Workload-Driven Modulation of Mixed-Reality Robot-Human Communication

Leanne Hirshfield
Institute of Cognitive Science
University of Colorado
Boulder, CO, USA
leanne.hirshfield@colorado.edu

Tom Williams
MIRRORLab
Colorado School of Mines
Golden, CO, USA
twilliams@mines.edu

Natalie Sommer
Newhouse MIND Lab
Syracuse University
Syracuse, NY, USA
nmsommer@syr.edu

Trevor Grant
Newhouse MIND Lab
Syracuse University
Syracuse, NY, USA
tjgran01@syr.edu

Senem Velipasalar Gursoy
Computer Science Department
Syracuse University
Syracuse, NY, USA
svelipas@syr.edu

ABSTRACT

In this work we explore how Augmented Reality annotations can be used as a form of Mixed Reality gesture, how neurophysiological measurements can inform the decision as to whether or not to use such gestures, and whether and how to adapt language when using such gestures. In this paper, we propose a preliminary investigation of how decisions regarding robot-to-human communication modality in mixed reality environments might be made on the basis of humans' perceptual and cognitive states. Specifically, we propose to use brain data acquired with high-density functional near-infrared spectroscopy (fNIRS) to measure the neural correlates of cognitive and emotional states with particular relevance to adaptive human-robot interaction (HRI). In this paper we describe several states of interest that fNIRS is well suited to measure and that have direct implications to HRI adaptations and we leverage a framework developed in our prior work to explore how different neurophysiological measures could inform the selection of different communication strategies. We then describe results from a feasibility experiment where multilabel Convolutional Long Short Term Memory Networks were trained to classify the target mental states of 10 participants and we discuss a research agenda for adaptive human-robot teams based on our findings.

ACM Reference Format:
Leanne Hirshfield, Tom Williams, Natalie Sommer, Trevor Grant, and Senem Velipasalar Gursoy. 2018. Workload-Driven Modulation of Mixed-Reality Robot-Human Communication. In *Proceedings of International Conference on Multimodal Interaction: Workshop on Modeling Cognitive Processes from Multimodal Data (ICMI:MCPMD'18).* ACM, New York, NY, USA, 8 pages.

ICMI:MCPMD'18, October 2018, Boulder, CO, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.

1 INTRODUCTION

For robots to engage in effective and natural interactions with human teammates, they must be able to effectively communicate their intentions. As for human-human communication, *natural language* has emerged as the primary channel for human-robot communication. Natural language is infinitely flexible, requires minimal training, and requires no hardware beyond a common microphone and speaker.

When planning natural language communication with human teammates (e.g., regarding objects, locations, people, and goals in the robot and humans' shared context), robots must choose between a number of different communication strategies. First, the robot must choose whether or not to communicate at all: a robot may have frequent opportunity to pose questions, for example, but may need to weigh the time sensitivity and potential information gain of such questions against the availability and temperament of its potential addressee. If the robot does choose to go forward with communication, it may decide whether to describe a target referent using an anaphoric form like "it", or to do so using a full noun phrase like "the red box", trading off between concision and specificity. Finally, the robot may elect to expend energy to accompany its language with a deictic gesture, in order to draw its teammate's gaze (and thus, visual attention) towards its target referent.

These decisions must be made by any language-capable robot operating in *pure reality* environments. However, the human-robot interaction community has recently seen an explosion of work envisioning the potential for human-robot interaction in *Mixed Reality* environments [57]. *Mixed Reality* is defined as the subset of the Virtual Reality continuum that is neither entirely real nor entirely virtual [35], comprised of *Augmented Reality*, in which computer-generated visualizations are overlaid over a user's view of the real world, and *Augmented Virtuality*, in which real-world entities are overlaid over a user's view of a virtual world.

In our own recent work, we have explored the parallels between augmented reality annotations and traditional forms of robotic gesture, developing the first framework for categorizing the space of gestures available in Mixed-Reality human-robot interactions, including both traditional physical gestures, augmented reality annotations that achieve gestural goals [55, 58], and combinations thereof. In that work, we presented a preliminary analysis of how these gestural categories differ with respect to dimensions such as perspective,

embodiment, capability, privacy, cost, and legibility, suggesting different contexts in which different types of gestures may or may not be appropriate.

In this work, we consider how the use of another technology, neurophysiological sensing, may be helpful in making these difficult communicative decisions. Factors such as frustration may be helpful in determining whether a human teammate is in need of assistance; factors relating to cognitive load may be helpful in determining whether a human teammate has the capacity to accept new information; and factors like visual and auditory perceptual load specifically may be informative in choosing which communication modality will be most effective.

Advances in Augmented Reality and Neurophysiological sensing make this combination of technologies newly possible. Neurophysiological sensors such as *functional Near-Infrared Spectroscopy (fNIRS)*, which have become widely available, are lightweight, non-invasive, and offer improved spatial resolution compared to other technologies. Similarly, Augmented Reality Head-Mounted Displays (HMDs) are becoming lightweight, untethered, and consumer-grade, and in fact are already being deployed in a variety of industrial fields, such as warehousing [48]. There has been recent work on neurophysiological control of robots [e.g., 8, 18, 36], or neurophysiologically informed adaptations in robot behavior [49]. There has also been some recent work combining Augmented Reality and neurophysiological technologies within single-helmet hardware configurations [6]. However, that latter work has focused on the use of Augmented Reality for the generation of language *replacing* holograms. In contrast, we are interested in how Augmented Reality annotations can be used as a form of *language-accompanying* Mixed Reality gesture, and how neurophysiological measurements can inform the decision as to whether or not to use such gestures, and whether and how to adapt language when using such gestures.

In this paper, we propose a preliminary investigation of how decisions regarding robot-to-human communication modality in mixed reality environments might be made on the basis of humans' perceptual and cognitive states. Specifically, we propose to use brain data acquired with high-density fNIRS to measure the neural correlates of cognitive and emotional states with particular relevance to adaptive human-robot interaction (HRI). This paper makes four contributions to the research domain: (1) First, we describe several states of interest that fNIRS is well suited to measure and that have direct implications to HRI adaptations; (2) Second, we leverage the framework developed in our previous work to explore how different neurophysiological measures could inform the selection of different communication strategies; (3) Third, we describe results from a feasibility experiment where multilabel/multiclass Convolutional Long Short Term Memory (ConvLSTM) Networks were trained to classify the target mental states of 10 participants; and (4) Fourth, we discuss a research agenda for adaptive human-robot teams based on our findings.

2 NEUROPHYSIOLOGICAL STATES MEASURABLE USING FNIRS

Ideal task performance is dependent on optimizing humans' information processing capabilities, which are affected by the complex interplay between their perceptual processing load [33, 45, 53], their cognitive load [33, 47, 54], and their emotional state [13, 42, 44]. In the following sections we define each of these tenets of human information processing and we describe prior research using fNIRS to measure these states of interest. The non-invasive fNIRS device provides spatially accurate brain activity information like functional magnetic resonance imaging (fMRI) (about 1cm lower than that achieved by fMRI [19, 34]), but it can do so in ecologically valid experimental environments. The device holds great potential for non-invasive brain measurement in naturalistic settings due to its practical nature, ease of set-up, robustness to motion artifacts, and high spatial resolution [7, 24, 27]. The basis of fNIRS is the use of near-infrared light, which can penetrate through scalp and skull to reach the brain cortex. Optical fibers are placed on the surface of the head for illumination while detection fibers measure light which reflects back (Fig. 2), and concentration changes in oxy- and deoxy- hemoglobin can be distinguished [7]. The fNIRS has higher spatial resolution than EEG, making it possible to localize specific functional brain regions of activation, as could be done with the constrictive fMRI device [43].

2.1 Cognitive Load

Ideal task performance depends on optimizing mental workload, which refers to the limited information processing capabilities of the human brain, as demanded by a task [33, 54]. When task demands are too high for the brain's maximum processing capacity, performance decrements and task shedding often occur [5, 33]. 'Workload' is an umbrella term. When we compute arithmetic, compose a poem, or chat with a friend, we engage different cognitive resources to complete the task. It is possible to identify the neural correlates of different types of cognitive load with fNIRS and other brain imaging devices. We identify several constructs from the cognitive science domain below that are highly relevant to the HCI domain. The types of cognitive load our model currently considers are Response Inhibitions (RI), Working Memory (WM), Spatial Attention (SA), Visual Lexical Processing (VLP), Visual Search (VS). Detailed descriptions of these cognitive resources and the tasks used to express these types cognitive load in human participants are detailed in sections 4, 5, as well as expressed visually in figure 1.

2.2 Negative Affect

Emotional state also has a strong impact on information processing capabilities, with Negative Affect (NA) (e.g., frustration, stress) causing dramatic increases in cognitive load [22, 42], which can often be catastrophic to human performance [13, 14, 31, 52]. We will focus on affective states associated with high arousal and low valence, such as stress, frustration, and fear. Although there have been many techniques proposed in the research to define, operationalize, and measure affective states, one of the most common approaches is to describe affect along the two orthogonal dimensions of valence (ranging from unpleasant to pleasant) and arousal (ranging from low to high excitement). We focus on NA in this paper (low valance and moderate to high arousal) as affective states such as frustration, stress, and fear are particularly detrimental to human performance. fNIRS has been used to measure NA [3, 41].

2.3 Perceptual Modality

It has also been shown that people process stimuli from their environment through their five senses of sight, hearing, touch, smell, and taste. At this time we focus on measuring perceptual load on people's visual and auditory resources, as those are the primary modalities employed by HCI and HRI. It is possible to measure the load on a given perceptual resource (auditory or visual) with fNIRS, which can then be used to help multimodal systems to determine which output modality to use in a given scenario [45].

3 USEFULNESS OF MEASURABLE STATES FOR MIXED REALITY COMMUNICATION

Now that we have described mental states shown to be robustly measurable using fNIRS, we can discuss how those states may be useful to inform communication decisions in Mixed-Reality environments.

In previous work, we defined a taxonomy of mixed-reality gestures [58]. In that taxonomy, five primary categories of gesture were defined: *Egocentric* gestures (physical gestures performed within the speaker's perspective); *Allocentric* gestures (augmented gestures picking out the speaker's target referent within the viewer's perspective); *Perspective-Free* gestures (augmented gestures projected onto the environment from a third-party perspective); *Ego-Sensitive Allocentric* gestures (augmented gestures that connect the speaker to its referent within the viewer's perspective); and *Ego-Sensitive Perspective-Free* gestures (augmented gestures that connect the speaker to its referent using a projection onto the environment from a third-party perspective).

In this section, we will consider only egocentric and allocenric gestures. Specifically, we will consider how the mental states discussed in the previous section could inform the following decisions: (1) Should the robot pursue communication with its teammate? (2) Should the robot use a fully descriptive or concise referring form? (3) Should the robot accompany its language with a gesture? (4) Should the robot use an egocentric (physical) or allocentric (augmented) gesture? In this section, we will provide preliminary answers to these questions based on our own intuitions; throughout the rest of the paper we will describe a framework for robustly assessing mental states; in immediate future work, we plan to leverage that framework to experimentally investigate these research questions, using our preliminary intuitive answers as testable predictions in those experiments.

3.1 Should the robot pursue communication with its teammate?

We believe that measurements of WM Load and Negative Affect may be informative in deciding whether or not to communicate with a human in the first place. That is, when interacting with a teammate with high WM Load or NA, it may be advantageous for a robot to postpone communicative actions. Communicating new information to a teammate who already has high WM load may disrupt that teammate's ability to maintain the items currently in WM, and as such the robot risks harming both the teammate's Situational Awareness [15] and task performance. Communicating new information to a teammate who has NA may be regarded as annoying, and further decrease that teammate's affect. Accordingly, a robot may be better off avoiding communication with a high NA team member unless the robot has reason to believe that the information to be communicated will increase teammate affect or that the value of communicating the information is worth the potential drop in teammate affect.

3.2 Should the robot use a fully descriptive or concise referring form?

We believe that measurements of WM Load, RI, and Perceptual Modality may be informative in deciding whether to use a fully descriptive or concise referring form when communicating with human teammates. When communicating with a teammate with high WM Load, is it more advantageous to use a fully descriptive or a concise referring form? The use of a fully descriptive form will decrease the likelihood of the listener being able to maintain the items currently in WM, suggesting that a concise form should be chosen. On the other hand, concise referring forms are often chosen because the speaker (implicitly) believes that their target referent is already activated [21], a status strongly correlated with maintenance in WM [12, 56]. If this is *not* the case, then a concise form will not only be ineffective, but may cause an unnecessary context shift for the listener accompanied by a decrease in ability to maintain the current contents of their WM, resulting in potential harm to the listener's task performance. When communicating with a teammate with low RI, it may be advantageous to use a full referring form. As a full referring form both more clearly distinguishes the target referent and takes more time to process. Using such a form could decrease the likelihood of the listener instinctively and erroneously acting towards an incorrect referent. Finally, the perceptual modality of a teammate's workload may directly inform what type of referring form should be used to communicate with them. That is, if a teammate has high auditory load, a more concise referring form may be preferable to avoid overloading that channel.

3.3 Should the robot accompany its language with a gesture?

We believe that measurements of VS and Perceptual Modality may be informative in deciding what type of gesture to use when communicating with humans. If a robot's teammate is performing a VS task and the robot's target is not relevant to their teammate's VS, language unaccompanied by gesture may be preferable to gesture-accompanied language, so as information can be communicated with minimal disruption of the VS. Finally, the perceptual modality of a teammate's workload may directly inform what mode of communication should be used with them. When communicating with a worker with high auditory load, it may make more sense to communicate information primarily visually (i.e., through gesture); when communicating with a worker with high visual load, it may make more sense to communicate information through spoken language (i.e., unaccompanied by gesture).

3.4 Should the robot use an egocentric (physical) or allocentric (augmented) gesture?

We believe that measurements of RI and VS may be informative in deciding what type of gesture to use when communicating with humans. Because egocentric (physical) gestures may be less legible than allocentric (augmented) gestures, and thus may take more effort

to interpret, using such a form could decrease the likelihood of the listener instinctively and erroneously acting towards an incorrect referent. If a robot's teammate is performing a VS and the robot's target referent is relevant to that task, a gesture that immediately draws the listener's eye towards that target (e.g., an allocentric gesture) may be preferable.

4 MULTILABEL CLASSIFICATION OF PERCEPTUAL MODALITY, COGNITIVE LOAD, AND EMOTIONAL STATE

None of the cognitive and emotional states described above are mutually exclusive. In fact, these states often occur in concert with one another (i.e., a person may feel frustrated while having a high WM load, and while processing information via her visual channel). Researchers have explored the effects of combining two or more simplified tasks in order to view brain activity realistic multitasking scenarios and found that if two or more specific regions of the brain are activated by each task separately, then combining these tasks will cause greater activation in the areas of the brain that were recruited to complete each task alone [37, 38]. Other complementary research has found that when multiple tasks are viewed separately and then combined, an additional area in the prefrontal cortex, implicated in dual tasking, is activated [17, 39]. The task of predicting cognitive load, perceptual processing modality, and emotional state is therefore well suited for multilabel classification, where patterns of brain activity can be associated with multiple labels at once.

We have designed a multiclass/multilabel fNIRS classifier building on recent work by Leon that used multilabel classification of EEG data for motor-imagery based BCI applications [30]. Leon used a multilabel approach that considers the detection of single as well as combined motor imageries, (i.e., two or more body parts used at the same time) to direct a robotic arm. Since imagining the movement of particular limbs causes activation in specific brain regions, a signal processing scheme was developed based on the specific location of the activity sources that are related to each body part. This allowed Leon to utilize the Common Spatial Pattern (CSP) algorithm to the multi-class domain; as CSP has been powerful at discriminating sensorimotor rhythms [17]. Multilabeling was utilized to account for cases of combined motor imageries with an on('1')/off('0') label for each respective body part. Finally, Leon's approach was to break down the multiclass/multilabel classification task into a series of single label classification problems. In our case, we constructed a multiclass/multilabel algorithm by adapting a convolutional long short-term memory (ConvLSTM) network for classifying high('2')/medium('1')/absent('0') levels of (i) cognitive load, (ii) perceptual processing modality, and (iii) affect. As depicted in Fig. 1, the cognitive labels include response inhibition, working memory, spatial attention, visual lexical processing and visual search loads. The perceptual processing modality includes auditory and visual perceptive loads. The affect label is focused solely on negative affect. Each label is assigned one of three possible classes to indicate the level of use of a mental resource during the benchmark tasks described in section 5.1. In the case of high use, a '2' is assigned, medium use is assigned a '1' and no use receives a '0'. Fig. 1 depicts an example of one of our multiclass/multilabel assignments.

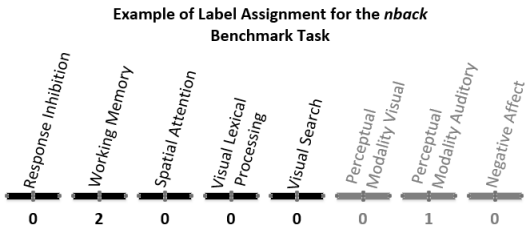

Figure 1: Eight possible labels are included in our multilabel output. The outputs are assigned one of three possible classes to indicate the level of use of the respective cognitive resource ('2': high use; '1': medium use; '0': no use). Black labels represent cognitive load, grey labels are for perceptual modality and negative affect. configurations

The model takes in training data consisting of benchmark tasks from the psychology literature designed to elicit specific types of perceptual load, cognitive load, and emotional states. The trained model can then be used to make predictions in real-time of users' perceptual and cognitive load, as well as their affective state. In the next section we describe a feasibility experiment that was run to test our multiclass/multilabel model.

5 FEASIBILITY EXPERIMENT

fNIRS data was collected on 10 participants using the Hitachi ETG-4000 near-infrared spectroscopy device with a sampling rate of 10Hz. As shown in Fig. 2, a probe design, measuring 52 channel locations, was created to cover the frontal cortex. Once the probe was placed on each participant's head, a Patriot Polhemus 3d digitizer was used to measure the location of each source/detector on that participant's brain. Using the digitizer information, NIRS_SPM was used to identify the Brodmann region that was measured by each of the 52 fNIRS channels.

For all tasks, participants were given on screen instructions and practice trials in order to ensure they understood how to complete the task. During the practice trials they were informed as to whether or not they had entered a correct response. Once they were finished with the practice trials they immediately performed the task, which took place in multiple blocks, containing multiple trials of each task. Between trials, after the participant's response period had ended, a variable inter stimulus interval (ISI), consisting of a cross fixation point, was presented between the trials. The length of the ISI was an exponential distribution (mean= 4s, min=2s, max=8s). The Multi-Attribute Task Battery (MATB) task was administered three different times to participants with the difficulty setting increasing from 'low', to 'medium', to 'high' throughout the experiment.

5.1 Training Data

5.1.1 Load: Working Memory, Task: N-Back. WM refers to an information processing system that temporarily stores information in order to serve higher order cognitive functions such as planning, problem solving, and understanding language [28]. Many studies in experimental psychology are based on Baddeley's model of WM [2, 47], which hypothesizes that there are separate storage spaces for

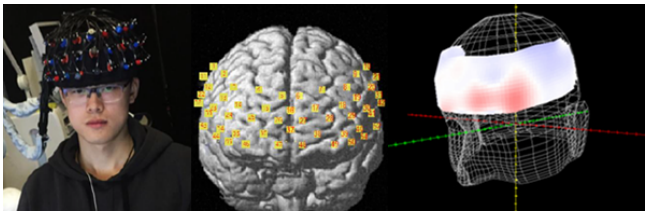

Figure 2: The 52 channel fNIRS configuration used in these experiments.

short term memory. Since task demands can easily exceed humans' capacity for holding items in working memory (e.g., Miller's magic number of 7 ± 2 [37] or Cowan's 4 [12]. The neural correlates of working memory have been measured in the prefrontal cortex, which is accessible with the fNIRS device [24, 32, 33, 47].

The N-Back (blocks=4, trials=20 n-value=1-4) task is designed to cause cognitive load on people's WM resources by requiring participants to hold a stream of characters in their mind and to respond when a new character presented to them matches one of the characters they are currently holding. Our N-Back task, based on Harvey et al. [23], presented participants with a series of letters, a single letter at a time, for a duration of 500ms each.

5.1.2 Load: Visual Search, Task: Visual Search. VS load involves visually searching for items within a set of distractor items. There are two kinds of VS; efficient and inefficient VS. If a target item is saliently different as compared to the distractor items surrounding the target item it can be found immediately, regardless of the number of distractor items [1, 9, 39]. Inefficient search occurs when the search item is not highly salient as compared to the distractor items. Although efficient VS occurs with minimal cognitive processing, inefficient visual search is believed to recruit functional brain regions in the prefrontal cortex, which controls selective attention, and to recall, with WM, what areas of the set of items have already been searched [1, 39]. fNIRS has been used to measure the neural correlates of VS load in the prefrontal cortex [9, 24, 33].

The VS task is was modeled after the task design developed by Wang, Cavanagh, and Green [51]. A circular array of nine letters consisting of a distractor (backwards Ns) and a target (normal facing Ns) was displayed to the participant for a variable amount of time (mean=950ms, max=1250ms, min=650ms). The participant's task was to determine as to whether or not the target was displayed within the array.

5.1.3 Visuo-Spatial Attention, Posner Curing Paradigm. VSA is a form of visual attention that involves directing attention to a location in one's current visual field without any accompanying eye or body movements. This form of visual attention has been theorized to be of importance in the search for salient stimuli within a goal orientated environment and involves correctly selecting visual stimuli that are relevant to the successful completion of the goal. [4]. Similar to other forms of attentional modulation, the parietal and frontal cortices have been implicated in the shifting of VSA from one stimuli to another [11]. Past evidence indicates that fNIRS is a suitable device for measuring VSA in the dorsal lateral prefrontal cortex [25, 40].

The Posner task design was based on Thiel, Zilles, and Fink's work in designing a posner cuing task that worked within an fMRI environment [50]. The task consisted of a diamond-shaped cue surrounded by two empty boxes to the left and right of the cue that appeared in the center of the screen. After a period of 100ms, one side of the diamond would then be highlighted for a period of 300ms, then a black circle would appear in one of the boxes to either the left or the right of the diamond-shaped cue. The participant's task was to indicate to which side of the cue to dot appeared.

5.1.4 Load: Response Inhibition, Task: Go No-Go. RI deals with the brain's suppression of automatic, but incorrect responses. An example of RI is seen when video gamers who want to make their characters jump on screen overcome the natural urge to thrust the game controller into the air to mimic jumping instead of pressing 'A' on the controller for the desired effect. Systems such as the Ninentdo Wii reduce the load on RI by enabling users to make characters jump by simply thrusting the Wii controller upward. HCI designers aim to create systems that allow users to interact naturally with their task environment, limiting the load on their RI resources [16, 24]. The neural correlates of RI have been repeatedly found in the anterior cingulate cortex [29, 46, 47], and several studies have demonstrated the utility of fNIRS for measuring patterns of activation in this region [24, 46].

A go no-go task was used to measure RI load. The development of stimulus materials was guided by Huettel, Mack, and McCarthy [26]. The participant was tasked with responding to a target stimulus and not responding to a distrator stimulus.

5.2 Stimulus Materials: Test Data

5.2.1 Triage Analysis Task. The Triage Analyst task acts as an ecologically valid representation of a cyber-security network analyst's position, and is based on the work of Greenlee et al [20]. The task involved the participant viewing an empty table in the center of the screen. The table headings were 'Source IP', 'Source Port', 'Destination IP', 'Destination Port'. Participants were informed prior to the task beginning that they did not require working knowledge of the terminology involved in order to complete the task. The table would populate with 'transmissions' on the 'network' the participant was monitoring. Starting from the top of the table, new 'transmissions' would fill the table until a maximum of five 'transmissions' were on the screen. A maximum of five 'transmissions' were present at one time. The participant was tasked with determining when either two different 'transmissions' on the table had the same destination information (both 'Destination IP' and 'Destination Port), or two different 'transmissions' on the table having the same source information.

Source IP	Source Port	Destination IP	Destination Port
103.17.22.62	82	198.176.21.9	14
56.254.13.15	11	198.176.21.9	14
226.12.22.132	63	108.71.226.62	77
203.11.46.26	89	251.102.18.3	65
42.113.56.5	44	56.225.11.89	43

Figure 3: Destination Intrusion

5.2.2 MATB. Our task involved a complex multi-tasking scenario using a variation of the Multi-Attribute Task Battery (MATB) [10, 38]. We used the Air Force's updated version of the Multi-Attribute Task Battery (AF_MATB) [38], and chose a difficulty level that required a good deal of mental effort and multi-tasking. With the difficulty of the task and the high level of multi-tasking required, the task was difficult to complete perfectly. Pilot tests showed that all subjects had to remain engaged during the entire task to receive an adequate performance score. The AF_MATB consists of six windows which provide information about four different subtasks (Fig. 4): System Monitoring, Communications, Resource Management, and Tracking. The last two windows, which contain Scheduling and Pump Status information, are resources that the user can use to improve performance during the task.

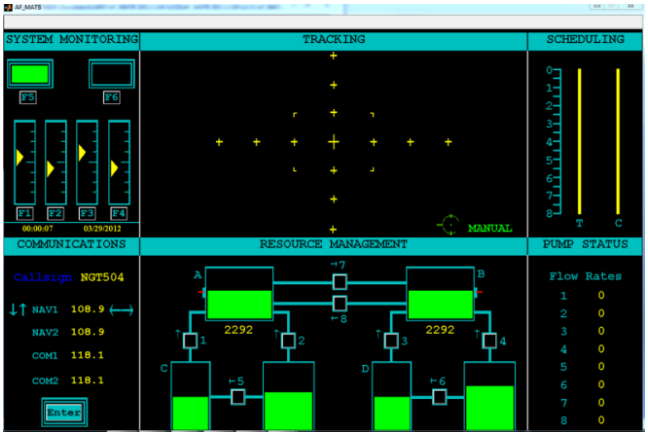

Figure 4: The Multi-Attribute Task Battery

6 PRELIMINARY RESULTS

Using fNIRS data gathered during the five previously described benchmark tasks, a classification model illustrated in Fig. 6 was trained for ten participants. Training was based on a Keras supported Python code that reshaped each time-instance oxy/deoxy channel readings into a 3D frame with the same spatial configuration as the experimental probe set-up. In order to adhere to the temporal nature of the data, frames were grouped into sequences that corresponded to the timesteps of the benchmark tasks. All groups of frames had eight mental resource multilabeled targets, each with a class value (2,1, or 0) representative of the level of use of the resource. All class values were set to 0 for the eight labels assigned to the reaction time task reference data. Training data was input into a ConvLSTM2D algorithm followed by two dense layers with relu activation and a dropout layer in between. Training was supported with adam optimization, categorical crossentropy loss and final output softmax activation. Our model was fit to the input data and respective labels using a batch size of 32 and 20 epochs. Each instance (i.e. row) of testing data was also reshaped and grouped together with the same configuration used for training data shown in Fig. 6. Testing was performed on triage data which resulted in eight output vector predictions, each with three probabilities representing the three cognitive resource levels (2, 1, or 0). Final label values

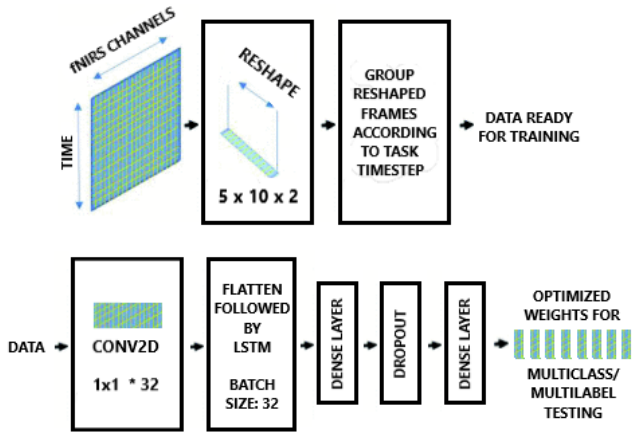

Figure 5: Structure of model used to obtain initial results in determining neurophysiological state levels of participants during our experimental studies

were based on the levels with the highest probabilities. Testing was also repeated for the medium-difficulty MATB experimental results of ten participants. Average label values were calculated generating values of 0.6 (s.d.= 0.92) and 0.1 (s.d.= 0.3) for the triage and MATB medium RI levels, respectively. Values for WM were 0.8 (s.d.= 0.98) for triage and 0.4 (s.d.= 0.8) for MATB medium. The average label value for VSA was lower for triage, 0.3 (s.d.= 0.64) than for MATB medium with a value of 0.6 (s.d.= 0.92). The Auditory Perceptual Modality had values of 0.1 (s.d.= 0.3) and 0.2 (s.d.= 0.6) whereas Visual Perceptual Modality ended up with values at 1.1 (s.d.= 0.3) and 1 (s.d.= 0.45) for triage and MATB medium, respectively. Finally, NA had a value of 0.1 (s.d.= 0.3) for triage and was non-existent for MATB medium. These values are presented in figure 6.

6.1 Perceptual Modality

As seen in Figure 6, the average values of class assignments for each mental resource for the ten participants that were tested can offer insight into the experienced workload for the ecologically valid triage analyst and MATB tasks. The predicted perceptual modality for both tasks is predominantly visual with a predicted output label of approximately 1, indicating a moderate amount of perceptual load. Also of note in the perceptual modality predictions generated by the model was that the average prediction for auditory perceptual load for the MATB task was two times as high as the rate as it was in the triage task. This result may be due to the communications task component of the MATB, which requires users to listen in for 'air traffic' messages periodically in order to successfully complete the task. However, more research must be completed in order to confirm and further explore this result.

6.2 Negative Affect

Our current model was not able to predict anything elucidating with respect to NA. The result of this may be that the particular tasks that were used both in the training or testing set were unable to elicit

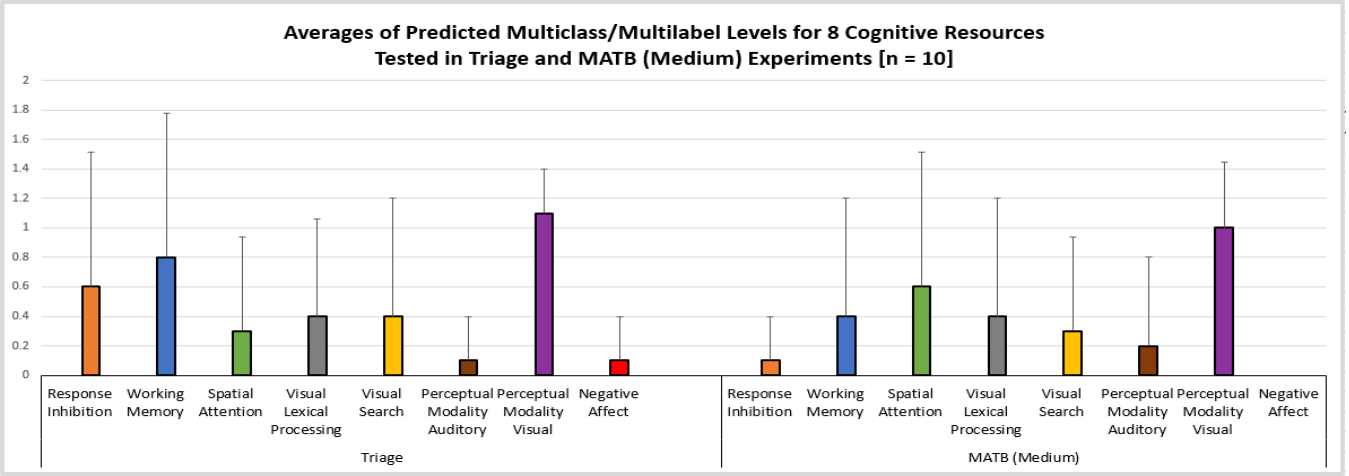

Figure 6: Triage and MATB (Medium) Preliminary Results

the level of frustration needed to detect a physiological response. Another reason may be that differences in cognitive ability between participants may have caused some participants to feel frustrated with a task, whereas other participants felt they could complete the task without any negative emotional interference. Though the current predictions leave much to explore when trying to predict NA, future work could perhaps see more accurate results with respect to NA by labeling NA in a manner consistent with a participant's self reported level of frustration, rather than relying on a value label based purely on the type of task.

6.3 Cognitive Load

The differences in the predictions of the types cognitive load that our model predicted participants were undergoing during ecologically valid tasks offer insight into what cognitive resources are utilized by a particular task. Our model predicted that the triage analyst task saw the heaviest usage of both WM as well as RI. These findings are consistent with the nature of the task, detailed in section 5.2.1, which requires both the use of RI on the part of the participant to not report an 'intrusion' on the network even when there may be salient information that might lead them to suspect there could be, as well as keep various pieces of information within WM while deciding whether or not an 'intrusion' has taken place. The predicted cognitive resource most heavily utilized during the MATB task (section 5.2.2), however, was SA. The utilization of this resource may be due in part to the user having to monitor multiple tasks simultaneously across the screen. Also of interest is the low predicted value of RI load on the participants during the MATB, being that the task requires the participant to monitor multiple gauges and to respond as quickly as possible when any on of the gauges begins to shift into an unsatisfactory range of values. Overall, the differences in the predicted values generated by the model give insight into the different cognitive resources each task requires in order to complete, and would be able to provide information to a robot agent about what types of cognitive load the human agent is most likely experiencing during a given task. A larger dataset with more training data, with a wider range of cognitive benchmark tasks that have been shown require the same cognitive resources to complete and more ecologically valid testing tasks could improve a multi-label classification system's ability to accurately predict both type and magnitude of cognitive load resources used in a given task.

7 RESEARCH AGENDA

We believe that the multilabel LSTM approach described in this paper has potential for the greater fNIRS research community. There is a need to begin combining fNIRS datasets for machine learning that provides enough training data to build robust classifiers that do not overfit to any one individual, task, or sensor configuration. The multilabel approach described here is a first step toward defining a labeling schema that can be used to as a common labeling technique for labeling a large range of experimental tasks from different research labs. The eight labels depicted in Fig. 1 are our first attempt at a labeling schema. Future work should iterate upon this to ensure the multilabels are robust enough to capture human perceptual and cognitive processing across a range of ecologically valid task environments. Our labeling schema was designed with Wickens' multiple resource theory and task demand vector work in mind [53]. Future work in this field should consider the Wickens' theoretical work and other work on human information processing to develop a multilabel schema that is robust and theoretically sound. With that labeling schema determined, empirical research of the future can help to refine the multilabeling schema. Ultimately, large datasets can be combined with enough multilabeled training data to achieve robust models to classify a range of perceptive, cognitive, and emotional states that can be used to inform human-robot interactions of the future.

With an agreed upon multilabeling schema for labeling fNIRS and other cognitive training data, empirical studies can be run to explore the effects of different mixed reality communications on humans, in order to answer the research questions laid out in Section. 3. An important step for the HRI research community will then be to develop and test adaptive systems that modify the interaction content and modality of autonomous agents based on the current mental state of the human teammate.

REFERENCES

[1] EJ Anderson, SK Mannan, M Husain, G Rees, Petroc Sumner, et al. Involvement of prefrontal cortex in visual search. *Experimental Brain Research*, 2007.

[2] Alan David Baddeley and Sergio Della Sala. Working memory and executive control. *Phil. Trans. R. Soc. Lond. B*, 351(1346):1397–1404, 1996.

[3] Danushka Bandara, Senem Velipasalar, Sarah Bratt, and Leanne Hirshfield. Building predictive models of emotion with functional near-infrared spectroscopy. *International Journal of Human-Computer Studies*, 110:75–85, 2018.

[4] Paolo Bartolomeo, Michel Thiebaut De Schotten, and Ana B Chica. Brain networks of visuospatial attention and their disruption in visual neglect. *Frontiers in human neuroscience*, 6:110, 2012.

[5] James P Bliss, John W Harden, and H Charles Dischinger Jr. Task shedding and control performance as a function of perceived automation reliability and time pressure. In *Proc. Human Factors and Ergonomics Society Annual Meeting*, 2013.

[6] Tathagata Chakraborti, Sarath Sreedharan, Anagha Kulkarni, and Subbarao Kambhampati. Alternative modes of interaction in proximal human-in-the-loop operation of robots. *arXiv preprint arXiv:1703.08930*, 2017.

[7] B Chance, Z Zhuang, Chu UnAh, C Alter, and L Lipton. Cognition-activated low-frequency modulation of light absorption in human brain. *PNAS*, 1993.

[8] John K Chapin, Karen A Moxon, Ronald S Markowitz, and Miguel AL Nicolelis. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature neuroscience*, 2(7):664, 1999.

[9] Willy NJM Colier, Valentina Quaresima, Rüdiger Wenzel, Marco C van der Sluijs, Berend Oeseburg, Marco Ferrari, and Arno Villringer. Simultaneous near-infrared spectroscopy monitoring of left and right occipital areas reveals contra-lateral hemodynamic changes upon hemi-field paradigm. *Vision research*, 41(1), 2001.

[10] J R Comstock Jr and Ruth Arnegard. The multi-attribute task battery for human operator workload and strategic behavior research. Technical report, NASA, 1992.

[11] Jennifer T Coull. Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. *Progress in neurobiology*, 55(4):343–361, 1998.

[12] Nelson Cowan. *Attention and Memory: An Integrated Framework*. OUP, 1998.

[13] Sidney D'Mello and Rafael A Calvo. Beyond the basic emotions: what should affective computing compute? In *CHI'13 Extended Abstracts*, 2013.

[14] Sidney D'Mello and Art Graesser. Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2):145–157, 2012.

[15] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 1995.

[16] MA Finch, James G Phillips, and James W Meehan. Cursor type and response conflict in graphical user interfaces. *Interacting with Computers*, 19(1), 2006.

[17] Riccardo Fusaroli, Joanna Rączaszek-Leonardi, and Kristian Tylén. Dialog as interpersonal synergy. *New Ideas in Psychology*, 32:147–157, 2014.

[18] Ferran Galán, Marnix Nuttin, Eileen Lew, Pierre Ferrez, Gerolf Vanacker, et al. A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. *Clinical neurophysiology*, 2008.

[19] Gabriele Gratton and Monica Fabiani. Fast optical signals: Principles, methods, and experimental results. *In vivo optical imaging of brain function*, 2009.

[20] Eric Greenlee, Gregory Funke, Joel Warm, Ben Sawyer, Victor Finomore, Vince Mancuso, et al. Stress and workload profiles of network analysis: Not all tasks are created equal. In *Adv. Hum. Fac. Cybersec.* 2016.

[21] Jeanette K Gundel, Nancy Hedberg, and Ron Zacharski. Cognitive status and the form of referring expressions in discourse. *Language*, 1993.

[22] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psych.* 1988.

[23] Philippe-Olivier Harvey, Philippe Fossati, Jean-Baptiste Pochon, Richard Levy, Guillaume LeBastard, et al. Cognitive control and brain resources in major depression: an fMRI study using the n-back task. *Neuroimage*, 2005.

[24] Leanne M Hirshfield, Rebecca Gulotta, Stuart Hirshfield, Sam Hincks, Matthew Russell, Rachel Ward, et al. This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy. In *CHI*, 2011.

[25] Jing Huang, Fang Wang, Yulong Ding, Haijing Niu, Fenghua Tian, Hanli Liu, and Yan Song. Predicting N2pc from anticipatory hbo activity during sustained visuospatial attention: a concurrent fNIRS–ERP study. *Neuroimage*, 113, 2015.

[26] Scott Huettel, Peter Mack, and Gregory McCarthy. Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nat. Neuro.*, 2002.

[27] Kurtulus Izzetoglu, Scott Bunce, Banu Onaral, Kambiz Pourrezaei, and Britton Chance. Functional optical brain imaging using near-infrared during cognitive tasks. *International Journal of human-computer interaction*, 17(2):211–227, 2004.

[28] Susanne M Jaeggi, Ria Seewer, Arto C Nirkko, Doris Eckstein, Gerhard Schroth, Rudolf Groner, and Klemens Gutbrod. Does excessive memory load attenuate activation in the prefrontal cortex? load-dependent processing in single and dual tasks: functional magnetic resonance imaging study. *NeuroImage*, 2003.

[29] Yves Joanette. Neuroimaging investigation of executive functions: evidence from fNIRS. *Psico*, 39(3):267–274, 2008.

[30] Cecilia Lindig León. *Multilabel classification of EEG-based combined motor imageries implemented for the 3D control of a robotic arm*. PhD thesis, Université de Lorraine, 2017.

[31] Regan Mandryk, M Atkins, and Kori Inkpen. A continuous and objective evaluation of emotional experience with interactive play environments. In *CHI*, 2006.

[32] Ryan McKendrick, Hasan Ayaz, Ryan Olmstead, and Raja Parasuraman. Enhancing dual-task performance with verbal and spatial working memory training: continuous monitoring of cerebral hemodynamics with nirs. *Neuroimage*, 2014.

[33] Ryan McKendrick, Raja Parasuraman, Rabia Murtza, Alice Formwalt, Wendy Baccus, Martin Paczynski, and Hasan Ayaz. Into the wild: neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. *Front. Hum. Neuro.*, 2016.

[34] Andrei V Medvedev. Shedding near-infrared light on brain networks. *Journal of Radiology and Radiation Therapy*, 2013.

[35] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *Transactions on Information and Systems*, 77(12):1321–1329, 1994.

[36] J R Millán, Frederic Renkens, J Mourino, and W Gerstner. Noninvasive brain-actuated control of a mobile robot by human eeg. *Trans. Biomed. Eng.*, 2004.

[37] George A Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 1956.

[38] William D Miller Jr. The US air force-developed adaptation of the multi-attribute task battery for the assessment of human operator workload and strategic behavior. Technical report, Consortium Research and Fellows Program, 2010.

[39] Gisela Müller-Plath. Localizing subprocesses of visual search by correlating local brain activation in fmri with response time model parameters. *Journal of neuroscience methods*, 171(2):316–330, 2008.

[40] Takayuki Nakahachi, Ryouhei Ishii, Masao Iwase, Leonides Canuet, Hidetoshi Takahashi, Ryu Kurimoto, Koji Ikezawa, Michiyo Azechi, Osami Kajimoto, and Masatoshi Takeda. Frontal cortex activation associated with speeded processing of visuospatial working memory revealed by multichannel near-infrared spectroscopy during advanced trail making test performance. *Behavioural brain research*, 215 (1):21–27, 2010.

[41] Shota Nishitani, Kazuyuki Shinohara, et al. Nirs as a tool for assaying emotional function in the prefrontal cortex. *Frontiers in human neuroscience*, 7:770, 2013.

[42] Raja Parasuraman and D Caggiano. Neural and genetic assays of human mental workload. *Quantifying human information processing*, pages 123–149, 2005.

[43] Raja Parasuraman and Matthew Rizzo. *Neuroergonomics: The brain at work*. Oxford University Press, 2008.

[44] Rosalind Wright Picard. *Affective computing*. MIT Press, 1997.

[45] Felix Putze, Sebastian Hesslinger, Chun-Yu Tse, YunYing Huang, Christian Herff, Cuntai Guan, and Tanja Schultz. Hybrid fNIRS-EEG based classification of auditory and visual perception processes. *Frontiers in neuroscience*, 8:373, 2014.

[46] Matthias L Schroeter, Stefan Zysset, Thomas Kupka, Frithjof Kruggel, and D Yves Von Cramon. Near-infrared spectroscopy can detect brain activity during a color–word matching stroop task in an event-related design. *Human Brain Map.*, 2002.

[47] Edward E Smith and John Jonides. Storage and executive processes in the frontal lobes. *Science*, 283(5408):1657–1661, 1999.

[48] Marie-Hélène Stoltz, Vaggelis Giannikas, Duncan McFarlane, James Strachan, Jumyung Um, and Rengarajan Srinivasan. Augmented reality in warehouse operations: Opportunities and barriers. *IFAC-PapersOnLine*, 2017.

[49] Daniel Szafir and Bilge Mutlu. Pay attention!: designing adaptive agents that monitor and improve user engagement. In *CHI*, 2012.

[50] Christiane M Thiel, Karl Zilles, and Gereon R Fink. Cerebral correlates of alerting, orienting and reorienting of visuospatial attention: an event-related fmri study. *Neuroimage*, 21(1):318–328, 2004.

[51] Qinqin Wang, Patrick Cavanagh, and Marc Green. Familiarity and pop-out in visual search. *Perception & psychophysics*, 56(5):495–500, 1994.

[52] Brian E Weeks. Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, 65(4):699–719, 2015.

[53] Christopher D Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.

[54] Christopher D Wickens. Multiple resources and mental workload. *Human factors*, 50(3):449–455, 2008.

[55] Tom Williams. A framework for robot-generated mixed-reality deixis. In *Proc.1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI*, 2018.

[56] Tom Williams and Matthias Scheutz. Reference in robotics: A givenness hierarchy theoretic approach. In *The Oxford Handbook of Reference*. 2018 (in press).

[57] Tom Williams, Daniel Szafir, Tathagata Chakraborti, and Heni Ben Amor. Virtual, augmented, and mixed reality for human-robot interaction. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018.

[58] Tom Williams, Nhan Tran, Josh Rands, and Neil T. Dantam. Augmented, mixed, and virtual reality enabling of robot deixis. In *VAMR*, 2018.