Physiologically Attentive User Interface for Robot Teleoperation

António Jorge Pires Gomes Tavares
antonio.tavares@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

October 2020

Abstract

The management of user attention is becoming a crucial challenge in the development of modern user interfaces, both in the Human-Computer Interaction and Human-Robot Interaction fields. User interfaces are shifting from being information-hungry devices to being attentive systems that consider their user's needs upon interaction. The interfaces developed for robot teleoperation can be particularly complex, often displaying very high amounts of information on their screens, which can induce a great deal of cognitive overload on the operators during life-critical missions. In this dissertation, a prototype for a Physiologically Attentive User Interface is presented, which is applied to an Urban Search and Rescue robot that provides a complex user interface. The system analyses physiological data, facial expressions, and eye movements to classify three emotional states (rest, stress, and workload) during robot teleoperation tasks. An attentive user interface is then assembled, which is modified dynamically according to the predicted emotional state in order to manage the user's focus during mentally demanding situations. This work contributes with the design of a user experiment, comprising emotion induction tasks that successfully trigger high and low cognitive overload states, along with effective strategies of managing user attention. Results from a user evaluation revealed no statistically significant differences in terms of the task performance and usability achieved when comparing this system to a classic user interface. The results were limited by the small number of subjects available for the study and the poor performance of the emotional state classifier.

Keywords: Robot Teleoperation, Attentive User Interface, Emotion Classification, Neural Networks

1. Introduction

Throughout the last few decades, society has witnessed a sharp rise in the level of technology available to humans. Not only the number of displays per human has increased substantially, but also the computational capabilities of our devices, such as PCs, laptops, smartphones, tablets, amongst many others, are now greater than ever. As the general population is becoming more and more surrounded by numerous interfaces that constantly distract us with pop-ups, such as notifications or messages, the attention span of users is getting reduced to the point where they have to filter information from multiple sources, often with the drawback of doing it at a superficial level, effectively limiting their ability to interact properly with each system [1]. This problem brings in the need for smarter interfaces that understand human needs and adapt to them, namely in the Human-Computer Interaction (HCI) field.

In order to tackle this problem, Vertegaal (2003) suggested a model for an Attentive User Interface (AUI), designed to be sensitive to the user's attention and act accordingly. These attentive user interfaces take advantage of overt properties of user attention, such as user presence, proximity and gaze direction, to determine which task or device the user is focused on and, consequently, his availability for interruptions [2]. However, while these measurements can help the understanding of the current visual focus of the user, they do not provide information about the state of mind, which can be just as or more relevant, since the user's physical activity is not necessarily an indicator of mental engagement. Fortunately, due to scientific advances in the field of Psychophysiology, it is possible to establish links between the human body and mind in order to acquire more reliable information about human cognitive and emotional states [3]. Therefore, the measurement of physiological signals allows the perception of covert states of mind that are not visible to an AUI. With this in mind, Chen (2006) proposed a framework for a Physiologically Attentive User Interface (PAUI) that resorts to physiological measures to respond actively to the user's needs. The use of a PAUI effectively extends the reach of an AUI, in the sense that it enables a deeper understanding of the emotional state of the user [4].

With the evolution of robotics and the increased extent of interactions between humans and robots, the problem of managing user attention also expanded to the field of Human-Robot Interaction (HRI), particularly in the robot teleoperation area. The ever-growing sophistication of the tasks performed by robots often comes with the drawback of requiring operators to deal with very complex interfaces that can potentially submit them to a great deal of stress and workload, effectively compromising their focus and performance on the given task. This issue is especially problematic in fields like Urban Search and Rescue (USAR) operations, where it is essential to guarantee that the operator's focus remains at its best, which would otherwise leave a person's life at stake.

In recent years, with the growth of computational power, the analysis of data and its employment in the development of artificially intelligent systems has led to increasingly effective algorithms for data classification that yield a vast array of applications, from self-driving cars, to medical diagnosis assistance, and even emotion recognition. The growing burst of available data is giving these algorithms more potential than ever, particularly in the field of Machine Learning (ML) and, more specifically, Deep Learning (DL). This promotes the development of intelligent systems that can accurately learn useful properties of data that are often not detectable by humans.

This article presents the development of a PAUI that aims to tackle the problem of managing user's attention and focus during robot teleoperation tasks. The PAUI retrieves physiological signals in real-time with the intent of classifying the emotional state of an operator during the teleoperation of a robot. The interface is then changed dynamically with respect to the user's predicted emotional state. The PAUI is developed over a pre-existing robot's Graphical User Interface (GUI) in order to reduce its complexity and increase the flexibility of the system. In this case, the PAUI was specifically applied to a USAR robot developed in 2005 called RAPOSA [5], and was elaborated upon the framework defined by Singh et al. (2018) [6] for a PAUI applied to this specific robot.

The objective of this work is to understand if the employment of a PAUI in this specific context can improve the usability of the interface when operators are under high cognitive stress and workload conditions, while improving their ability to stay focused when compared to a classic GUI approach. Taking this into account, the research aims to prove the following hypotheses:

- **H1:** *The employment of a PAUI in robot teleoperation tasks improves the efficiency of operators in comparison with a classic GUI approach.*
- **H2:** *The employment of a PAUI in robot teleoperation tasks improves the effectiveness of operators in comparison with a classic GUI approach.*
- **H3:** *The employment of a PAUI in robot teleoperation tasks improves the operator's ability to remain focused during missions in comparison with a classic GUI approach.*
- **H4:** *The employment of a PAUI in robot teleoperation tasks improves the level of usefulness and satisfaction experienced in comparison with a classic GUI approach.*

This work's contributions include: a prototype of the PAUI applied to the robot teleoperation field; the presentation of three emotion induction tasks that stimulate different emotional states on users; a USAR simulator that enables the re-creation of USAR environments; a report of the results obtained during the user study carried out for the evaluation of the system.

2. Related Work

The literature review presented in this article focuses on two different areas: the design of attentive user interfaces and the estimation of mental states with regard to covert and overt signals of user attention. The approaches adopted by researchers in these fields that are relevant to this work are further described.

2.1. Attentive User Interfaces

The concept of an AUI has been target of significant concern as the years go by. Vertegaal (2003) [2] proposed a framework for the development of an AUI, i.e. an interface that is sensitive to the user's necessities, which can be achieved through the measurement of covert characteristics of user attention, such as user presence, gaze direction, proximity and speech. These interfaces can then act on this information and decide when the user is available for interruptions, delivering them in a progressive way instead of forcing the information upon the user, potentially leading to a decrease in the user's focus.

Bulling (2016) [1] considered the management of user attention as a "critical challenge for next-generation human-computer interfaces". Human attention and focus is a limited resource that can play a very important role in the performance of the interfaces themselves [7]. Bulling addressed the problem of *continuous partial attention*, stating that the shifting of focus between various sources of information can effectively lead to a reduction in the overall focus of the user, since it limits the ability to concentrate on a specific task. Bulling defined Unobtrusiveness, Accuracy, Large scale, Longlivedness, Seamlessness and Context awareness as important categories that should be taken into account in the development of a new generation of *pervasive attentive user interfaces*.

While AUI's can accurately detect if a user is paying attention to a certain device, they cannot determine the actual level of engagement of the user with that device, since they rely on overt means of measuring a user's attention. For this reason, Chen and Vertegaal (2004) [8] proposed a prototype for a PAUI that is based on the use of LF (Low Frequency) spectral components and Electroencephalography (EEG) analysis. These signals allowed the classification of the user's mental and motor activity in order to differentiate four distinct user states that can be used to predict the user's availability for interruptions.

Significant research has also been done in improving user experience in the HRI field. Guo and Sharlin (2008) [9] developed a technique for capturing human arm movements and hand gestures that were used as input to control a robot in navigation and posture tasks, revealing an improvement in task performance when compared to the original keypad controls. Millan et al. (2004) [10] employed a Brain-Machine Interface (BMI) based on non-invasive electroencephalogram analysis in conjunction with advanced robotics to achieve brain-actuated control of a robot. Similarly, the Honda Research Institute Japan, Advanced Telecommunications Research Institute International (ATR), in cooperation with the Shimadzu Corporation [11], developed a BMI for the operation of a robot by human thoughts only, through the measurement of electric potential differences on the scalp through EEG and brain blood flow changes with near-infrared spectroscopy.

The necessity of an improvement of user interfaces and the simplification of the respective user interaction style has also been manifested in USAR operations. Baker et al. (2004) [12] studied more than a dozen USAR robot interfaces used in the American Association for Artificial Intelligence (AAAI) and the RoboCup Robot Rescue competition, concluding that these interfaces contain large amounts of information, most of which is disregarded by the operators for most of the time due to its irrelevance to the specific task at hands. Riley and Endsley (2004) [13] also expressed a concern for the lack of situational awareness in USAR operations. This study identified the workload induced due to a visually demanding task and poor integration of data on interfaces as some of the most problematic causes of degradation of task performance in search and rescue operations.

2.2. Mental State Estimation

The recognition of a user's emotional state can be a very useful input in the design of modern HCI and HRI systems, as it enables the development of affective computing strategies. Significant research has been done in using facial expressions to distinguish different emotions [14][15]. With the growth of more accurate means of measuring physiological signals, the employment of such measurements is proving to be a another good source of information on the emotional state of a person [16][17].

A study carried by Kim et al. (2004) [18] performed emotion prediction on 50 subjects in order to classify four emotional states (sadness, anger, stress and surprise) while acquiring data from Electrocardiography (ECG), Electrodermal Activity (EDA) and skin temperature variation. This study showed the clusters formed by each class had large variance within themselves, and significantly overlapped each other. A Support Vector Machine (SVM) was chosen as the classifier, obtaining a prediction accuracy of 61.76% for 4 classes and 78.43% when classifying only three classes (sadness, anger and stress). Wang et al. (2014) [19] investigated the potential use of EEG features for emotion classification by conducting a series of experiments that induced positive or negative emotions in six subjects who watched movie clips that targeted specific emotions. The extracted features were smoothed and dimensionality reduction methods were applied, which led to a classification accuracy of 91.77%. Zheng et al. (2014) [20] studied the usage of advanced DL models to classify emotional data from two classes (positive and negative) relying on extracted 62-channel EEG signals from subjects exposed to emotion-inducing video clips. This study concluded that Deep Belief Networks outperformed other common approaches such as SVM, KNN and GELM, obtaining an accuracy of 87.62%.

The works presented above show that a new generation of user interfaces that take in consideration their user's needs are emerging. The problem of managing user attention is manifesting itself in a wide range of areas, and has affected particularly the performance of USAR teams that use robots in their missions, whose interfaces can prove to be very complex and create a high visual demand on its users. With the rise of more sophisticated artificially intelligent systems, emotion recognition is becoming possible to perform with relatively high precision due to the extraction of physiological signals and facial expressions. It is believed that this information can be used to improve the operator's experience by updating the robot's interface with respect to his/her emotional state, leading to a decrease of the workload induced and an increase of the usability of the system.

3. Approach

In this work, a Physiologically Attentive User Interface (PAUI) was designed for application in the robot teleoperation field. The objective of the PAUI is to employ an artificial intelligent system trained to classify the operator's emotional state between three different emotional states (*Rest*, *Stress* and *Workload*). The emotional state prediction is then used to dynamically change the interface in real-time in order to improve the user's focus during high cognitive stress moments. The PAUI is established over the preexisting robot Graphical User Interface (GUI) with the purpose of reducing its complexity and displaying information in a clearer way, thus decreasing the emotional overload of the operator. Furthermore, the PAUI interacts with the old GUI by means of an automation tool (*SikuliX*) and extracts its image in order to render an AUI based on the original interface, thus eliminating the need to develop a new interface or a completely new solution from scratch when the source code is not available. Here, the PAUI is developed over the preexisting RAPOSA's user interface, which is a 15 years old interface with a complex display of information that can potentially overload the operator. Nonetheless, the solution presented offers flexibility to be adjusted to other interfaces.

3.1. System Architecture

The development of the PAUI comprises the combination of three different modules that worked together to make the concept of a PAUI applied in the robot teleoperation field feasible: the *Signal Extractor*, the *Emotional State Classifier* and the *Attentive User Interface*. Figure 1 presents the overall architecture of the system.

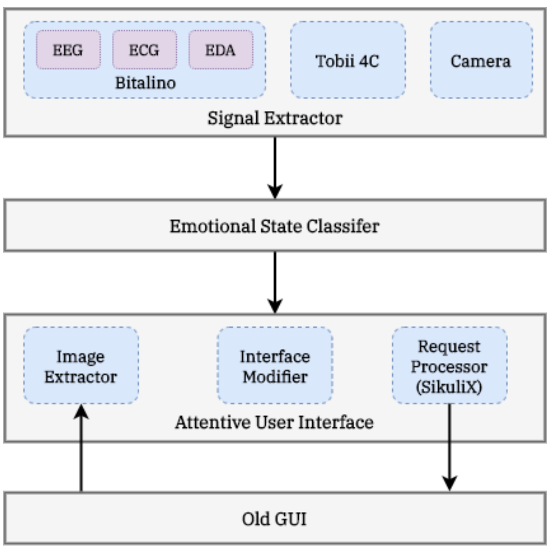

Figure 1: PAUI Architecture.

3.1.1 Signal Extractor

The signal extractor is responsible for the acquisition of data that can potentially yield valuable information about the mental state of a person. The extracted data is of three types: physiological signals; facial expressions and emotions; eye movements. For reading physiological signals, the Bitalino (r)evolution Plugged Kit BT by Plux was used to extract biosignals picked up by three sensors, namely EEG, EDA and ECG. For detecting eye movements, the device used was the Tobii Eye Tracker 4C by Tobii Technology. Finally, for facial expressions and emotions, the laptop integrated webcam was used alongside the AFFDEX SDK by Affectiva [21].

The extracted data is then processed in order to obtain a wide range of parameters that can later be used as input for training a classifier. Data processing was performed with the aid of the pre-existing PAUI framework developed by Singh et al. (2018) [6], which processes the extracted raw analog signals into more refined parameters. During execution, each device has a *thread* responsible for managing its data: the *Bitalino* thread which extracts physiological analog signals at 1000 Hz; the *Tobii* thread which monitors eye movements and fixation information at 90 Hz; and the *Camera* thread which extracts facial expressions and emotions at 30 Hz.

3.1.2 Emotional State Classifier

This module takes charge of classifying the user's emotional state based on the signals received from the signal extractor. As referred in the previous section, the system runs a thread for each data extraction device that is in charge of acquiring and processing its respective signals. Likewise, the emotional state classification also has a thread responsible for its management, which runs above the three signal extraction threads. This thread receives the processed signals from the Bitalino, Tobii and Camera threads and makes an average of the signals received during 1 second. The thread then uses the averaged signals as input for the classifier, which predicts the operator's emotional state.

In this work, the classifier used was an *artificial neural network*. Since the model's objective is to predict the user's emotional state from three different classes, where each instance must be assigned to only one class, the problem can be defined as a *single-label multiclass classification* problem. As the dataset collected in this work is class-balanced, the metric chosen to measure the success of the model was accuracy, and a *K-fold cross validation* with $K = 4$ [22] was adopted as the validation protocol, since the dataset acquired is small. Considering the data collected for the purposes of this work is time-series data, where the user's signals were collected over the course of time for each class, the data was split before shuffling [23] in a 75/25 ratio, using the first 75% of each class for the training set and the remaining 25% for the test set. A similar procedure was adopted for the validation test split. Additionally, the data was normalized with the mean and standard deviation computed for the training set, given that the features present in the dataset used for this work take up values in significantly different ranges.

Regarding the configuration of the model, the loss function used for the optimization of the model was *categorical cross-entropy*, using the *softmax* function as the activation function of the output layer of the model, as is common in multiclass classification problems [23]. The activation function chosen for the hidden layers was the *rectified linear unit* (ReLU), which is a popular choice of activation function that usually shows better convergence performance than the sigmoid and hyperbolic tangent activation functions, while also providing a solution to the *vanishing gradient problem* [24]. In relation to the stochastic optimization of the model, a state-of-the-art optimizer called *Adam* was used, which is an efficient optimization method that has low memory requirements and has been shown to be advantageous in practice for most problems, when compared to other stochastic optimization methods [25].

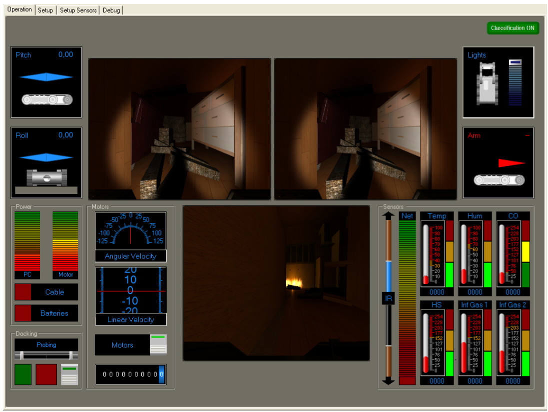

Figure 2: Original RAPOSA's GUI.

3.1.3 Attentive User Interface

The AUI module is responsible for managing the user's attention and ease the process of operating the robot in high cognitive stress situations. The AUI carries out this task by redefining the pre-existing interface, reducing its complexity and displaying only the relevant information at any given time. For this purpose, the AUI extracts image data (screenshots) from the old interface and uses it to render the new interface, which can be modified depending on the predicted emotional state. The AUI also issues requests to the old interface with the aid of a task automation tool called *SikuliX*, which enables the automation of mouse and keyboard operations, effectively allowing the AUI to interact actively with the old interface. With the aim of hiding the old interface from the user's view while maintaining the interaction between both interfaces, the old interface runs on a virtual machine that is connected to the host machine, where the AUI operates. The connection between both machines is established through a *host-only network*, allowing the transfer of screenshots from the old GUI to the AUI through the TCP communication protocol.

Depending on the predicted emotional state, the AUI can take different actions. If the predicted state is rest, the original GUI is rendered, since the operator can still manage his/her attention well enough to handle the complexity of the original interface and stay focused during the robot teleoperation task. Figure 2 shows the appearance of the original RAPOSA's graphical interface.

When the classifier predicts the emotional state of the operator as stress, the rendered interface remains to be the original GUI, but the AUI issues a request to the old GUI for an increase of the maximum speed of the robot. This request is executed by a SikuliX script that automates a series of clicks programmed to change the maximum speed value setting in the old GUI. This attentive measure can be helpful in situations where some areas of the environment are clogged with difficult obstacles, which can cause the operator to lose precious time and enter a stressful state. For this reason, an increase in the maximum speed of the robot can help the operator compensate for these moments when the environment is easier to traverse, where he/she can make use of the extra speed without having to manually change it in the Setup tab of the interface. While its value could theoretically be set to the limit at all times, doing so leads to significant overheating of the robot at the hardware level. For this reason, the AUI takes up the responsibility of managing the robot's maximum speed, allowing it to reach values closer to the limit when the user is under stress, while alleviating the robot when the user does not need its full potential.

In case the predicted state is workload, the interface is redefined to a simpler format that emphasizes the camera view of the robot and only shows the most relevant elements of the old interface. Furthermore, some elements are only displayed when the values they present go beyond a certain level. In this case, the battery levels are only shown when their values drop below 40%, and the sensor readings are shown when their values cross the danger zone. These measures contribute to maintaining the user's focus on the task, since the interface only demands more attention when it is absolutely essential. Figure 3 shows the new interface displayed in workload situations.

The AUI is the only module of the PAUI that needs to be tuned for each specific interface, since the procedures adopted for the management of user's attention are specific to the interface in question. As such, the procedures presented previously are specific to the RAPOSA's interface.

4. Evaluation

In order to evaluate the proposed solution, two experimental sessions were carried out: a data collection session and an evaluation session. The necessity of acquiring a new dataset arises from the fact that physiological data varies significantly from per-

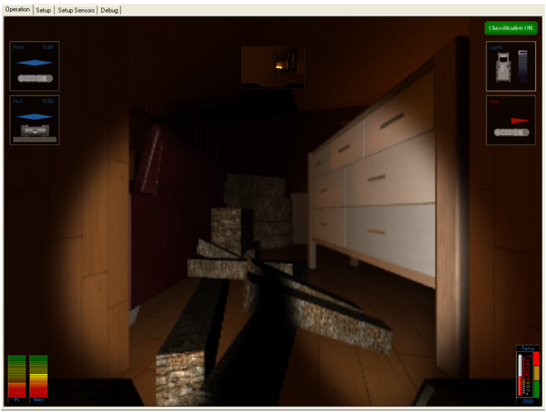

Figure 3: Redefined AUI rendered in workload situations.

son to person (see section 4.7.1). This means that the system could only be evaluated on the same subjects from whom data was collected, which made the use of the dataset collected previously by Singh et al. (2018) [6] unfeasible. Moreover, the high variability inherent to physiological data would require an extensive work to overcome the difficulty of training a classifier using data from all subjects. Even though the usage of person-specific classifiers requires the training of a model for each operator, it was the approach adopted since the emotional state classification is not the main focus of this work.

On a separate note, RAPOSA was not available for usage in this work due to the need of heavy maintenance. This led to the development of a USAR environment simulator using Unity, making the evaluation of the proposed solution possible. The simulator allowed the creation of a *mock-up* of the RAPOSA's original interface and controls, simulating the robot's operation as close as possible to reality. Apart from maintaining the main functionalities of the original system, the simulator enabled the development of custom Search and Rescue scenarios that were used to induce stress and workload to users while performing a robot teleoperation task.

4.1. Subject Grouping

For the purposes of the second experimental session (evaluation session), the subjects were split in two independent groups, where one group tested the PAUI approach and the other group tested the GUI approach. Although this method leads to higher variance in the results obtained, it would be unfeasible to have each subject test both approaches, due to the influence of carryover effects. In order to mitigate the variance in the results as much as possible, the Immersive Tendencies Questionnaire (ITQ) [26] was used, which reliably reflects each subject's tendency to become more involved in virtual environment tasks. The scores obtained in the questionnaire were then used to split the subjects in two balanced groups, where each group tested a different approach in the second experimental session.

4.2. Apparatus Used and Setup

The devices used to extract data from users in the experimental sessions were referred in section 3.1.1. Regarding the attachment of electrodes, the best suggested placement by Němcová et al. (2016) [27] was used for ECG: positive lead under right clavicula, negative lead under left musculus pectoralis major and reference lead under left clavicula. For EEG, the positive and negative leads were placed at forehead and the reference lead at the left earlobe. For EDA, only two electrodes are required, which were placed in the left hand palm [6]. Apart from the data acquisition devices, a gamepad was used to control the robot.

4.3. Performed Tasks

In order to evaluate the proposed solution, a task meant to induce each of the three targeted emotional states was designed using the USAR simulator. All three tasks required the subject to drive the robot in a home-like setup while attempting to complete a certain objective, where the home is adapted to fulfill the needs of each task.

During the *rest task*, each participant was required to drive along a room inside an empty house environment for five minutes. This task did not require great mental or physical effort, thus leaving the participant in a restful state, possibly with windows of boredom.

In the *stress task*, participants were asked to find four victims inside a house on fire before a timer ran out. With the aim of triggering a stressful state, a loud beep was played as each second passed, there were numerous obstacles spread around the house that made the teleoperation of the robot much more difficult, and the user suffered a time penalty if the robot's batteries discharged completely or if the robot's temperature got too high.

In the *workload task*, participants were required to find 10 objects (represented by red cubes) inside a house while answering workload inducing questions. These questions included basic arithmetic operations, requests to read values of sensors in the interface, questions about the surroundings of the robot and basic logic problems. It should also be noted that participants were reminded that there was not a time limit to answer each question, in order to avoid inducing unnecessary stress.

4.4. Metrics

In order to evaluate the hypotheses formulated, two types of metrics were considered: task performance metrics and user experience metrics. Task performance refers to the quality of the tasks achieved by users, where speed and accuracy are typically the most important metrics. Additionally, it was also of interest to measure the level of attention that users retained when performing the tasks. Therefore, the task performance metrics defined were:

- Completion time of the stress task;
- Number of objects found during the workload task;
- The relative change of the mean engagement values from the rest task to the workload task, in percentage. According to McMahan et al. (2015) [28], the engagement index that is extracted from the EEG sensor reflects a person's ability to sustain attention and gather information.

Regarding user experience metrics, the USE Questionnaire [29] was used to measure the level of usefulness and satisfaction reported by the subjects towards each approach.

It should be noted that, even though it does not contribute as a metric, the Discrete Emotions Questionnaire (DEQ) [30] was also filled by the participants for purposes of discussion, which was aimed at understanding what type of emotions were felt by the user over the course of each task.

4.5. Procedure

Both experimental sessions had a very similar procedure, which took aproximatelly 1 hour to complete. Upon arrival to the testing office, the participant was given a description of the experimental procedure. Additionally, an overview of the robot's GUI, its controls and its capabilities was presented. Subsequently, the participant proceeded to fill a demographics questionnaire, followed by the ITQ (this step was exclusive to the first experimental session). Afterwards, the eye tracking device was calibrated and the ECG, EEG and EDA electrodes were attached to the subject according to the configurations referred in section 4.2.

After going through all the necessary preparations, the subject went through a training session in order to get acquainted with the teleoperation of the robot, which usually took 10 to 15 minutes to complete. After leaving the training session, the rest, stress and workload tasks were executed in a random order. Upon the completion of each task, the user went through a 2 minute break period, in order to relax and come back to a normal mental state.

Regarding the second experimental session, after finishing the teleoperation tasks, the subject was also asked to fill the USE Questionnaire and the DEQ, followed by questions relative to the attentive elements of the interface, in case the subject belonged to the PAUI group. It should also be noted that, in case the subject belonged to the group that tested the classical GUI approach, the emotional state classifier still predicted the emotional state of the subject in real-time for the purposes of evaluating the classifier itself, but the system did not act on this information.

4.6. Participants

The subject group used in this study included 6 voluntary participants (4 male, 2 female) aged between 20 and 24 years old ($M = 22.833$, $SD = 1.472$). All participants were Portuguese university students, with a background in engineering.

4.7. Results

This section presents the measures adopted to visualize the data acquired, as well as the comparison of the classification accuracies obtained from different neural network model iterations. Furthermore, the ITQ results that led to the group division are also presented, as well as the statistical test results obtained from the metrics recorded in the evaluation session.

4.7.1 Model Training

After collecting the parameters extracted in the first experimental session, a visualization of the data was carried out in order to understand better the patterns present in the data. In conformity with the findings of other researchers [17] [19], it was found that there is an inherently high variability in the extracted physiological data, particularly from subject to subject, which arises from the fact that different subjects might be prone to different emotional reactions due to their personalities and experiences. The variance of physiological data between subjects was made evident through the application of a t-distributed Stochastic Neighbour Embedding (t-SNE) to a dataset formed by the acquired data of all six subjects, which formed clusters that tend to group the data from each subject separately (see figure 4).

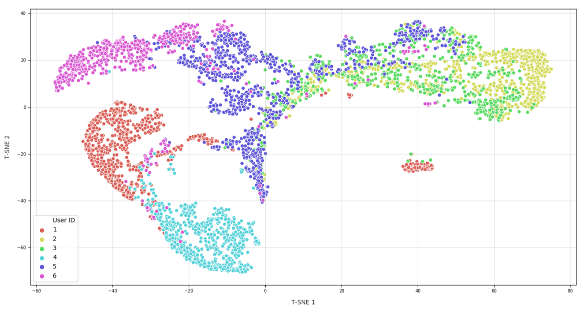

Figure 4: Embedded space that results from the application of t-SNE to the data collected from each user during the stress task, where each user is represented by a different color.

It should also be noted that the eye tracking features were not considered in the emotional state classifier training, due to its high percentage of missing values and low correlation with the classes.

Given the high dimension of the feature space (45-dimensional array), a Principal Component

Analysis (PCA) was applied to the dataset in order to reduce its dimensionality. Due to the inherent tendency of physiological signals to contain large amounts of noise [20], it was decided that it would be best to keep most of the explained variance present, in order to retain most of the valuable information. With this purpose, 95% of the total variance in the data was kept, which is explained by the first 31 principal components.

After projecting the dataset on the new feature space, a neural network model was trained for each user. The fine-tuning of the hyper-parameters led to the optimal values of 0.0001 for the optimizer's learning rate and 32 for the batch size. The optimal architecture achieved was a network with three hidden layers (40, 30 and 20 hidden neurons), where three different regularization methods were tested: L^1 regularization, L^2 regularization, and dropout. With the application of early stopping, L^2 regularization proved to be the method that yielded the highest accuracy, with a model trained over 500 epochs. The training of a model for each user and respective evaluation on the test set yielded an average accuracy of 80.9%. When the models were tested in real-time during the second session, the classification performance dropped significantly for each participant, to an average of 50.3% (a discussion on the decrease of the classification performance is presented in section 5). The accuracies obtained for each participant in the test set and in real-time are compared in table 1.

User ID	Test Set (%)	Real-Time (%)
1	87.9	51.1
2	81.9	49.8
3	77.3	48.3
4	81.6	52.3
5	83.1	51.7
6	73.6	48.5

Table 1: Comparison of the classification accuracies obtained in the test set and in real-time classification, for each user.

4.7.2 Group Division

Regarding the division in groups for the second experimental session, a score was given to each subject based on the ITQ results, following the author's scoring recommendations. A set containing the scores of all subjects ($M = 74.167$, $SD = 10.496$) was then partitioned into two subsets, where the sum of each subset's scores was as close as possible, in order to generate two balanced groups that contain subjects with both high and low ITQ scores. The group with the highest classification accuracy (on average) was chosen to test the PAUI approach, since the focus of this work is to gain insight on the advantages of using a PAUI against the usage of a traditional GUI approach.

4.7.3 Statistical Tests

With the aim of analysing the gathered data on the task performance obtained under both approaches, the Shapiro-Wilk test was employed to check data for normality. In case the data was normally distributed, an *independent t-test* was employed to determine if statistically significant differences were observed between both groups. For non normally distributed data, a *Mann-Whitney U test* was carried out.

The execution of a Shapiro-Wilk test for the completion time and relative changes of the engagement levels revealed that both are normally distributed. As the number of objects found is a discrete value, it is assumed that the distribution that generates the data is not normal. An independent t-test was applied to the completion time, showing no statistically significant differences between the PAUI approach (M = 743.3, SD = 55.7) and the GUI approach (M = 773.3, SD = 5.8) (t(4) = -0.541, p = 0.617). Similarly, the independent t-test applied to the relative changes of the engagement levels did not reveal any statistically significant differences in the values obtained in the PAUI approach (M = 175.8, SD = 149.2) and the GUI approach (M = 41.6, SD = 40.4) (t(4) = 1.504, p = 0.207). Furthermore, the application of a Mann-Whitney U test showed that there were no statistically significant differences in number of objects found by participants using the PAUI approach (M = 6.0, SD = 1.0) or the GUI approach (M = 5.67, SD = 2.08) (U = 3.5, p = 0.658).

Regarding the USE Questionnaire, it was assumed that the results it yielded did not follow a normal distribution, since they were drawn from 7-point Likert scales. The execution of a Mann-Whitney U test showed a tendency towards statistical significance on the sensation of effectiveness perceived by the participants (U = 0.5, p = 0.072), as well as the ease of use of the interface (U = 0.5, p = 0.077), where both elements have shown an improvement in the PAUI approach, when compared to the GUI approach. All the other questions led to statistically insignificant differences.

Additionally, the results from the DEQ revealed that users reported strong feelings associated with relaxation during the rest task, while obtaining higher scores for emotions related to anxiety during the stress and workload tasks. Furthermore, all three participants that tested the PAUI approach reported that, even though the system could show more robustness (in terms of classification accuracy), the changes performed by the interface were helpful and eased the performance of the tasks.

5. Discussion

The results presented show a significant drop in terms of performance of the emotional state classifiers in real-time in comparison to the accuracies obtained in the test set, from 80.9% to 50.3%. This decline of the classification accuracy can be explained due to the inherently high variability present in physiological signals not only between subjects, as stated previously in section 4.7.1, but also within each subject. Additionally, the low signal-to-noise ratios exhibited in physiological data may have contributed to the loss of the model's generalization power, along with the difficulty of attributing the correct class label to each data point.

Regarding the task performance metrics, the results obtained from the statistical tests were not conclusive. Still, the relative change of the engagement levels was higher for the PAUI users, which could be indicative that the attentive properties of the system improved the users' ability to remain focused. Nonetheless, the small number of subjects that were available for the experimental sessions makes it difficult to draw conclusions regarding hypotheses **H1**, **H2** and **H3**.

When it comes to the USE Questionnaire, a tendency towards statistical significance was verified for the improvement of the feeling of effectiveness and ease of use felt by the PAUI users, in comparison with the GUI users. These results indicate the effectiveness of the attentive strategies adopted by the PAUI, since they focus on enhancing the effectiveness of the users by presenting information on the interface in a clear and easy way to understand. However, the lack of statistical significance cannot assure the check of the hypothesis **H4**.

Although not statistically significant, the small improvements obtained in the task performance metrics are indicative that more significant differences could be achieved with the aid of a more powerful classifier, preferably trained with a larger and more refined dataset collected over the course of a longer teleoperation session. Moreover, the lack of statistical significance was expected due to the very small number of subjects available for the experiment (only 3 subjects for each condition). This presented a major limitation in the results obtained, which can be improved in future studies that rely on a larger number of subjects. Furthermore, the design of the experiment proved to be effective in replicating the pretended emotional states on all the participants. The sensations reported by the participants clearly show the effectiveness of the emotion induction tasks, which can be used in the future for studies that use both the USAR simulator or the real robot. Additionally, all three participants that tested the PAUI approach were able to experience moments where the classifier was predicting the correct emotional state. Despite the inconsistency caused by the poor classification performance, the PAUI users observed that the changes performed by the PAUI were helpful and allowed a better understanding of the environment, indicating that the attentive measures adopted by the system can lead to the improvement of the user's focus in difficult situations.

Summing up, the small number of subjects available and the poor classification performance were the main limitations of results obtained, leading to their statistical insignificance. The success of both the emotion induction strategies adopted and the attentive measures of the PAUI is indicative of potential improvements in future works that comprehend a significant number of subjects, along with enhanced classifiers. However, due to the lack of significance found in the statistical tests carried out for this study, the initial research statement cannot be validated.

6. Conclusions

Through the execution of a user study with 6 people, it was possible to analyze quantitative and qualitative measures of the user's task performance and preference towards the usage of a physiologically attentive system in comparison to the classic interface approach. The study carried out did not find any significant differences in terms of task performance between both groups, although there was a tendency for an improvement in the feeling of effectiveness and ease of usage, as reported by users in the questionnaires.

The major contribution of this thesis was the development of the first prototype of the PAUI applied to the pre-existing RAPOSA's interface, where a great emphasis was placed on creating effective methods of managing user's attention in moments of cognitive workload, taking into account not only the concerns expressed by previous research in the design of attentive user interfaces, but also the problems that are recurrent in interfaces developed for USAR robots. The planning of tasks designed to induce specific emotional states (rest, stress and workload) was also an important step in realizing what types of strategies can be implemented in re-creating real life USAR environments and successfully inducing the sensations felt by operators when going through these situations. The development of a USAR simulator also contributed largely to the success of the previous point, by allowing the modelling of custom 3D environments and the implementation of a large set of features that simulate the teleoperation process of the real robot, thus producing a faithful representation of the reality that can be used in the future studies.

Furthermore, the results obtained considering the small number of subjects available, allied with the limited capabilities of the emotional state classifier are indicative of potential improvements in future work. The achievement of a robust emotional state classifier is by itself a challenging task, and the results clearly show the need of obtaining a larger dataset that can accurately represent the underlying distribution that generates physiological data. Additionally, new strategies for managing user attention can be implemented by giving emphasis to the eye tracking device, which can give insight into the regions of the interface where users tend to stare at the most. Moreover, it could be beneficial to ease the process of adapting the system to other case studies in order to increase the flexibility of the solution. Finally, the execution of future evaluation studies with the real robot could be helpful in providing users an improved understanding of the robot teleoperation process, allowing the employment of attentive measures that offer a greater contribution to the task performance achieved.

References

[1] A. Bulling, "Pervasive Attentive User Interfaces," *IEEE Computer*, vol. 49, no. 1, pp. 94–98, 2016.

[2] R. Vertegaal *et al.*, "Attentive user interfaces," *Communications of the ACM*, vol. 46, no. 3, pp. 30–33, 2003.

[3] A. C. Dirican and M. Göktürk, "Psychophysiological measures of human cognitive states applied in human computer interaction," *Procedia Computer Science*, vol. 3, pp. 1361–1367, 2011.

[4] D. Chen, *The Physiologically Attentive User Interface Towards A Physiological Model of Interruptability.* Citeseer, 2006.

[5] C. Marques, J. Cristóvao, P. Alvito, P. Lima, J. Frazao, I. Ribeiro, and R. Ventura, "A search and rescue robot with tele-operated tether docking system," *Industrial Robot: An International Journal*, vol. 34, no. 4, pp. 332–338, 2007.

[6] G. Singh, S. Bermúdez i Badia, R. Ventura, and J. L. Silva, "Physiologically attentive user interface for robot teleoperation: real time emotional state estimation and interface modification using physiology, facial expressions and eye movements," in *11th International Joint Conference on Biomedical Engineering Systems and Technologies.* SCITEPRESS-Science and Technology Publications, 2018, pp. 294–302.

[7] R. Vertegaal, J. S. Shell, D. Chen, and A. Mamuji, "Designing for augmented attention: Towards a framework for attentive user interfaces," *Computers in Human Behavior*, vol. 22, no. 4, pp. 771–789, 2006.

[8] D. Chen and R. Vertegaal, "Using mental load for managing interruptions in physiologically attentive user interfaces," in *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, 2004, pp. 1513–1516.

[9] C. Guo and E. Sharlin, "Exploring the use of tangible user interfaces for human-robot interaction: a comparative study," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 121–130.

[10] J. R. Millan, F. Renkens, J. Mourino, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1026–1033, 2004.

[11] B. Zhang, J. Wang, and T. Fuhlbrigge, "A review of the commercial brain-computer interface technology from perspective of industrial robotics," in *2010 IEEE International Conference on Automation and Logistics.* IEEE, 2010, pp. 379–384.

[12] M. Baker, R. Casey, B. Keyes, and H. A. Yanco, "Improved interfaces for human-robot interaction in urban search and rescue," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 3. IEEE, 2004, pp. 2960–2965.

[13] J. M. Riley and M. R. Endsley, "The hunt for situation awareness: Human-robot interaction in search and rescue," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 3. SAGE Publications Sage CA: Los Angeles, CA, 2004, pp. 693–697.

[14] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 1, pp. 96–105, 2006.

[15] I. Hupont, S. Baldassarri, and E. Cerezo, "Facial emotional classification: from a discrete perspective to a continuous emotional space," *Pattern Analysis and Applications*, vol. 16, no. 1, pp. 41–54, 2013.

[16] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 102–115, 2011.

[17] P. C. Petrantonakis and L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 5, pp. 737–746, 2011.

[18] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and Biological Engineering and Computing*, vol. 42, no. 3, pp. 419–427, 2004.

[19] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.

[20] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME).* IEEE, 2014, pp. 1–6.

[21] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 3723–3726.

[22] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, and S. Ridella, "The 'K' in K-fold Cross Validation," in *ESANN*, 2012.

[23] F. Chollet, *Deep Learning with Python.* Manning Publications, 2017.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[26] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[27] A. Němcová, L. Maršánová, and R. Smíšek, "Recommendations for ECG acquisition using BITalino," in *Conference Paper EEICT*, vol. 1, 2016, pp. 543–47.

[28] T. McMahan, I. Parberry, and T. D. Parsons, "Evaluating player task engagement and arousal using electroencephalography," *Procedia Manufacturing*, vol. 3, pp. 2303–2310, 2015.

[29] A. M. Lund, "Measuring usability with the USE questionnaire," *Usability Interface*, vol. 8, no. 2, pp. 3–6, 2001.

[30] C. Harmon-Jones, B. Bastian, and E. Harmon-Jones, "The Discrete Emotions Questionnaire: A New Tool for Measuring State Self-reported Emotions," *PloS one*, vol. 11, no. 8, p. e0159915, 2016.