Human Factors* 43.3 (2001), pp. 366–380.

[20] Jeremy C Rietschel et al. "Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty". In: *Biological psychology* 90.2 (2012), pp. 127–133.

[21] Amirhossein H Memar and Ehsan T Esfahani. "EEG correlates of motor control difficulty in physical human-robot interaction: A frequency domain analysis". In: *2018 IEEE Haptics Symposium (HAPTICS)*. IEEE. 2018, pp. 229–234.