Brainput: Enhancing Interactive Systems with Streaming fNIRS Brain Input

Erin Treacy Solovey[13], Paul Schermerhorn[2], Matthias Scheutz[3], Angelo Sassaroli[3], Sergio Fantini[3], Robert J.K. Jacob[3]

[1]MIT
Cambridge, MA 02139, USA

[2]Indiana University
Bloomington, IN, 47406, USA

[3]Tufts University
Medford, MA, 02155, USA

erinsol@mit.edu,pscherme@indiana.edu,{matthias.scheutz,angelo.sassaroli,sergio.fantini,robert.jacob}@tufts.edu

ABSTRACT

This paper describes the *Brainput* system, which learns to identify brain activity patterns occurring during multitasking. It provides a continuous, supplemental input stream to an interactive human-robot system, which uses this information to modify its behavior to better support multitasking. This paper demonstrates that we can use non-invasive methods to detect signals coming from the brain that users naturally and effortlessly generate while using a computer system. If used with care, this additional information can lead to systems that respond appropriately to changes in the user's state. Our experimental study shows that *Brainput* significantly improves several performance metrics, as well as the subjective NASA-Task Load Index scores in a dual-task human-robot activity.

Author Keywords

fNIRS; near-infrared spectroscopy; multitasking; brain computer interface; human-robot interaction

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

When communicating with other people, we do not simply use words, but also accompanying visual and auditory cues that give the other person additional insight to our thoughts. At the same time, several physiological changes occur that may or may not be detected by the other person. When we communicate with computers, we also generate these additional signals, but the computer cannot sense such signals, and therefore completely ignores them. Detecting these signals and incorporating them into the user interface could improve the communication channel between the computer and the human with little additional effort required of the user. This communication improvement would lead to technology that is more supportive of the user's changing cognitive state. Such improvements in bandwidth are increasingly valuable, as technology has become more powerful and pervasive, while our cognitive abilities have not improved significantly.

In particular, users increasingly are faced with numerous simultaneous demands and information overload. For example, jobs such as air traffic inherently require multitasking, despite human limitations in this ability. Moreover, the rapidly evolving field of human-robot interaction (HRI) has begun to require complex user interfaces, especially in contexts where human operators are working in teams with one or more autonomous robots.

When the user is unable to handle multiple simultaneous demands, we observe high stress and performance degradation. To determine the proper way to support the user, it is critical to understand the different types of multitasking that may occur, as they do not all affect the user in the same way. Recognizing signals generated naturally by the user that differentiate different types of multitasking could lead to higher productivity, better task performance, and improved experience when the signals are utilized to make the system responsive to users' needs.

Ideally, any sensing of the user's changing cognitive state would be done automatically, in real time, with little inconvenience to the user. Some researchers have approached this problem by monitoring performance data or interaction history (e.g. keystrokes) to assess the user's current state, while others use computer vision to detect facial expressions or other behavioral measures. Physiological measures are also emerging as continuous indicators of cognitive state changes [5, 17, 18]. Brain imaging and brain sensing techniques aim to get close to the source by looking at changes in brain activity during task performance [7, 10]. Currently, electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) are the primary brain sensing methods that have potential for use in realistic HCI settings [28] due to low cost, easy setup, and portability.

There have been early demonstrations that fNIRS signals could be sent in realtime from hardware to a user interface [3, 6]. Using fNIRS in a non-interactive situation, Solovey et al. [29] presented background experiments that showed that three specific cognitive multitasking states could be distinguished from each other. These results suggested that an interactive system that is aware of the user's changing

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Figure 1. *Brainput* provides a passive, implicit input channel to interactive systems, with little effort from the user.

cognitive state during multitasking would be possible, and parts of a feasibility demonstration were described.

Here we present the full working system and an experimental evaluation of its efficacy. We demonstrate that we can use non-invasive methods to detect signals coming from the brain that users naturally and effortlessly produce while using a computer system. *Brainput* learns to identify brain activity patterns occurring during multitasking. It then provides a continuous, supplemental input stream to an interactive human-robot system, which uses this information to modify its behavior to better support multitasking.

The contributions of this paper are as follows:

1) A description and demonstration of a working system that uses fNIRS as a passive, implicit input channel to an interactive system;

2) An evaluation experiment in which performance data shows improved performance using this input modality in a human-robot system;

3) Evidence from subjective questionnaires showing that this input modality improved the users' perceived workload and experience in the human-robot system.

BACKGROUND AND RELATED WORK

Passive, Implicit Input Channel

Brainput was designed to be a passive, implicit and supplementary input channel utilizing the multitasking state classification from the fNIRS brain signal. Other physiological sensors have been similarly explored to identify aspects of the user's state that can be useful in the context of the user interface. Of particular relevance are those that use the physiological data implicitly to augment other aspects of the system. In a gaming context, Nacke et al. found that indirect use of physiological signals was best for adapting dramatic effects such as environment variables [18]. Physiological signals have also been used to automatically pause and bookmark an audio stream during an interruption [21]. These examples demonstrate some of the potential of passive, implicit, and supplemental input in various contexts.

Until recently, most brain-computer interfaces were designed for disabled users, and employed brain signals as the primary input [2, 13, 23, 25, 32]. While these systems provide this group of users with a valuable communication channel, they likely will not see wider adoption due to the low bandwidth compared to other available methods for non-disabled users. With lower costs for non-invasive brain sensing, we recently have seen a growing interest in employing brain sensors for a wider audience (for an overview, see [12]). Much of this work has also used brain sensing as explicit input to the system to make selections or control the interface, (e.g. in a game context [15, 19] or with a multitouch table [33]), although there have been examples of passive brain sensing to be used either as implicit input or for evaluation of user interfaces [7, 10, 11, 16]. Girouard, Solovey and Jacob [6] brought offline analysis of fNIRS signals into a real-time system with the goal of using it to build passive brain-computer interfaces. *Brainput* goes beyond this work by improving the processing, training and classification algorithms, and building and evaluating a viable new input technique to improve the user performance and experience.

Anterior Prefrontal Cortex and Multitasking

While fNIRS has been applied to various locations on the head, the most successful placement is on the forehead (Figure 1) because of the absence of hair, which absorbs light and degrades the fNIRS signal. Thus, the anterior prefrontal cortex, which lies behind the forehead, is the main target for fNIRS brain sensing in this paper. This area of the brain is responsible for many high level processes and many activities can activate the area. Here, we look specifically at detecting brain activity changes during multitasking as we would like to improve user performance and experience in such difficult situations.

Koechlin et al. [14] described three distinct but related multitasking states and these form the basis of the work described in this paper:

1) *Branching* occurs when the user must "hold in mind goals while exploring and processing secondary goals" [14]. This is illustrated by the following scenario: *A user is tackling a complex programming task but is interrupted by an incoming email from her boss that is time sensitive*. Branching processes are triggered frequently in multitasking environments and pose a challenge to users. Automatically sensing this state is the focus in this paper.

2) *Delay Task* occurs when secondary task is ignored and therefore requires little attentional resources. For example, imagine that *a user is tackling a complex programming assignment and at the same time getting instant messages that the user notices, but ignores*. In this case, the secondary task does not require an attentional shift, but instead simply delays the primary task.

3) *Dual Task* entails frequent task switching without the need to maintain information about the previous task (e.g. *A user is monitoring and responding to high priority software support issues that are logged by clients as well as responding to important emails, and regularly switches between the two tasks*.)

Using fMRI for brain imaging, Koechlin et al. demonstrated that these three multitasking activities had different signatures in the anterior prefrontal cortex [14], the area that is best for measuring with fNIRS.

Solovey et al. later showed that these states could be distinguished using fNIRS as well [29]. Figure 2 shows the mean and standard error of the fNIRS signal in *branching* (Blue), *delay* (Red) and *dual task* (Green) from that experiment. The mean is across ten trials of each multitasking activity for each of twelve subjects. The figures in top row show the pattern for oxygenated hemoglobin (Oxy-Hb) and the bottom row shows the deoxygenated hemoglobin (Deoxy-Hb). These are the two measures that we get with fNIRS. The figures on the left are from the sensor on the left side of the head and the figures on the right are from the right side of the head. In addition, [29] showed that these cognitive states may be generic processes that occur in more than one domain. These results have direct implications for HCI as *branching* states would be prime candidates for triggering the support of an adaptive user interface. The analysis in [29] was done offline, after all of the data was collected. However, it shows promise that these signals could be differentiated in realtime.

This paper builds on these results by using the known multitasking activities described by Koechlin et al. [14] and [29] as stimuli for creating individual sets of fNIRS training data during multitasking for each user. This training data is used to build a classification model for each individual that is used to later distinguish between multitasking states the user is experiencing during other tasks and activities. For a full description of the stimuli, please refer to [14] and [29], as we have identical stimuli which should stimulate *branching, dual task,* and *delay task* states in the user.

Brain Sensing for Human-Robot Interaction

Brain-computer interfaces (BCIs) have previously been incorporated into robot architectures, although these have typically been EEG-based systems (e.g., for controlling mobile robots [1, 4] or an intelligent wheelchair [22], among others). fNIRS provides advantages over the more prevalent EEG due to its easy setup and robustness to noise. Past research has proposed the use of fNIRS-based BCIs [20, 31]. However, these projects focus on using brain data for direct one-way control of robot movement, and are less concerned with the interaction between the human and robot. In addition, the reliability of such active control schemes will vary greatly depending on context, and may be particularly difficult to apply in high-stress or high-load contexts. By using fNIRS to passively identify an operator's cognitive state, we can exploit that very phenomenon to improve interaction efficiency.

Augmenting active communication channels is particularly important in HRI, because people have a tendency to ascribe human-like abilities of comprehension to autonomous robots (possibly due to their apparent agency, or to their depiction in popular culture) that are, at this point, unrealistic. Hence, any additional information that can help the robot to understand the operator's intentions will be of great value.

BRAINPUT

The *Brainput* system is made up of several components.

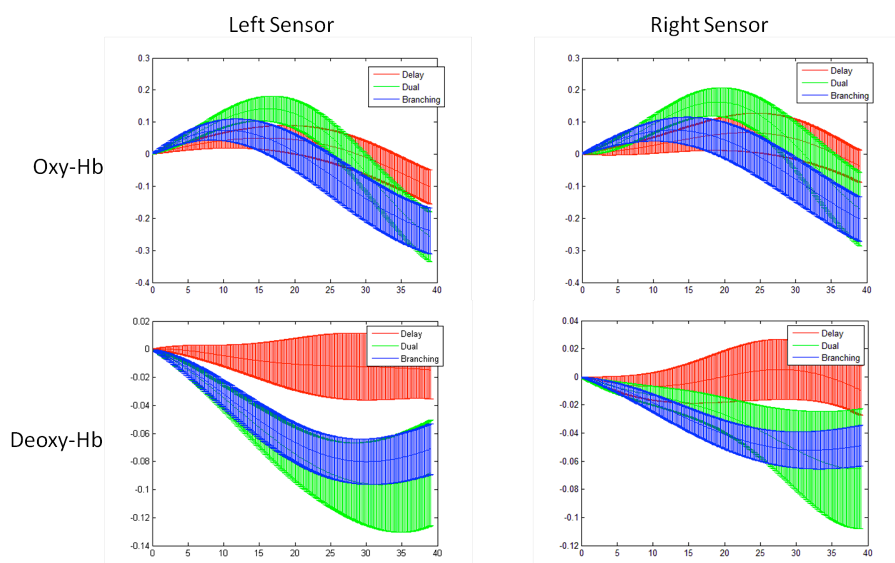

Figure 2. Mean and standard error of fNIRS signal during 40-second trial in *branching* (Blue), *delay* (Red) and *dual task* (Green) from experiment reported in [29]. Mean is across ten trials of each multitasking activity for 12 subjects.

First, the fNIRS data is processed by Boxy software from ISS, Inc. Two components of the Online Fnirs Analysis and Classification System (OFAC) [6] are used to get data from Boxy into a Matlab as well as markers indicating class labels from Presentation software. A new engine for preprocessing and training a machine learning model was built using Matlab and Weka [8] which sends classification results to the robot system continuously via sockets.

EXPERIMENT

To evaluate the effectiveness of using *Brainput* in an adaptive system, we created three adaptation schemes for comparison that are triggered by the brain input stream. Participants completed a robot navigation task three times, each employing a different adaptive behavior as a response to the brain input stream.

Participants

This study included eleven participants (three male), between the ages of 18 and 22 (mean 20.7). All participants were right-handed, had no history of brain injury and had normal or corrected-to-normal vision. Informed consent was obtained for all participants. This experiment was approved by our institutional review board.

Equipment

We used a multichannel frequency domain OxiplexTS from ISS Inc. (Champaign, IL) for fNIRS data acquisition. Two probes were placed on the forehead to measure the two hemispheres of the anterior prefrontal cortex (Figure 1). The source-detector distances were 1.5, 2, 2.5, and 3cm. Each distance measures a different depth in the cortex. Each source emits two near-infrared wavelengths (690 nm and 830 nm) to detect and differentiate between oxygenated and deoxygenated hemoglobin. The sampling rate was 6.25 Hz.

Training Phase

Before working with the robot, each participant completed a training phase to gather fNIRS data in known multitasking exercises. This data was used to build a machine learning model for classifying these cognitive multitasking states.

We used the original multitasking exercises described by Koechlin et al. [14] as simple calibration tasks to elicit *delay, dual task* and *branching* states. The exercises involved processing rules based on letters appearing on the screen. Each stimulus was either an uppercase or lowercase letter from the word "tablet." The expected response from the user was different depending on the case of the letter, so switching between uppercase and lowercase letters would be similar to balancing two tasks. The rules for responding to the stimuli were designed to trigger *branching, dual task,* and *delay*. These are described in detail in [14]. By using known activities that induce known cognitive states, we can train a machine learning classifier to recognize these states later, in new contexts.

Each run of the known multitasking activities lasted approximately forty seconds, and the fNIRS brain signal over this 40-second window became the training example. Once a machine learning model was built from this training data, the *Brainput* system continuously classified 40-second slices of fNIRS data, in a sliding window that moved with each new sample. The sampling rate for the fNIRS system was 6.25 Hz. For more details, see [30].

Experimental Tasks

The main task for the study is a multi-robot version of the task introduced in [27]. Participants remotely supervised two robots (the blue robot and the red robot) that were exploring different areas of a virtual environment. Participants were told that the two robots had collected information that needed to be transmitted back to the control center. The robots could help the participant search for an appropriate transmission location by measuring and reporting the signal strength in its current position. Transmissions were only possible in locations with signal strength of at least 2400 (values ranged from 1300 to 2500, and the single target region in each robot's area covered roughly 1.25 % of the environment).The user had a console to view the environment from each robot's point of view (Figure 3) and could issue commands to the robots such as "go straight," or "turn right" (Figure 4) and the appropriate robot would follow the commands. They could also ask the robot for the signal strength of the current location by clicking "take a reading" and the robot would report the current signal strength. This required the robot to stop and also consumed resources so the robot could not be measuring signal strength at all times. In addition to the command interface, the red robot received cognitive load estimates from the fNIRS system.

Participants were told that the task would last for five minutes and that the task was considered a failure if either

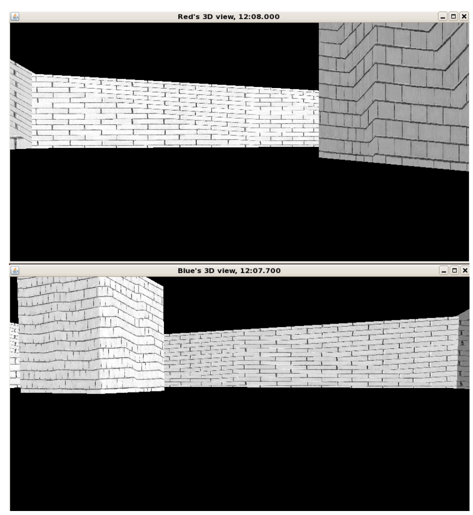

Figure 3. 3D view from robots' perspectives in navigation task. The red robot's view is above the blue robot's view.

robot did not find a transmission location in time. The robots moved continuously throughout each task run, except when (a) pausing to measure signal strength, (b) in a collision state with a wall or obstacle, or (c) at the target location. Participants were instructed to avoid collisions with obstacles and walls, and were advised not to leave either robot idle, as it may go into a hibernation state to save power. These constraints helped to ensure that the participants engaged in multitasking between the two robot consoles and did not focus on finding one robot's transmission point and then moving to the second.

Conditions

We compared our adaptive system with two alternate conditions in the experiment. One condition simply turned off autonomy as a baseline. The other used the *Brainput* inversely to probe more deeply into whether the brain input had any effect, similar to the experimental design in Pope et al [24]. Thus, there were three conditions in the study, varying only in the adaptive behavior that was triggered in the red robot by the fNIRS brain input:

1) In the *adaptive condition*—our system—the red robot went into autonomy mode whenever a branching state was detected—indicating that the user was tending to multiple tasks and maintaining information about the primary task over time. The red robot exited autonomy mode when a non-branching state was detected, requiring the human to give instructions to the robot about where to explore.
2) In the *non-adaptive condition*, the brain input was ignored, and the red robot never acted autonomously.
3) In the *maladaptive condition*, the rule was reversed from the *adaptive condition*. When a non-branching state was detected, the red robot began working autonomously and stopped when a branching state was detected, waiting for commands from the participant. The autonomy mode provided the same assistance as in the *adaptive* condition, and thus should still allow the user to focus on the blue robot. The only difference is the timing of the onset of the autonomy mode. This condition allows us to investigate the effect of the mapping between *Brainput* and the adaptive behavior.

Figure 4. Navigation controls for robot navigation task. There was a separate control for each of the robots, positioned to the left of the robot's 3D view (Figure 3).

In autonomy mode (regardless of which condition), the red robot would take over the search task, periodically sensing the signal strength and making appropriate course adjustments to ensure progress toward the target location. Note, however, that even in autonomy mode the robot could be interrupted by the operator (e.g., when asked to take a reading of the signal strength), but would return to the autonomous behavior after completing the requested action.

The blue robot never acted autonomously, as we wanted to ensure that the human operator always had a task to perform. The red robot staying in autonomous mode throughout the entire task would not be ideal as the human needs to be aware of the robot's location and progress to provide corrective feedback (as the robot's search behavior is not optimal), to ensure that the message is transmitted before time is up.

Experimental Procedure

Before the experiment, each participant completed a practice session without the fNIRS sensors, first with the robot navigation task and then with the multitasking exercises. This allowed the participants to familiarize themselves with each of the tasks. For the robot practice session, neither robot was autonomous as we simply wanted the participant to learn how to use the console and see the robot in action. In the multitasking practice exercises, the three distinct multitasking exercises (branching, dual task, and delay) were presented in a counterbalanced order. Each was repeated until the participant achieved greater than 80% accuracy in the task.

After the practice sessions, fNIRS sensors were applied to the forehead and the machine learning training session began. The participant completed known multitasking exercises to allow us to build an individual model of fNIRS activity for classification. The participant was presented with an initial rest screen, which was used to collect a one-minute baseline measure of the brain activity at rest. After that, the user had to complete two sets of ten 40-second trials. There were ten trials of *branching*, five of *delay* and five of *dual task,* which were presented in random order. We wanted to be able to distinguish branching from other non-branching states. Between each trial, the user was presented with the instructions for the next trial, followed by a 10-second rest screen.

Once the training session was complete, the data was used to build a training set for a machine learning model. Any trial where the participant achieved lower than 70% accuracy was not used as this indicated that they were not actually performing the task. For the main experimental task, we were most interested in detecting branching states, as the workload level can be quite high with the demand from both context switching and working memory load. The *dual task* and *delay* trials were combined into one category for the machine learning model as *non-branching*. If the number of *branching* and *nonbranching* training examples were not equal, the smaller set was oversampled so that the clas-

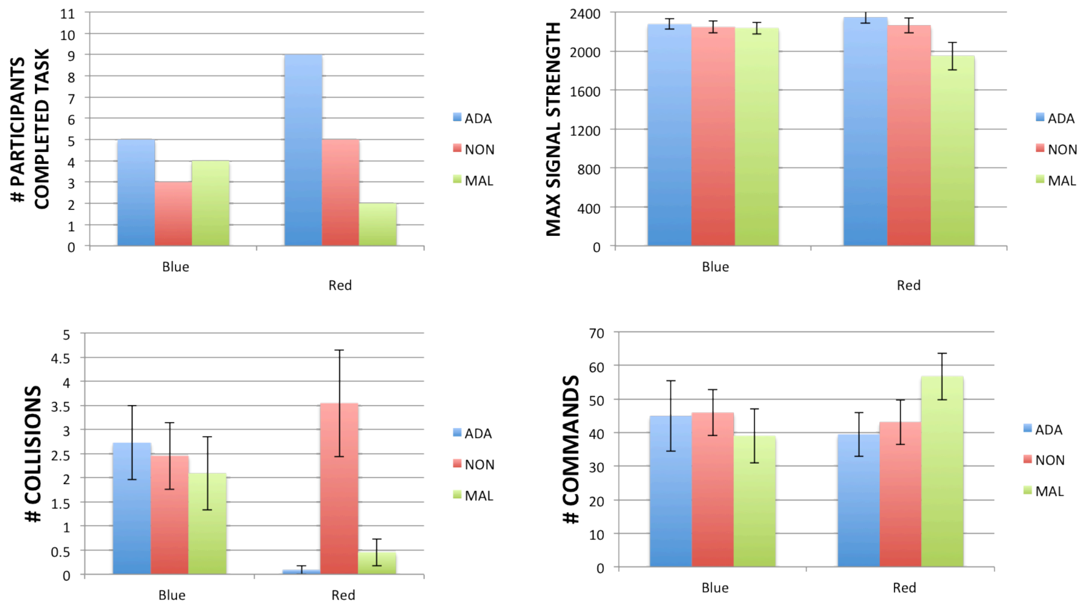

Figure 5. Performance results in relation to the two robots (blue and red) for the three adaptive conditions (adaptive, non-adaptive, and maladaptive). The top left figure shows the number of participants (out of 11 total) that completed each of the tasks. The other three figures show the means and standard error of the mean (SEM) for the maximum signal strength, number of collisions, and number of commands issued.

ses were balanced. The machine learning model was built using Weka's [8] SMO package for Support Vector Machines. For more details, see [30].

Once the model was built, the participant did one five-minute session of the robot navigation task in each of the three conditions. After each session, the participant filled out a NASA Task Load Index questionnaire (NASA-TLX) to provide their subjective assessment of task load. The first five participants also provided voluntary additional comments about their experience with each of the robots. To formalize this, the second set of six subjects also filled out a questionnaire on their perceptions of the robot in each of the three conditions. During the navigation tasks, the system logged all commands issued, fNIRS multitasking classifications received, and events such as collisions with obstacles.

Design and Analysis

The study used a within-subjects design. The independent variable is the robot's adaptive condition: *adaptive, non-adaptive* and *maladaptive*. All participants performed one five-minute session in each of the conditions. The condition order was counterbalanced. To evaluate whether the multitasking state information was valuable in the navigation task and produced differences between the three conditions, we investigated the following dependent measures: NASA-TLX questionnaire results, the robot perception questionnaire, and task performance from the log files, including number of completed tasks, number of commands issued, number of collisions, and maximum signal strength found.

RESULTS

Performance Results

We examined several aspects of task performance to see how they were affected by the adaptive condition (Figure 5).

First, since the autonomy mode should help the participant find the transmission location, we expected that we may observe higher task completion in both the *adaptive* and *maladaptive* conditions, over the *non-adaptive* condition where the user always had to control both robots. For the blue robot, we did find this (5, 4, 3 participants, respectively) but the result was not statistically significant. For the red robot, we did find a higher completion rate in the *adaptive* condition (9 out of 11 participants) than in the *non-adaptive* condition (5 out of 11 participants), as expected. However, the *maladaptive* condition had a lower completion rate (2 out of 11) than both the *adaptive* and *non-adaptive* conditions, indicating that the autonomy is helpful only when it is well-matched to the user's cognitive state. With a Cochran's Q test, we found a significant difference among the three adaptive conditions ($\chi^2(2) = 10.57$, $p < 0.01$). A pairwise comparison using continuity-corrected McNemar's tests with Bonferroni correction revealed that significantly more participants completed the task in the *adaptive* condition than in the *maladaptive* condition ($p<0.1$, $\phi = 0.48$).

To get a more fine-grained look at task completion, we investigated the maximum signal strength found (Figure 5, top right). Since the main goal was to find a transmission point above 2400, this could give an indication of how close the participants came to completion. Nonparametric analysis was used since the Shapiro-Wilk normality test showed that the data from each condition was not from a normal distribution. The *adaptive condition* resulted in the highest median for the maximum signal strength of the red robot (2416.0). The *maladaptive condition* resulted in the lowest median for the maximum signal strength of the red robot (2108.0). The *non-adaptive condition* was in the middle (2336.0). A Friedman nonparametric repeated measures ANOVA confirmed that the difference in the medians was statistically significant ($p<0.001$). Dunn's multiple comparisons post-hoc test was conducted and showed a significant difference between the *adaptive* and *maladaptive* conditions ($p<0.001$). There was no statistically significant difference in the medians of the maximum signal strength of the blue robot across the three conditions.

We then looked at the number of collisions in each of the conditions (Figure 5, bottom left). Participants were told to avoid collisions with walls and obstacles, as it would damage the robot. This ensured that both robots were attended to throughout the tasks. The Shapiro-Wilk Normality Test showed that the data from each condition was not taken from a normal distribution. A Friedman nonparametric repeated measures ANOVA was performed to compare the medians. We did not find any statistically significant difference for the blue robot. However, for the red robot, we did find that the *non-adaptive* condition resulted in a higher number of collisions than the two adaptive conditions ($p=0.005$). Collisions during the *non-adaptive* condition may indicate a performance degradation in the participant since a non-autonomous robot would likely walk into a wall if ignored, since there was no way to pause or stop the robot from moving.

Finally, as a measure of effort during the tasks, a repeated-measures analysis of variance was carried out to determine whether the adaptive mode had any effect on the number of commands issued (Figure 5, bottom right). For the red robot, there was a statistically significant main effect of adaptive condition, $F(2,20) = 3.691$, $p = 0.04$, but post-hoc analysis did not reveal any statistically significant results. There was no statistically significant difference for the blue robot, $F(2,20) = 1.153$, $p = 0.34$.

NASA-TLX Results

The goal of implementing adaptive behavior is to decrease the user's workload level. To investigate the success of this, we analyze the results of the NASA-TLX questionnaire [9]. This survey was designed to take into account individual differences in perceptions of workload. The questionnaire asks the participant to rate the workload level of the task in several categories. In addition, the user evaluates pairs of workload categories and indicates the one that contributes most to the workload. This is used to generate a set of weights that are applied to the other workload ratings and that reduce inter-subject variability. The means and standard error for the NASA-TLX scores from our study are shown in Figure 6. A repeated-measures analysis of variance on the NASA-TLX score showed that there was a sta-

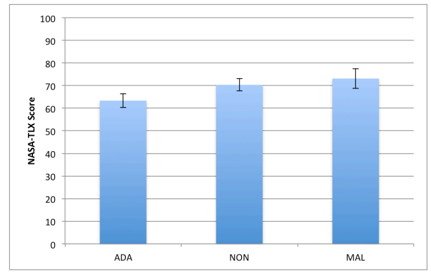

Figure 6. Mean and standard error in NASA-TLX results. There was a statistically significant main effect of adaptive condition, F(2,20) = 4.65, p = 0.02.

tistically significant main effect of adaptive condition, $F(2,20) = 4.65$, $p = 0.02$. A Tukey's pairwise comparison revealed the significant differences between *adaptive* and *maladaptive* ($p < 0.05$). Each dimension of workload (mental demand, physical demand, temporal demand, performance, effort, and frustration) was analyzed separately using Friedman's non-parametric repeated measures ANOVA. The adaptive mode had a significant effect on performance ($p<0.05$) and frustration ($p<0.05$).

Perceptions of Adaptive Behaviors

The first five participants in the study provided informal comments about the robots and the different conditions. Many commented on the behavior of the red robot. For example, after the *adaptive condition*, one participant said, "Although red robot occasionally disobeyed my commands, for the most part it was cooperative and found the transmission spot. Blue robot was still very cooperative." The same participant had this comment after completing the *maladaptive condition*, "Blue robot was much more cooperative than the red robot, which frequently disobeyed my commands and would go in its own direction." From these comments, we can see that the participant found the red robot to be mostly helpful in the *adaptive* condition, but that it disobeyed in the *maladaptive* condition. To capture a clearer picture of the perceptions of the robots, we had the next six participants complete a questionnaire on their perceptions of the robots. The results are illustrated in Figure 7.

This is a small sample of users, and so results of the questionnaire are preliminary. However, some patterns are beginning to emerge. First, we see that the participants seemed to agree that in the *maladaptive condition*, the robots appeared to make their own decisions and that the robots appeared to disobey the user's commands. There was less agreement on those points for the *adaptive condition*, even though the red robot was autonomous in this condition as well. This indicates that when the robot was autonomous at appropriate times (based on the *branching* classification from fNIRS), it was less noticeable to the user.

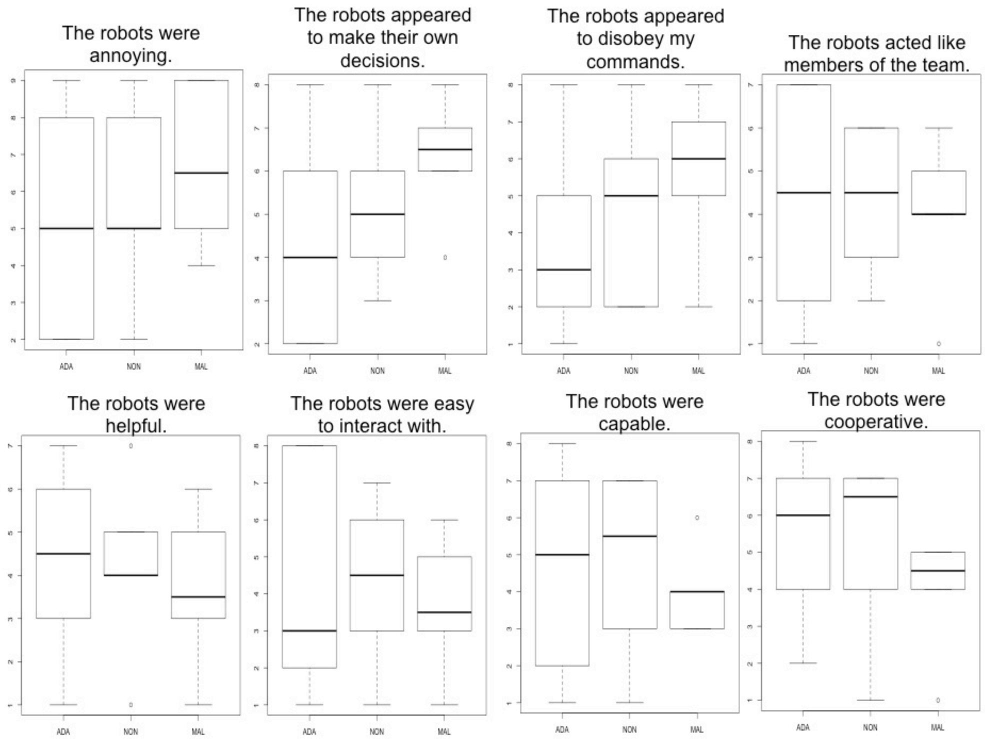

Figure 7. Preliminary results from robot perception questionnaire for the three conditions: *adaptive* (ADA), *non-adaptive* (NON) and *maladaptive* (MAL) (N=6). The scale was from 1 (strongly disagree) to 9 (strongly agree).

In the *maladaptive condition*, the participants indicated that the robots were more annoying than in the other conditions. This would make sense since they also felt that the robots were disobeying their commands. The lower score for "*The robots were cooperative*" in the *maladaptive condition* also corresponds with the other statements.

It is interesting to note that the ratings did not show strong differences between the *adaptive* and the *non-adaptive conditions* for these statements: "*The robots acted like members of the team*," *The robots were annoying*," "*The robots were capable*," and "*The robots were cooperative*." This provides evidence that the user hardly noticed the *adaptive* behavior when it was consistent with the user's needs. This is consistent with what we found in previous single-robot studies with both real and simulated robots [26].

DISCUSSION

Overall, our results suggest that *Brainput* provided measureable benefits to the user, with little additional effort required of the user. This study also confirmed that we can train a machine learning classifier on a set of known tasks and later successfully classify brain activity in unrelated activities that elicit similar brain processes.

The NASA-TLX results indicate that the *adaptive condition* had the lowest task load rating and the *maladaptive condition* had the highest, indicating that appropriate adaptive behavior helps to reduce workload, while adverse adaptations can actually make the system perform worse. In addition, the completion rates and maximum signal strength for the red robot were highest in the *adaptive condition* and lowest in the *maladaptive condition*, indicating that the adaptive behavior triggered by the *Brainput* correlates to performance improvements.

The red robot's autonomous behavior in both the *adaptive* and *maladaptive conditions* was appropriate: it made progress toward the target location. For this reason, our initial expectation was that task performance (in terms of number of successful task runs) would be improved in both adaptive conditions. As noted above, this was not the case for the *maladaptive condition*. This raises the question of how a properly functioning cooperative teammate could decrease performance. One likely explanation is hinted at by the number of commands issued in each condition. Participants issued more commands to the red robot in the *maladaptive condition* than in the other conditions, and the *maladaptive condition* is the only one in which the red robot received more commands than the blue robot. It seems that subjects were less accepting of autonomous behavior occurring during non-branching phases than during branching phases, and expended effort trying to "correct" the robot in that condition. This is reflected also in the subjective assessments: the robots were rated as more annoying and disobedient, and less helpful and cooperative in the *maladaptive condition*. These results demonstrate that basing the autonomy onset on the cognitive multitasking state has a positive impact on subjective task load.

As the name implies, the strategy adopted by the red robot in the *maladaptive condition* is not being proposed as a potentially viable candidate for future robotic architectures. Instead, the *maladaptive condition* is included to serve as a direct contrast to the *adaptive condition*, similar to the comparisons of positive and negative feedback loops in Pope et al. [24].

Comparing the *adaptive* and *non-adaptive conditions* demonstrates that robot autonomy can improve task performance, but that is not surprising, having been shown in prior work (e.g., [26]). What is unclear, however, is whether the *Brainput*-initiated autonomy transitions correspond to meaningful cognitive state transitions in participants. Periods of autonomy might seem likely to be helpful in a task like this regardless of when they occur, so comparisons between the adaptive and non-adaptive conditions cannot, by themselves, support claims regarding the legitimacy of *Brainput* classifications. However, contrasting the *adaptive* and *maladaptive* results makes it immediately apparent that *Brainput* has successfully identified a distinction in cognitive states: if *Brainput* were not detecting a genuine difference in cognitive load, one would expect no difference between the *adaptive* and *maladaptive* conditions, and could attribute all of the performance benefits to the proportion of the time spent in autonomy mode. Instead, participants respond significantly differently to autonomy initiated when *Brainput* indicates a branching state than to autonomy initiated when *Brainput* indicates a non-branching state. This constitutes strong evidence that the system is properly categorizing the fNIRS data—*Brainput* implicitly provides information, distinguishing between times in which autonomous operation is beneficial, and those in which autonomous operation is detrimental to the task.

FUTURE WORK

Multitasking has become integral in many aspects of our lives, so there are opportunities to explore *Brainput* in other tasks and domains. In any activity involving multitasking or information overload, we could expect to see improvements in the user's performance and experience. Some examples of other domains are complex data analytics, air traffic control and management of multiple unmanned vehicles.

Evaluating the system's performance with larger groups of simulated robots will give some indication of how the BCI mechanism scales up and (since additional robots is likely to increase the "ambient" cognitive load) allow us to test the granularity at which we are able to categorize operator load; it may be that there are useful distinctions to be made even within the high-load category. Also, it is likely that interacting with real, physical robots will lead to some differences in operators' overall cognitive states; evaluating *Brainput* with real robots will be important for determining its applicability to real-world problems.

Finally, there may be other cognitive states that could be exploited to improve human-computer interaction efficiency; an exploration of the system's ability to distinguish other states could, due to the passive approach to utilizing BCIs, lead to new enhancements at little to no cost to the operator.

CONCLUSIONS

Here, we take a different approach for brain-computer interfaces that augments traditional input devices such as the mouse and keyboard and that targets a wider group of users. We use brain sensor data as a passive, implicit input channel that expands the bandwidth between the human and computer by providing extra information about the user.

We successfully integrated the *Brainput* system into a robot architecture and demonstrated that it can successfully be used to reduce human workload in interactions with multiple robots. This is the first study to show the potential for fNIRS-based (as opposed to EEG-based) brain input in human-robot interaction. Moreover, the task used in this study is closely related to elements of many related real-world scenarios, including military operations using unmanned aerial vehicles for reconnaissance, search and rescue operations using robots to explore areas unsafe for humans, among others. In each of these scenarios, human operators must interact effectively with multiple robots in high-load, high-stress conditions in which the cost of failure is high.

Brainput gives the interactive system a valuable additional information channel that can be used to improve team performance without adding to the operator's load.

ACKNOWLEDGMENTS

We thank Audrey Girouard, Leanne Hirshfield, Krysta Chauncey, Margarita Parasi, Douglas Weaver, Orit Shaer, Michael Horn, Michael Poor, Megan Strait, Dan Afergan, Garth Griffin, Ruth Agbaji, Yunjie Tong, and the rest of the Tufts HCI Lab, Remco Chang, Desney Tan, Dan Morris, Missy Cummings and MIT HAL. We also thank the NSF for support of this research (NSF Grant Nos. IIS-0713506 and IIS-1065154 and 1136996 awarded to Computing Research Association for the CI Fellows project).

REFERENCES

1. Barbosa, A., Achanccaray, D. and Meggiolaro, M.A., Activation of a mobile robot through a brain computer interface. *Proc. IEEE ICRA*, (2010), 4815-4821.
2. Blankertz, B., et al. The Berlin brain-computer interface: Machine-learning based detection of user specific brain states. In G. Dornhege, J. d. R. Millán, T. Hinterberger, D. J. McFarland & K. Müller (Eds.), *Toward Brain-Computer Interfacing*: MIT Press (2007).
3. Coyle, S., Ward, T. and Markham, C. Brain-computer interface using a simplified functional near-infrared spectroscopy system. *J. of Neural Eng*, 4 (3). 219.
4. Escolano, C., et al. A telepresence robotic system operated with a P300-based brain-computer interface: Initial

tests with ALS patients. In *Proc. IEEE EMBC*, (2010), 4476-4480.

5. Fairclough, S.H. Fundamentals of physiological computing. *Interact. Comput*., 21 (1-2). 133-145.
6. Girouard, A., Solovey, E.T. and Jacob, J.K. Designing a Passive Brain Computer Interface using Real Time Classification of Functional Near-Infrared Spectroscopy. *IJAACS* (2010).
7. Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P. and Rao, R.P.N. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proc. CHI'08*, ACM Press (2008).
8. Hall, M., et al. The WEKA data mining software: an update. *SIGKDD Explor. Newsl*., 11 (1). 10-18.
9. Hart, S.G. and Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theorical research. in Hancock, P., Meshkati, N. ed. *Human Mental Workload*, 1988, pp 139 - 183.
10. Hirshfield, L.M., Solovey, E.T., Girouard, A., Kebinger, J., Sassaroli, A., Tong, Y., Fantini, S. and Jacob, R.J.K. Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload with Functional Near Infrared Spectroscopy. In *Proc. CHI'09*, ACM Press (2009).
11. Hirshfield, L.M., et al. This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy. In *Proc. CHI'11,* ACM Press (2011).
12. Jackson, M.M. and Mappus, R. Applications for Brain-Computer Interfaces. in Tan, D. and Nijholt, A. eds. *Brain Computer Interfaces: Applying our Minds to Human-Computer Interaction*, Springer (2010), 89-104.
13. Kennedy, P.R., Bakay, R.A.E., Moore, M.M., Adams, K. and Goldwaithe, J. Direct control of a computer from the human central nervous system. *IEEE Trans. on Rehabilitation Engineering*, 8 (2). 198.
14. Koechlin, E., Basso, G., Pietrini, P., Panzer, S. & Grafman, J. The role of the anterior prefrontal cortex in human cognition. *Nature* (1999), 148-151.
15. Kuikkaniemi, K., et al. The influence of implicit and explicit biofeedback in first-person shooter games. In *Proc CHI'10*. ACM Press (2010).
16. Lee, J.C. and Tan, D.S. Using a low-cost electroencephalograph for task classification in HCI research. In *Proc. UIST*, ACM Press (2006).
17. Mandryk, R., Atkins, M. and Inkpen, K. A continuous and objective evaluation of emotional experience with interactive play environments. *Proc. CHI*, (2006).
18. Nacke, L.E., Kalyn, M., Lough, C. and Mandryk, R.L. Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proc. CH1*, ACM Press (2011).
19. O'Hara, K., Sellen, A. and Harper, R. Embodiment in brain-computer interaction. *Proc. CHI*, (2011).
20. Okumura, M. and Zhiwei, L., On NIRS-based brain-robot interface. In *Proc. IEEE ROBIO* (2007), 864-869.
21. Pan, M.K.X.J., Chang, J.-S., Himmetoglu, G.H., Moon, A., Hazelton, T.W., MacLean, K.E. and Croft, E.A. Now where was I?: physiologically-triggered bookmarking. In *Proc. CHI*, ACM Press (2011).
22. Perrin, X., Chavarriaga, R., Colas, F., Siegwart, R. and Millán, J.d.R. Brain-coupled interaction for semi-autonomous navigation of an assistive robot. *Robot. Auton. Syst.*, 58 (12). 1246-1255.
23. Pfurtscheller, G., Flotzinger, D. and Kalcher, J. Brain-computer interface: a new communication device for handicapped persons. *J. Microcomput*. Appl., 16 (3). 293-299.
24. Pope, A. T., et al. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40, (1995), 187-195.
25. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N. and Wolpaw, J.R. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE *Trans. on Biomedical Engineering*, 51 (6). 1034.
26. Schermerhorn, P. and Scheutz, M. Disentangling the Effects of Robot Affect, Embodiment, and Autonomy on Human Team Members in a Mixed-Initiative Task. In *Proc. ACHI* (2011), 236--241.
27. Scheutz, M., Schermerhorn, P. and Kramer, J. The utility of affect expression in natural language interactions in joint human-robot tasks. *Proc. HRI* (2006).
28. Solovey, E.T., et al. Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines. In *Proc. UIST* (2009).
29. Solovey, E.T., et al., Sensing Cognitive Multitasking for a Brain-Based Adaptive User Interface. In *Proc. CHI*, ACM Press (2011).
30. Solovey, E. T. Real-time fNIRS Brain Input for Enhancing Interactive Systems. Doctoral Dissertation, Tufts University (2012).
31. Tsubone, T., Tsutsui, K., Muroga, T. and Wada, Y. Estimation of Force Motor Command to Control Robot by NIRS-Based BCI. in Masumi, I., Kenji, D., Hiroyuki, M. and Takeshi, Y. eds. *Neural Information Processing*, Springer-Verlag (2008), 969-978.
32. Wolpaw, J.R., McFarland, D.J., Neat, G.W. and Forneris, C.A. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78 (3). 252.
33. Yuksel, B.F., Donnerer, M., Tompkin, J. and Steed, A. A novel brain-computer interface using a multi-touch surface. In *Proc. CHI*, ACM Press (2010).