Physiologically Attentive User Interface for Improved Robot Teleoperation

António J. Tavares *

Institute for Systems and Robotics, Instituto Superior Técnico, 1049-001 Lisboa, Portugal, antonio.tavares@tecnico.ulisboa.pt

José Luís Silva

ITI/LARSyS, Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisboa, Portugal, jose.luis.silva@iscte-iul.pt

Rodrigo Ventura

Institute for Systems and Robotics, Instituto Superior Técnico, 1049-001 Lisboa, Portugal, rodrigo.ventura@isr.tecnico.ulisboa.pt

User interfaces (UI) are shifting from being attention-hungry to being attentive to users' needs upon interaction. Interfaces developed for robot teleoperation can be particularly complex, often displaying large amounts of information, which can increase the cognitive overload that prejudices the performance of the operator. This paper presents the development of a Physiologically Attentive User Interface (PAUI) prototype preliminary evaluated with six participants. A case study on Urban Search and Rescue (USAR) operations that teleoperate a robot was used although the proposed approach aims to be generic. The robot considered provides an overly complex Graphical User Interface (GUI) which does not allow access to its source code. This represents a recurring and challenging scenario when robots are still in use, but technical updates are no longer offered that usually mean their abandon. A major contribution of the approach is the possibility of recycling old systems while improving the UI made available to end users and considering as input their physiological data. The proposed PAUI analyses physiological data, facial expressions, and eye movements to classify three mental states (rest, workload, and stress). An Attentive User Interface (AUI) is then assembled by recycling a pre-existing GUI, which is dynamically modified according to the predicted mental state to improve the user's focus during mentally demanding situations. In addition to the novelty of the proposed PAUIs that take advantage of pre-existing GUIs, this work also contributes with the design of a user experiment comprising mental state induction tasks that successfully trigger high and low cognitive overload states. Results from the preliminary user evaluation revealed a tendency for improvement in the usefulness and ease of usage of the PAUI, although without statistical significance, due to the reduced number of subjects.

CCS CONCEPTS • Human-centered computing~Human computer interaction (HCI)~Interactive systems and tools~User interface programming • Computing methodologies~Artificial intelligence

Additional Keywords and Phrases: Attentive User Interface, Recycling User Interfaces, Human-Robot Interaction, Mental State Classification, Neural Networks, Robot Teleoperation

ACM Reference Format:

* Place the footnote text for the author (if applicable) here.

1 INTRODUCTION

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 must filter information from multiple sources, often with the drawback of doing it at a superficial level, effectively limiting their ability to interact appropriately with each system [1]. This problem brings in the need for smarter interfaces that understand human needs and adapt to them. To tackle this problem, Vertegaal 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 mental state, 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 Psychophysiology, it is possible to establish links between the human body and mind to acquire more reliable information about human cognitive and mental states [3]. The measurement of physiological signals allows the perception of covert mental states that are not visible to an AUI. Therefore, Chen 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 user's mental state [4].

The problem of managing user attention also expanded to the field of Human-Robot Interaction (HRI). The case study used in this work focuses on the remote teleoperation of robots, where situational awareness is still a major concern due to the limitations it can bring to teleoperation. This concern is particularly important in time-critical tasks, such as robot search and rescue. The ever-growing sophistication of robot tasks often comes with the drawback of requiring operators to deal with very complex interfaces. This can expose them to high stress and workload levels, compromising their focus and performance on the given task. This problem is crucial while performing critical tasks, where it is essential to guarantee that the operator's focus remains at its best. In the USAR context, the user interfaces are generally very demanding in terms of the visual feed they provide, which can cause operators to miss important information or rapidly reach cognitive exhaustion [5].

This paper presents the development of a novel PAUI that aims to tackle the problem of managing users' attention and focus (applied in a case study during robot teleoperation tasks), which is addressed through the application of affective computing techniques to the manipulation of the user interface in real time. The PAUI retrieves physiological signals with the intent of classifying the mental state of an operator during the teleoperation of a robot. The interface is then changed dynamically with respect to the user's predicted mental state. The PAUI is developed over a pre-existing robot's Graphical User Interface (GUI) through the extraction of images from said GUI, allied with the usage of a task automation tool that automates mouse and keyboard operations. This approach enables the adaptation of a PAUI to any pre-existing GUI without needing access to its source code, requiring only small efforts in terms of adjustment to the desired system. The novelty of this paper resides in the usage of attentive strategies that aim to reduce cognitive overload and improve the performance of end users by focusing on the adaptation of the user interface in real time according to the user's needs. The case study of the approach is illustrated with a PAUI specifically applied to a USAR robot developed in 2005 called RAPOSA [6] (see Fig. 1) and was elaborated upon the framework defined by Singh et al. [7] for a PAUI applied to this specific robot.

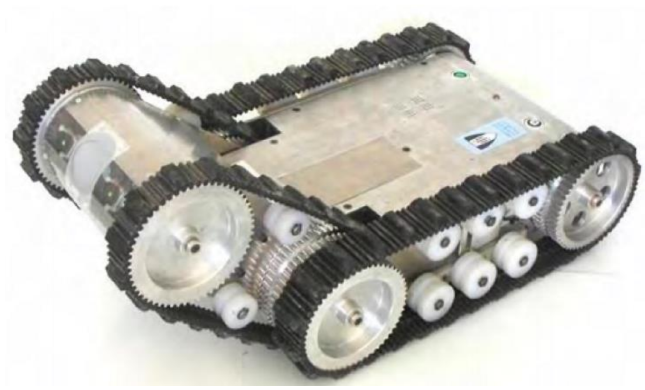


Figure 1: RAPOSA robot [6].

This work focuses on the development of a PAUI prototype applied to the robot teleoperation field, contributing with details about the strategies implemented to manage user's attention, the techniques adopted in the development of the mental state classifier, as well as the methodology undertaken to apply the PAUI to a specific case study (RAPOSA). Furthermore, the necessity of stimulating different emotions in the users resulted in the planning of tasks that successfully induced three distinct mental states (rest, workload, and stress). The workload is usually felt by operators (stress can be considered as an intense workload) [42], but additional mental states could have been considered. It should be noted that the rest mental state might always have some level of workload in real USAR. These tasks fall within the USAR context, accurately representing the circumstances commonly faced by operators in this field. Additionally, a preliminary evaluation was performed to offer insight as to whether the employment of a PAUI in this specific context can or cannot 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. The experiment conditions also led to the development of a simulator that enabled the creation of the robot's main functionalities and features, along with the modeling of home-like environments aiming to depict potential real USAR missions (home-like was the environment selected among different possible real USAR environments).

This paper is structured as follows. Section 2 presents related work. Section 3 describes the proposed solution to the problem of managing the user's attention and recycling the pre-existing GUI (applied to the field of robot teleoperation), offering a detailed view of the entire architecture of the system. The preliminary evaluation of the proposed solution is presented in Section 4, providing the user study results, followed by their respective analysis and discussion. Finally, Section 5 concludes the paper, highlighting its most important findings and indicating possible future improvements.

2 RELATED WORK

The literature review focuses on two areas: the design of attentive user interfaces, explored in sub-section 2.1, and the estimation of mental states regarding covert and overt signals of user attention in sub-section 2.2. Recycling existing GUIs was partially based on approaches of other works [47] [48].

2.1 Attentive User Interfaces

Vertegaal [2] proposed a framework for developing 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, progressively delivering them instead of forcing the information upon the user, potentially leading to a decrease in the user's focus.

Bulling [1] considered the management of user attention as a "critical challenge for next-generation human-computer interfaces". Human attention and focus are a limited resource that can play a very important role in the interfaces' performance [8]. 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, Long-livedness, Seamlessness, and Context-awareness as important categories that should be considered in the development of a new generation of pervasive attentive user interfaces. A study performed by Wintersberger et al. [9] analyzed the usage of an AUI and device integration in automated vehicles, where users can shift their attention between driving and non-driving related tasks. The AUI issues Take-Over Requests (TORs) only between tasks presented on the non-driving related device. Furthermore, the system enables the users to respond to the TOR directly on the handheld device in emergency situations. The study concluded that the usage of the AUI in automated driving environments contributes to reduced variance in response times, along with decreased stress levels in drivers. For an extensive discussion of the frameworks used to define AUIs and a review of the research on AUIs in the driving context, see the recent work of DeGuzman et al. [45].

Castellanos-Cruz et al. [10] studied the usage of eye gaze interfaces for controlling the guidance of telerobotic systems, with the aim of allowing physically impaired children to interact with toys. This study compared the employment of an AUI and an explicit eye gaze interface that predicts the target toy as the one the user fixates on for 500ms. On the other hand, the AUI uses a neural network trained to recognize the target toy based on the user's eye gaze and movements performed at a haptic robotic controller. While the results of a user study yielded a prediction accuracy of 86.4% for the AUI and 100% for the explicit eye gaze interface, the AUI was faster and more intuitive to use.

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 [11] 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 to differentiate four distinct user states that can be used to predict the user's availability for interruptions.

Significant research has also been done to improve user experience in the HRI field. Millan et al. [13] 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 [14] 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. [15] studied more than a dozen USAR robot interfaces, concluding that these interfaces contain large amounts of information, most of which is disregarded by the operators most of the time due to its irrelevance to the specific task at hands. Riley and Endsley [16] also expressed concern about 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. More recently, Delmerico et al. [12] stated that in rescue robotics research efforts should still focus on ease of use for improved adoption and real-time situational awareness. Towards those efforts, Roldán et al. [53] proposed an adaptive (to changes of scenario and robot) interface to real-world multi-robot scenarios that show improvement in workload, situational awareness, and performance.

It can be assumed that the problem of managing user attention is being addressed in the HRI field and has particularly affected the performance of USAR teams that use robots in their missions due to the difficulty and complexity associated with operating these robots. Although efforts have been made to improve the user experience in the operation of robots in general, USAR assistance robots still exhibit user interfaces that often express a developer's point of view and do not consider the cognitive load forced upon its operators or simply present too much information that is not relevant for most of the course of the mission. Moreover, a simple re-design of an existing interface might not prove to be an advantageous approach since some experienced operators might miss some of the information that was removed [15]. Operators may need to perform rescue operations during extended periods of time while interacting with these systems, leading to a degradation of their task performance, which can be critical in the context of scouting survivors. As such, this paper aims to address this problem by exploring mental state estimation techniques that enable dynamic manipulations of pre-existing robot interfaces in real-time with respect to the user's mental state.

2.2 Mental State Estimation

Recognizing a user's mental 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 on using facial expressions to distinguish different emotions [17] [18] [41]. With the growth of more accurate means of measuring physiological signals, the employment of such measurements is proving to be another good source of information on the mental state of a person [19] [20] [21] [22].

A study carried out by Kim et al. [23] performed emotion prediction on 50 subjects to classify four mental states (sadness, anger, stress, and surprise) while acquiring data from Electrocardiography (ECG), Electrodermal Activity (EDA) and skin temperature variation. This study showed that 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 four classes and 78.43% when classifying only three classes (sadness, anger, and stress). Wang et al. [24] 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. [25] 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%. Balan et al. [26] compared the performance of multiple machine learning and deep learning techniques (kNN, SVM, LDA, random forest, and deep neural networks) in the classification of six basic emotions (anger, disgust, fear, joy, sadness, and surprise), with and without feature selection, using the DEAP dataset. The performance analysis yielded a maximum accuracy of 98.02% for anger, 100% for joy, 96% for surprise, 95% for disgust, 90.75% for fear, and 90.08% for sadness, where the maximum accuracies for anger, disgust, fear, and sadness were obtained without feature selection.

For completeness, emotions recognition works performed between 2009 and 2016 using EEG signals can be consulted in the survey made by Alarcão et al. [49]. In the survey, the main aspects involved in the recognition process (e.g., features extracted, classifiers) were analyzed. More recently, Rahman et al. [50] also reviewed and compared existing studies (up to 2021) on emotion detection using EEG signals.

A combination of behavioral and physiological measures is another approach that has been followed. For instance, Singh et al. [46] use this combination for mental workload estimation in the context of pilot unmanned aerial vehicles teaming applications. Heard et al. [51] suggested several sensor-related metrics that seem to be relevant to mental workload assessment. Furthermore, Debie et al. [52] reviewed studies related to multimodal data fusion (e.g., electroencephalography, electrocardiography, eye tracking) to estimate the cognitive workload.

3 APPROACH

In this work, a PAUI was designed to address the following challenges: 1) coping with the overload of information caused by user interfaces in mobile robot teleoperation (illustrated in the USAR context); 2) adapting a pre-existing user interface to the PAUI format without having access to the original interface's source code, which is essential since most user interfaces are licensed commercially and cannot be modified directly. Regarding the first challenge, the solution proposed is to employ a machine learning approach based on an artificial neural network classifier trained to classify the operator's mental state between three different mental states (Rest, Stress, and Workload). The mental state prediction is then used to dynamically change the user interface in real time to improve the user's focus during cognitive overload moments. As stated by Lavie et al. [43], the levels of adaptivity should also be implemented with care due to the potential negative effects they might pose. Nevertheless, adaptive user interfaces already showed usability and safety benefits, such as reduced cognitive load [44] [53]. The PAUI is established over the pre-existing robot GUI with the purpose of reducing its complexity and displaying information more clearly, thus decreasing the cognitive overload of the operator. For the second challenge, the PAUI interacts with the old GUI using an automation tool (SikuliX [27]) and extracts its image to render an AUI based on the original interface, which can then be modified through a rendering API. Therefore, the proposed solution does not require access to the original GUI's source code, which eliminates the need to develop a new interface or a completely new solution from scratch and enables the rearrangement of very old systems, such as RAPOSA, that are already established and can be hard to reprogram, even if access to the source code is granted in the first place [28] [29]. Here, the PAUI is applied to the case study of RAPOSA, which provides a 15-year-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. Although the PAUI is on top of the pre-existing GUI, our approach does not significantly increase the time delay that might have affected the operators' situational awareness during operations.

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 Mental State Classifier, and the Attentive User Interface. Fig. 2 presents the overall architecture of the system.

3.1.1Signal Extractor

The signal extractor is responsible for acquiring data that can potentially yield valuable information about a person's mental state. 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. The Tobii Eye Tracker 4C by Tobii Technology was used to detect eye movements. Finally, for facial expressions and emotions, the laptop-integrated webcam was used alongside the AFFDEX SDK by Affectiva [30].

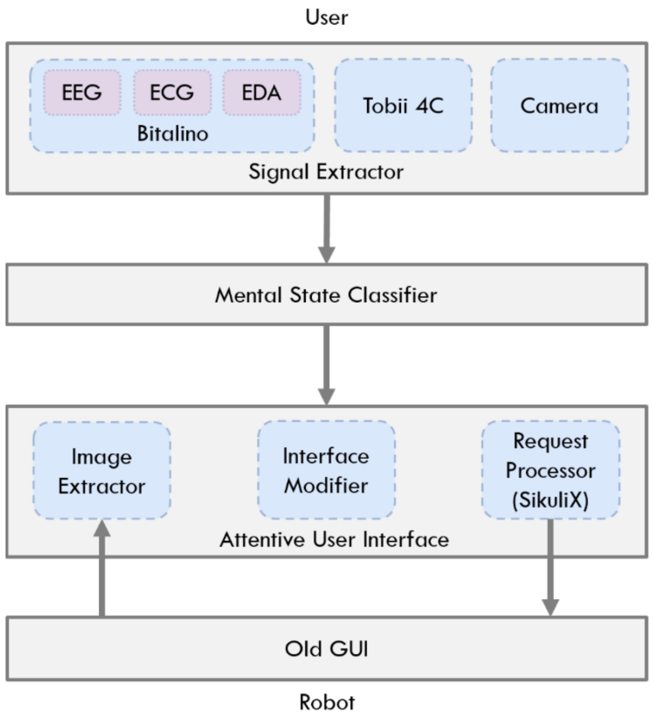


Figure 2: PAUI Architecture.

The extracted data is then processed to obtain a wide range of parameters (a complete list of the processed parameters is given in Table 1, Appendix A) that can later be used as input for training a classifier. Data processing was performed with the aid of the pre-existing PAUI framework previously developed by some of the co-authors of this paper [7], 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. A complete list of all the parameters resulting from processing the extracted data can be found in Table 1, Appendix A.

3.1.2 Mental State Classifier

This module takes charge of classifying the user's mental state based on the signals received from the signal extractor. In this work, the classifier predicts the mental state from three possible classes: rest, stress, and workload. These states were chosen as a basis for this work, as they target three possible sensations felt by operators in teleoperation situations, which were also used in the work that preceded this paper [7]. More states could be of interest, such as transitioning states between each of these, which might be explored in future works. A full rest might not be present in real USAR operations but it is a mental state usually present in most case studies.

As referred in the previous section, the system runs a thread for each data extraction device in charge of acquiring and processing its respective signals. Likewise, the mental 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 averages the signals received during 1 second. The thread then uses the averaged signals as input for the classifier, which predicts the operator's mental state.

In this work, the classifier used was an artificial neural network. This classifier is trained offline with data collected for that purpose and is then used in real-time to classify the user's physiological state.

Since the model's objective is to predict the user's mental 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. The dataset is divided into three classes: rest, stress, and workload. The data corresponding to each class results from the concatenation of the collected data, ordered by timestamp. As the dataset collected in this work is class-balanced, the metric chosen to measure the model's success was accuracy. Since the amount of data available for this study was relatively small, a K-fold cross-validation was adopted as the validation protocol, with a value of K = 4 being chosen according to the recommendations of Anguita et al. [31]. Considering the data collected for the purposes of this work is time-series data, where the user's signals were collected over time for each class, the data was split before shuffling to avoid exposing the network to observations from the whole array of time during training [32]. The split was done in a 75/25 ratio, using the first 75% of each class for the training set and the remaining 25% for the test set. The validation set split is done similarly to the train test split, where each fold i generates a validation set that contains data from the ith part of each class. After doing the validation split for each fold, the remaining training data is then shuffled before proceeding to train the network. Given that the features present in the dataset used for this work take up values in significantly different ranges, a data normalization procedure was carried out. It should be noted that the mean and standard deviation used to normalize the validation and test sets were computed using the training set. This measure prevents the leakage of information about the validation and test sets to the model during training.

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 [32]. 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. Concerning 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, such as Nesterov accelerated gradient, AdaGrad and RMSProp [33]. The default initialization values for all the hyper-parameters except the learning rate were used: 0.9 for the momentum decay hyper-parameter (β1), 0.999 for the scaling decay hyper-parameter (β2) and 10-8 for the smoothing term (ϵ) [34].

Over the course of the model's training, three different regularization methods were tested to reduce the overfitting of the model:

1) **Dropout**: The output of a random subset of neurons is set to zero, as well as their connections with other neurons;
2) **L2 regularization**: A cost on the square of the model's weight values is introduced in the cost function, leading to a preservation of the parameters that contribute significantly to the minimization of the cost function, whereas weights that might not be so informative tend to decay to smaller values:

$$\hat{J}(w) = J(w) + \frac{1}{2}\lambda|w|^2 \qquad (1)$$

Here, J^(w) denotes the regularized cost function, J(w) is the original cost function, w represents the model weights, and λ is a hyper-parameter that sets the amount of regularization to be applied.

3) **L1 regularization**: A cost proportionate to the absolute value of the weights is applied, causing some weights to shrink to values very close to zero, leading to a sparser solution:

$$\hat{J}(w) = J(w) + \lambda|w| \qquad (2)$$

During the training process, three values of λ were tested: 0.001, 0.005, and 0.01. The value that yielded better results was 0.01, both for L1 regularization and L2 regularization. Higher values for λ were also tested, which led to underfitting due to the aggressive shrinkage of coefficients to zero. In the dropout case, a dropout layer was added after each hidden layer of the network, which randomly dropped 20% of the neurons of that layer. Additionally, early stopping was also used to find the optimal number of training epochs by stopping the training process when the model performance on the validation set stopped improving, i.e., the validation loss ceased to decrease. The neural network architecture (i.e., the number of hidden layers and hidden neurons), along with the choice of regularization methods and the results obtained for each of them, are explored further in section 4.7.1.

3.1.3 Attentive User Interface

The AUI module is responsible for managing the user's attention and easing the process of operating the robot in cognitive overload 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 mental state (see Fig. 3).

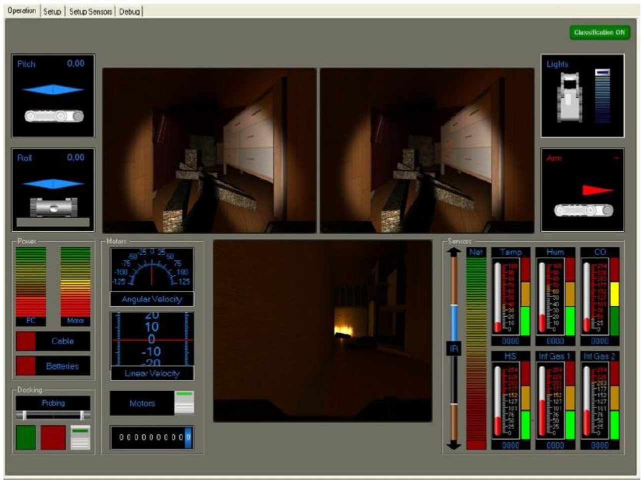
The AUI also issues requests to the old interface with the aid of a task automation tool called SikuliX, which enables the automation of mouse and keyboard operations, effectively allowing the AUI to interact actively with the old interface. Intending to hide the old interface from the user's view while maintaining the interaction between both interfaces, the old interface runs on a virtual machine 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 mental 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. Fig. 3(a) shows the appearance of the original RAPOSA's graphical interface.

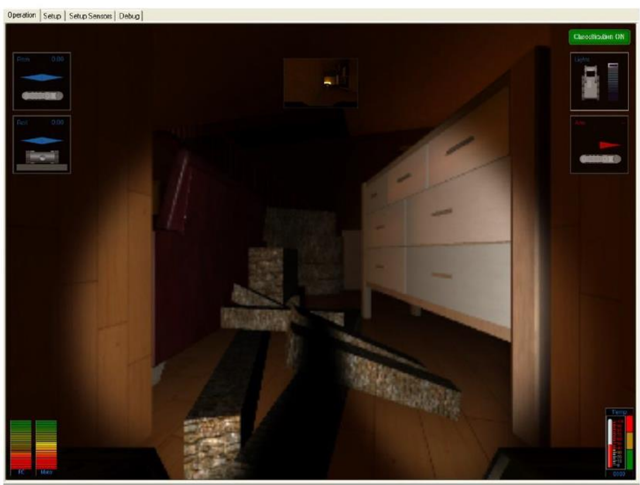
When the classifier predicts the mental state of the operator as stress, the rendered interface remains to be the original GUI, but the AUI issues a request to the old GUI for an increase of the maximum speed of the robot. This request is executed by a SikuliX script that automates a series of clicks programmed to change the maximum speed value setting in the Setup tab of the old GUI, taking an average of 540ms to execute (see Fig. 4).

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 robot's maximum speed 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 always set to the limit, 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 robot's camera view 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. Fig. 3(b) shows the new interface displayed in workload situations.



(a) Original RAPOSA's GUI.



(b) Redefined AUI rendered in workload situations.

Figure 3: Simplified visualization of the PAUI concept. On the top, the complexity of the original interface is shown, with only the relevant elements of the interface highlighted in light blue. When the operator is under cognitive workload, the system updates the interface to a simplified version aiming to improve the operator's task performance, shown at the bottom.

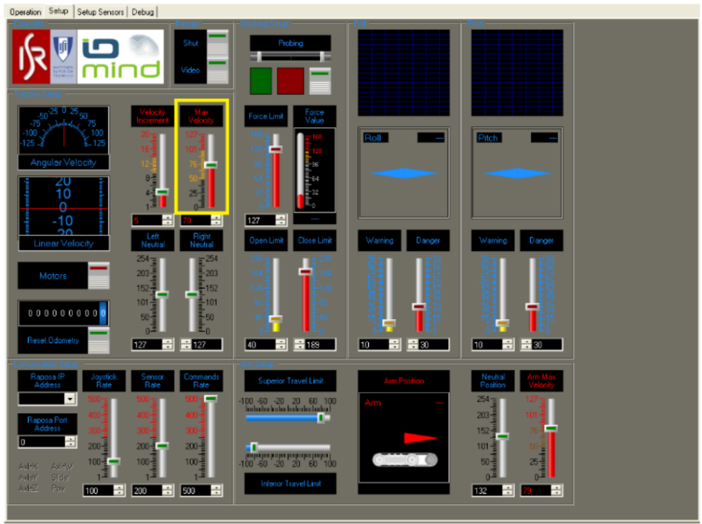


Figure 4: Setup tab of the RAPOSA's interface. When the classifier predicts the mental state as stress, a SikuliX script is executed, which accesses the old GUI, switches to the Setup tab, clicks in a desired position in the maximum velocity slider, and switches back to the Operation tab.

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 the user's attention are specific to the interface in question. As such, the procedures presented previously are specific to the RAPOSA's interface. As far as the authors acknowledged, there is a lack of studies in this specific area, leading the attentive strategies applied for each mental state to be chosen mainly out of intuition. Regarding the stress state, it was decided not to redefine the interface since a stressful and anxious state is more likely to be caused by a difficult teleoperation environment rather than by the interface itself. As such, the user might benefit more from a raised maximum speed than an improved interface. On the other hand, the workload state is likely to be triggered when the interface is being a cause of distractions (e.g., sensor danger levels and low batteries). As such, the simplification of the interface and the display of certain interface elements (e.g., sensors and batteries) only beyond certain levels can contribute to maintaining the user's focus on the task since the interface only demands more attention when it is absolutely essential.

This paper is a follow-up research building upon the framework presented in [7], it follows a very similar architecture, and only data processing was aided by pre-existing work. This paper provides much additional work and many significant scientific contributions beyond what was done before, namely:

- a detailed description of the new mental state classifier (using a new dataset) that was not presented in other works. In the previous work, the classifier was an SVM. In this work, the classifier used was an artificial neural network (contrary to previous work, eye tracking features were not considered in the mental state classifier training) that was employed with the aim of improving classification accuracy, considering Deep Learning techniques have been successful at solving problems in recent years;
- a prototype of a PAUI for robot teleoperation, offering details about the strategies implemented to manage the user's attention as well as the methodology to assemble the components together to create a PAUI on top of the pre-existing robot's interface. Although being applied to the RAPOSA's interface, the system can be adapted to other interfaces. In the previous work, the focus was on the Emotional State Estimator, as the pipeline was not

completed. Therefore, the Emotional State Estimator was the unique component evaluated (with two games, i.e., Rigs of Rods and a modified Tetris game). The other components were not implemented;

- design of emotion induction tasks. The necessity of stimulating different emotions in the users resulted in the planning of tasks that successfully induced three distinct emotional states (rest, stress, and workload). These tasks fall within the USAR context, accurately representing the circumstances commonly faced by operators in this field;
- USAR Simulator, i.e., the development of a Search and Rescue simulator that enabled the re-creation of the robot's main functionalities and features, along with the modeling of home-like environments that depict real Search and Rescue missions;
- user study performed to evaluate the research statement hypotheses, including a detailed procedure of the experimental sessions, followed by the results obtained and the respective discussion.

4 PRELIMINARY EVALUATION

As referred to in section 1, the objective of the preliminary evaluation is to infer about the potential of the PAUI to improve the usability of the interface when operators are submitted to high cognitive overload situations while improving their ability to stay focused. As such, the following hypotheses were drawn:

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*.

To evaluate the proposed solution, two experimental sessions were carried out: a data collection session and an evaluation session. The purpose of the data collection session was to collect a dataset of facial expressions, eye movements, and physiological signals of each user that could be used to train the mental state classifier. The classifier was then used to evaluate the PAUI approach in the evaluation session. The necessity of acquiring a new dataset in the data collection session arises from the fact that physiological data varies significantly from person 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. [7] 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 mental state classification is not the main focus of this work.

On a separate note, the current global pandemic impeded the access to hardware (RAPOSA), which led to the development of a USAR environment simulator using Unity, making the evaluation of the proposed solution possible. The Unified System for Automation and Robot Simulation (USARSim) could have been a simpler solution, but the authors did not identify it. The simulator allowed the creation of a mock-up of the RAPOSA's original interface and controls, as well as the development of custom Search and Rescue scenarios that were used to induce stress and workload to users during a robot teleoperation task. Additionally, the simulated environments facilitated the creation of stress and workload situations that would not be so easily accessible in a physical environment. The evaluation using a simulated environment has limitations compared to an evaluation using the physical environment (e.g., not being the most adequate to measure classification accuracy) nevertheless, this led, as stated, to additional contributions. Furthermore, evaluation results should be interpreted in concordance.

4.1 Subject Grouping

For the purposes of the second experimental session (evaluation session), the subjects were split into two independent groups for a between-subjects study design, where one group tested the PAUI approach and the other group tested the GUI approach, i.e., the original robot's GUI without attentive measures with respect to the user's mental state. 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. One solution to the usage of a within-subjects approach would be to have each subject test both approaches but with a different environment in each. However, this could lead to biased results since one environment could potentially be harder to operate in than the other, which led to the employment of a between-subjects approach. To mitigate the variance in the results as much as possible, the Immersive Tendencies Questionnaire (ITQ) [36] was used, which reliably reflects each subject's tendency to become more involved in virtual environment tasks, possibly displaying a better performance. The scores obtained in the questionnaire were then used to split the subjects into 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 to in section 3.1.1. Regarding the attachment of electrodes, the best-suggested placement by Nemcová et al. [37] was used for ECG: positive lead under the right clavicula, negative lead under left musculus pectoralis major, and reference lead under the left clavicula. For EEG, the positive and negative leads were placed at the forehead, and the reference lead at the left earlobe. For EDA, only two electrodes are required, which were placed in the left-hand palm [7]. Apart from the data acquisition devices, a conventional gamepad was used to control the robot without force feedback.

4.3 Performed Tasks

To evaluate the proposed solution, a task meant to induce each of the three targeted mental 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 (see Fig. 5). 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. Intending to trigger 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 (see Fig. 6).

.

Figure 5: Aerial view of the simulated home environment during the rest task (first floor).

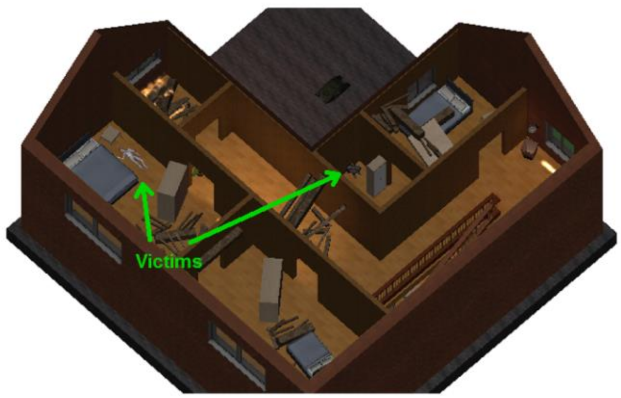


Figure 6: Aerial view of the simulated home environment during the stress task (second floor).

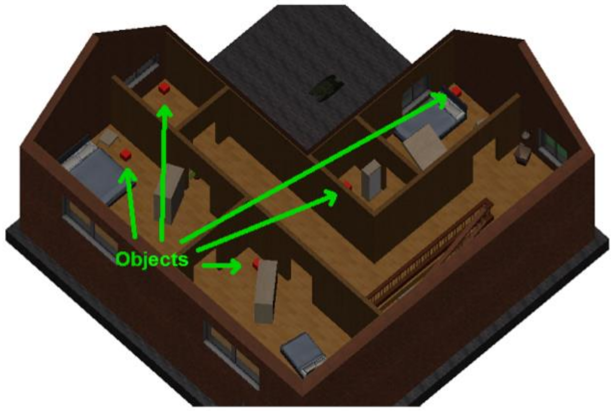


Figure 7: Aerial view of the simulated home environment during the workload task (second floor).

In the workload task, participants were required to find ten objects (represented by red cubes) inside a house (see Fig. 7) while answering workload inducing questions. These questions included basic arithmetic operations, requests to read the values of sensors in the interface, questions about the robot's surroundings, and basic logic problems. It should also be noted that participants were reminded that there was no time limit to answer each question, to avoid inducing unnecessary stress.

In the stress and workload tasks, the differentiation between victims and objects aims to induce different sensations since the usage of victims might lead the user to perceive the task as more dangerous or important, whereas the objects used in the workload task simply represent checkpoints that need to be verified.

4.4 Metrics

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. [38], 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 Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire [39] were 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) [40] 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 approximately 1 hour to complete. Upon arrival at 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 were presented. Subsequently, the participant filled a demographics questionnaire, followed by the ITQ (this step was exclusive to the first experimental session). Afterward, the eye tracking device was calibrated, and the ECG, EEG, and EDA electrodes were attached to the subject according to the configurations referred to in section 4.2.

After going through all the necessary preparations, the subject went through a training session 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 randomly. Upon the completion of each task, the user went through a 2-minute break period to relax and come back to a normal state.

Regarding the second experimental session, after finishing the teleoperation tasks, the subject was also asked to fill out 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 GUI approach, the mental state classifier still predicted the mental 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 six voluntary participants (4 male, 2 female) aged between 20 and 24 years old (M = 22.833, SD = 1.472). The participants were not paid, were Portuguese university students with a background in engineering without prior experience in robot teleoperation, and received informed consent.

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 to better understand the patterns present in the data. Inconformity with the findings of other researchers [20] [24], 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 Neighbor 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 Fig. 8).

It should also be noted that the eye tracking features were not considered in the mental 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 to reduce its dimensionality. Due to the inherent tendency of physiological signals to contain large amounts of noise [25], it was decided that it would be best to keep most of the explained variance present 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.

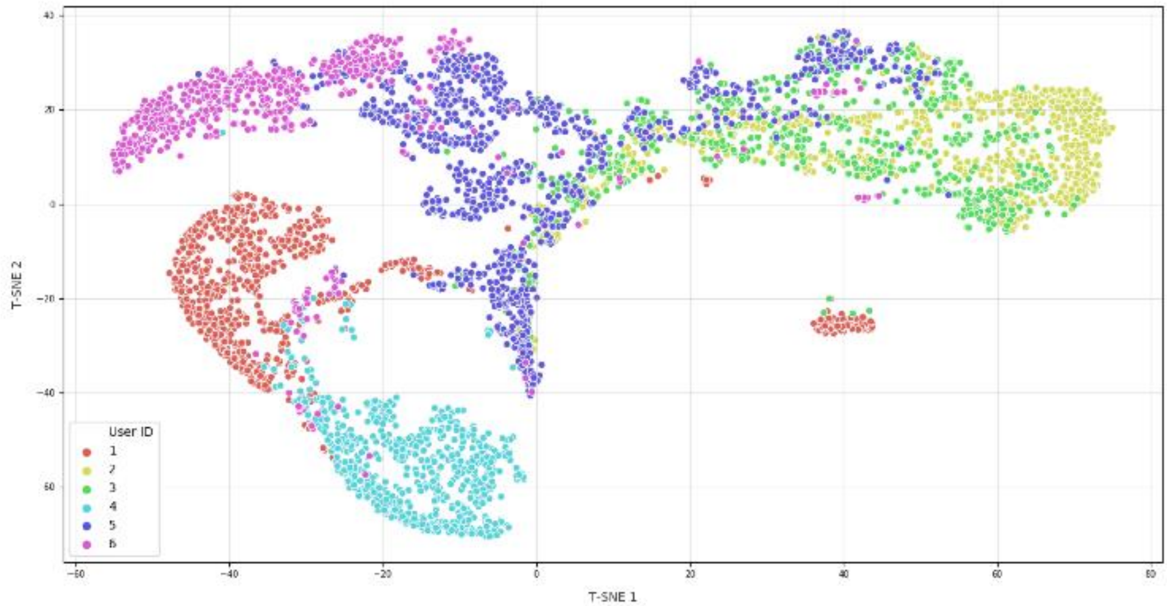


Figure 8: Embedded space that results from the application of t-SNE to the data collected from each user during the stress task, where each user is represented by a different color.

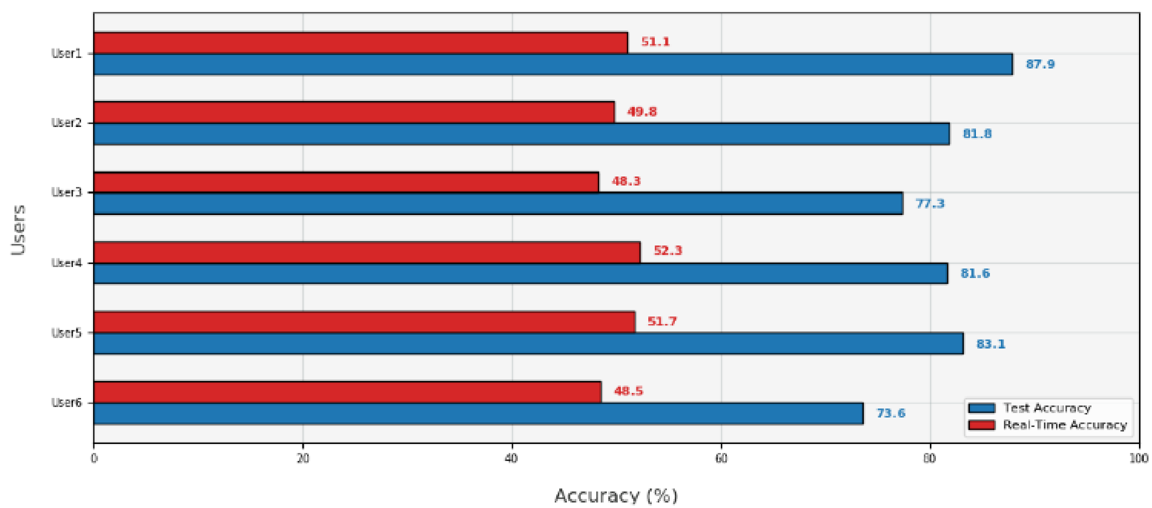


Figure 9: Comparison of the classification accuracies obtained in the test set and in real-time classification for each user.

After projecting the dataset on the new feature space, a neural network model was trained for each user. This neural network has 45 inputs, one for each fearture and 3 outputs, one for each class (see section 3.1.2). Roughly, each network was trained with 2700 samples, corresponding to 75% of the collected data along the 1-hour session.

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 fully connected network with three hidden layers (40, 30, and 20 hidden neurons), where three different regularization methods were tested: L1 regularization, L2 regularization, and dropout. With the application of early stopping, L2 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 4.8). The accuracies obtained for each participant in the test set and in real-time are compared in Fig. 9.

4.7.2 Group Division

Regarding the division into 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 to generate two balanced groups that contained 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 into the advantages of using a PAUI against the usage of a traditional GUI approach.

4.7.3 Statistical Tests

With the aim of analyzing the gathered data on the task performance obtained under both approaches (i.e., PAUI group vs. GUI group), 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. As such, an independent t-test was applied to the completion time and to the relative changes of the engagement levels, whereas a Mann-Whitney U test was applied to the number of objects found. All three tests revealed no statistically significant differences in the results obtained between the PAUI approach and the GUI approach:

- **Completion Time**: Independent t-test ($t(4) = -0.541$, $p = 0.617$);
- **Engagement Levels**: Independent t-test ($t(4) = 1.504$, $p = 0.207$);
- **Objects Found**: Mann-Whitney U test ($U = 3.5$, $p = 0.658$).

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 no statistically significant 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.

4.8 DISCUSSION

The results presented show a significant drop in terms of performance of the mental state classifiers in the evaluation session, when compared to the accuracies obtained in the test set after the data collection session, 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 compared to the GUI users. These results can be indicative of 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. Nonetheless, it should also be pointed that user reported emotions can be slightly different from real emotions, which can make the results somewhat unreliable, especially in experiments with a small number of participants such as the one presented. Additionally, even though the attentive strategies adopted focused on improving the user's interaction with the interface, it is also possible for users to have a biased preference for the PAUI simply because it is dynamic. As such, due to the lack of statistical significance, hypothesis H4 cannot be assured.

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. However, potentially there is only a small pool of USAR robotics specialists; thus having a customized interface would be beneficial. Furthermore, the design of the experiment proved to be effective in replicating the pretended mental 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 future studies that use both the 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 mental 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. Another consideration is that, even though the PAUI is to be used by experts, the users that were available for evaluating the system were not experts. This can have an impact on the perceived usefulness of the attentive strategies of the interface since expert users might miss some of the interface elements in redefined formats, such as the one applied in the workload state. Ultimately, it would be of interest if the system's attentive measures could be customized by users, as it would solve the problem of the changes being useful for some users but not for others.

Summing up, the small number of subjects available and the poor classification performance were the main limitations of the 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.

5 CONCLUSIONS

This work presented the development of a PAUI prototype applied to the pre-existing RAPOSA's interface, where a significant emphasis was placed on creating effective methods of managing user's attention in moments of cognitive workload, considering 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 and, in particular for USAR robots. The planning of tasks designed to induce specific mental 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 modeling 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.

Through the execution of a preliminary 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 compared 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.

Furthermore, the results obtained considering the small number of subjects available, allied with the limited capabilities of the mental state classifier, are indicative of potential improvements in future work. The achievement of a robust mental state classifier is by itself a challenging task, and the results clearly show the need to obtain 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 emphasizing the eye tracking device, which can give insight into the regions of the interface where users tend to stare most. Moreover, it could be beneficial to ease the process of adapting the system to other case studies to increase the flexibility of the solution. Finally, the execution of future evaluation studies with the real robot could help provide users with a broader teleoperation experience since all the robot's features are available, allowing the employment of attentive measures that offer a greater contribution to the task performance achieved.

ACKNOWLEDGMENTS

[Removed]

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] R. Murphy, S. Tadokoro, D. Nardi, A. Jacoff, P. Fiorini, H. Choset, and A. Erkmen, Search and Rescue Robotics, 2008, pp. 1151–1173.

[6] C. Marques, J. Cristóvão, P. Alvito, P. Lima, J. Frazão, 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.

[7] 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.

[8] 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.

[9] P. Wintersberger, C. Schartmüller, and A. Riener, “Attentive User Interfaces to Improve Multitasking and Take-Over Performance in Automated Driving: The Auto-Net of Things,” International Journal of Mobile Human Computer Interaction (IJMHCI), vol. 11, no. 3, pp. 40–58, 2019.

[10] J. L. Castellanos-Cruz, M. F. Gómez-Medina, M. Tavakoli, P. M. Pilarski, and K. Adams, “Comparison of Attentive and Explicit Eye Gaze Interfaces for Controlling Haptic Guidance of a Robotic Controller,” Journal of Medical Robotics Research, vol. 4, no. 03n04, p. 1950005, 2019.

[11] 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.

[12] Delmerico, J., Mintchev, S., Giusti, A., Gromov, B., Melo, K., Horvat, T., ... & Scaramuzza, D. (2019). The current state and future outlook of rescue robotics. Journal of Field Robotics, 36(7), 1171-1191.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] J.Wagner, J. Kim, and E. André, “From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification,” in 2005 IEEE International Conference on Multimedia and Expo. IEEE, 2005, pp. 940–943.

[22] J. Chen, D. Jiang, and Y. Zhang, “A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification,” IEEE Access, vol.7, pp. 118 530–118 540, 2019.

[23] 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.

[24] 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.

[25] 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.

[26] O. B˘alan, G. Moise, L. Petrescu, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, “Emotion Classification Based on Biophysical Signals and Machine Learning Techniques,” Symmetry, vol. 12, no. 1, p. 21, 2020.

[27] R. Hocke. SikuliX. (2017) [Online]. Available: http://www.sikulix.com,Accessed on: Jun. 28, 2020.

[28] J. L. Silva, J. D. Ornelas, and J. C. Silva, "Make it ISI: interactive systems integration tool," in Proceedings of the 8th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, 2016, pp. 245-250.

[29] R. Maia, J. C. Silva, and J. L. Silva, "Towards Graphical User Interface Redefinition without Source Code Access: System Design and Evaluation," in Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, 2019, pp. 1-6.

[30] 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.

[31] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, and S. Ridella, “The ’K’ in K-fold Cross Validation,” in ESANN, 2012.

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

[33] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[34] A. Géron, Hands-On Machine Learning with Scikit-Learn & TensorFlow. O’Reilly Media, 2017.

[35] ISO (2018) Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts (ISO/DIS 9241-11:2018).

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

[37] A. Nemcová, L. Maršánová, and R. Smíšek, “Recommendations for ECG acquisition using BITalino,” in Conference Paper EEICT, vol. 1, 2016, pp.543–47.

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

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

[40] 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.

[41] Pedro, T. M. S., & Silva, J. L. (2021). Towards higher sense of presence: a 3D virtual environment adaptable to confusion and engagement. IEEE Access, 9, 8455-8470.

[42] Shao, S., Zhou, Q., & Liu, Z. (2021). Study of mental workload imposed by different tasks based on teleoperation. International Journal of Occupational Safety and Ergonomics, 27(4), 979-989.

[43] Lavie, T., & Meyer, J. (2010). Benefits and costs of adaptive user interfaces. International Journal of Human-Computer Studies, 68(8), 508-524.

[44] Tchankue, P., Wesson, J., & Vogts, D. (2011, November). The impact of an adaptive user interface on reducing driver distraction. In Proceedings of the 3rd international conference on automotive user interfaces and interactive vehicular applications (pp. 87-94).

[45] DeGuzman, C. A., Kanaan, D., & Donmez, B. (2022). Attentive user interfaces: Adaptive interfaces that monitor and manage driver attention. In User Experience Design in the Era of Automated Driving (pp. 305-334). Springer, Cham.

[46] Singh, G., Chanel, C. P., & Roy, R. N. (2021). Mental workload estimation based on physiological features for pilot-UAV teaming applications. Frontiers in Human Neuroscience, 15.

[47] Silva, J. L., Ornelas, J. D., & Silva, J. C. (2016, June). Make it ISI: interactive systems integration tool. In Proceedings of the 8th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (pp. 245-250).

[48] Maia, R., Silva, J. C., & Silva, J. L. (2019, October). Towards graphical user interface redefinition without source code access: System design and evaluation. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 1-6).

[49] Alarcao, S. M., & Fonseca, M. J. (2017). Emotions recognition using EEG signals: A survey. IEEE Transactions on Affective Computing, 10(3), 374-393.

[50] Rahman, M. M., Sarkar, A. K., Hossain, M. A., Hossain, M. S., Islam, M. R., Hossain, M. B., ... & Moni, M. A. (2021). Recognition of human emotions using EEG signals: A review. Computers in Biology and Medicine, 136, 104696.

[51] Heard, J., Harriott, C. E., & Adams, J. A. (2018). A survey of workload assessment algorithms. IEEE Transactions on Human-Machine Systems, 48(5), 434-451.

[52] Debie, E., Rojas, R. F., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., ... & Abbass, H. A. (2019). Multimodal fusion for objective assessment of cognitive workload: a review. IEEE transactions on cybernetics, 51(3), 1542-1555.

[53] Roldán, J. J., Peña-Tapia, E., Garcia-Aunon, P., Del Cerro, J., & Barrientos, A. (2019). Bringing adaptive and immersive interfaces to real-world multi-robot scenarios: Application to surveillance and intervention in infrastructures. Ieee Access, 7, 86319-86335.

A APPENDIX – EXTRACTED PARAMETERS

Table 1: List of processed parameters that result from the extracted signals.

Sensor	Parameters
ECG	• Heart Rate • Standard Deviation of Normal to Normal • Root Mean Square of the Successive Differences • Very Low Frequency (0.0033 Hz - 0.04 Hz) • Low Frequency (0.04 Hz - 0.15 Hz) • High Frequency (0.15 Hz - 0.4 Hz)
EEG	• Delta (0.5 Hz – 3.5 Hz) • Theta (3.5 Hz – 8 Hz) • Alpha (8 Hz – 13 Hz) • Beta (13 Hz – 30 Hz) • Gamma (30 Hz – 45 Hz) • Engagement (Beta / (Alpha + Theta)
EDA	• Skin Conductance Level • Skin Conductance Response

Tobii	• Number of Fixations • Total Time • Total Fixation Duration • Average Fixation Duration • Repeated Fixations • Biggest Fixation At • Maximum Visited Counts • Maximum Visited At
Camera (Emotions)	• Joy • Fear • Disgust • Sadness • Anger • Surprise • Contempt • Valence
Camera (Expressions)	• Smile • Inner Brow Raise • Brow Raise • Brow Furrow • Nose Wrinkle • Upper Lip Raise • Lip Corner Depressor • Chin Raise • Lip Pucker • Lip Press • Lip Suck • Mouth Open • Cheek Raise • Dimplier • Eye Widen • Jaw Drop • Lip Tighten • Lip Stretch • Smirk • Eye Closure • Attention
Camera (Face Orientation)	• Pan • Tilt • Yaw