Dynamic Difficulty Using Brain Metrics of Workload

**Daniel Afergan[1], Evan M. Peck[1], Erin T. Solovey[2], Andrew Jenkins[1], Samuel W. Hincks[1],
Eli T. Brown[1], Remco Chang[1], Robert J.K. Jacob[1]**

[1]Tufts University
Medford, MA USA

[2]Drexel University
Philadelphia, PA USA

{afergan,epeck02,ajenki01,shincks,ebrown,remco,jacob}@cs.tufts.edu, erin@cs.drexel.edu

ABSTRACT

Dynamic difficulty adjustments can be used in human-computer systems in order to improve user engagement and performance. In this paper, we use functional near-infrared spectroscopy (fNIRS) to obtain passive brain sensing data and detect extended periods of boredom or overload. From these physiological signals, we can adapt a simulation in order to optimize workload in real-time, which allows the system to better fit the task to the user from moment to moment. To demonstrate this idea, we ran a laboratory study in which participants performed path planning for multiple unmanned aerial vehicles (UAVs) in a simulation. Based on their state, we varied the difficulty of the task by adding or removing UAVs and found that we were able to decrease errors by 35% over a baseline condition. Our results show that we can use fNIRS brain sensing to detect task difficulty in real-time and construct an interface that improves user performance through dynamic difficulty adjustment.

Author Keywords

BCI; passive brain-computer interface; dynamic difficulty; fNIRS; near-infrared spectroscopy; workload; UAV

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Task difficulty can induce a wide variety of cognitive and emotional states that can impact performance [16]. A problem that exceeds an individual's skill set can be overwhelming and cause anxiety, while a problem that does not utilize a person's skills may result in boredom and apathy. When users are in states of anxiety or boredom, they do not focus on their task, learn less, are less productive, and more prone to errors [30]. Minimizing anxiety and boredom during a task can keep people in a state of *flow*, helping a user maintain focus on the current task and perform well.

To keep people in an optimal state, *dynamic difficulty adjustment (DDA)* systems automatically modify the difficulty of the task in real-time. By monitoring user state and adapting the system when it detects detrimental states, a dynamic difficulty system improves performance and helps users maximize their amount of productive work. However, DDA systems face two primary challenges. First, DDA systems must infer a correct model of the user's skill level and willingness to engage in a task. This can be particularly difficult when performance metrics are hard to define or are not indicative of skill, or as skill improves. Second, DDA systems must subtly change difficulty level while not obtrusively interfering with the task in order to not interrupt presence and degrade performance [24, 41].

While there has been considerable research on DDA systems, there are few examples of DDA systems outside of gaming environments, where behavioral metrics are often a good indicator of performance and difficulty. Finding these indicators in the real world is challenging, but the benefits of systems with dynamic difficulty are widespread. For example, in computer-assisted learning environments, the system could calibrate the pace of learning or materials to the understanding of each individual. These principles can also be applied towards task allocation in order to use physiological signals to calibrate the amount of work given to a user or balance task load across team members.

In this paper, we demonstrate a set of adaptation techniques and triggers to address some of the traditional problems with using physiological input to DDA systems in the real world. As an example of a task allocation environment, we construct a system that aids operators in path planning for unmanned aerial vehicles (UAVs) and use input from *functional near-infrared spectroscopy (fNIRS)* brain sensing to model the operator's state. By using the confidence in its output, we construct an adaptive system that is more robust to the noise of physiological input. Finally, by being careful with our manipulation of visual elements, we prevent disrupting the user's mental model of the system. The UAV operator scenario is an ideal testbed to control task allocation because the path planning system is composed of discrete elements, each of which carries a level of independent workload.

In particular, we show that we can use fNIRS to model a user's working memory load in real time and use this signal as an indicator of difficulty in multitasking scenarios. We propose that fNIRS has specific advantages in being applied to DDA systems. It is a lightweight and non-intrusive sensor,

and can be affixed to the forehead while allowing users to still function normally. Additionally, it has been found to be resistant to movement artifacts in comparison to other sensors [35]. Because of this, it is ideal for a passive input, where users still perform a task normally but the system knows and adjusts to their overall cognitive state.

By applying fNIRS to a DDA system that supports UAV route planning, we find that we are able to improve task performance. We suggest that this is because we aid users in avoiding harmful cognitive states, thereby maintaining a state of *flow*. In this paper, we make the following claims:

- **We show that we can build a brain-computer interface that uses task allocation as means for dynamically matching the task to the user.** We apply functional near-infrared spectroscopy to detect signals in the prefrontal cortex that correlate with working memory load, using this signal as a proxy for the user's engagement with a task. We demonstrate that this system decreases operator error by 35% over a baseline condition, holding speed and other factors constant.

- **We suggest adaptive strategies that circumvent traditional problems with biocybernetic systems and make adaptation feasible for DDA systems in the real world, such as UAV path planning.** To minimize the impact of misclassifications, we only trigger the adaptive mechanism when the system has a high confidence in its modeling. To prevent disrupting the user's mental model of the system, we only modify those visual elements that the user has no reason to monitor.

Application: UAV Path Planning
For our testbed scenario, we use an unmanned aerial vehicle (UAV) software simulator. UAV operation lends itself well to task allocation because difficulty adjustments because operators go through periods of boredom and high workload, both of which lead to operator attention inefficiencies, delays, and degradations in performance measures [14]. Unmanned aircraft are an increasingly common solution used by the military for surveillance, intelligence, and combat. They are also frequently used for humanitarian assistance and disaster relief to assess damage and conduct search and rescue missions, as well as for transporting cargo and passengers.

Workload for a UAV operator is inherently dynamic. As autonomy capabilities improve, there is increasing interest in giving individual operators responsibility for multiple autonomous vehicles. Because of varying demands for each UAV, sometimes a lower number of UAVs will cause more work and we cannot measure operator workload based solely on number of vehicles [12]. A successful dynamic difficulty system would better optimize distribution of workload and the utilization of human resources to maximize performance. In addition, control systems could be alerted when more or fewer UAV operators are needed. However, such a system requires the user's state be captured in real-time, as performance metrics alone might not accurately reflect the mental demands of a UAV operator's current scenario. Although this study is based on a UAV simulation, we posit that it is generalizable to most systems with variable levels of workload that can be modified by task allocation.

BACKGROUND AND RELATED WORK

Flow and Dynamic Difficulty
As defined by psychologist and philosopher Mihaly Csikszentmihalyi, *flow* is the state of complete immersion in a task induced by achieving an ideal balance between the skills of the user and the challenge of the user's task [10]. When this balance is met, a user will typically experience a clear sense of goals, focused concentration, loss of self-consciousness, altered sense of time, increased sense of control, increased motivation, and a merging of action and awareness [10]. As depicted in Figure 1, keeping users in the flow state provides a positive emotional experience and avoids boredom, frustration, anxiety, or apathy [15, 26].
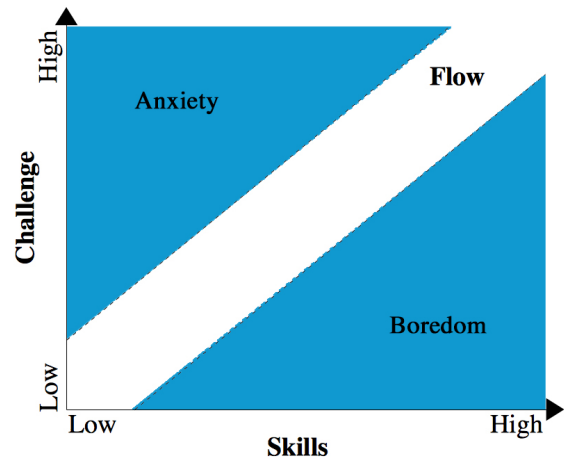

Figure 1. A two-dimensional model of flow

To capitalize on these benefits, designers have attempted to automatically maintain a user's flow state through the use of *dynamic difficulty adjustment (DDA)*, or adapting the challenge of a task automatically based on a model of the user's current skills. Research has shown that game players, for example, experience faster performance gains and feel more in control when the difficulty increases based on their skill [24, 39]. Scenarios that adjust difficulty according to performance keep the player more immersed than those that simply adjust difficulty over the course of gameplay [32]. Outside of the gaming industry, DDA systems can be used to accommodate the transition of users from novice to expert [6] — a central application of adaptive interfaces.

Dynamic Difficulty Psychophysiological Metrics
Psychophysiological signals can be used to determine player boredom and anxiety [8], and many researchers have looked at how these affect performance and flow [15, 32, 33]. Chanel et al. found that playing at different levels of difficulty will give rise to different emotional states and those emotional states (and the underlying conditions) can be assessed using physiological measures [8]. These signals can also be used

passively to modify a scenario by determining a user's affective, or emotional, state [15]. Participants' skills improve more and the participants find the simulations more enjoyable when the level of challenge is altered due to affective or mental state rather than their performance [33]. In addition, research has found that boredom and frustration/anxiety are distinguishable from each other and from flow, and that they can be classified with high levels of accuracy [8, 33].

The sympathetic and parasympathetic nervous sytem are susceptible to effects of cognitive state. Unobtrusive sensors such as electrocardiograph (EKG) electrodes or Galvanic Skin Response (GSR) can be used to find changes in a participant's physiological state. Heart rate variability and interbeat interval (IBI) measured from EKG can correspond to arousal levels. In addition, IBI standard deviation and GSR mean have been found to rise and peak as subjects' challenge states change from boredom to flow, but then fall again once the challenge becomes too difficult and subjects reach a state of anxiety, indicating that these physiological signals fall once the user becomes disengaged [16]. These signals are a good complement to fNIRS. However, they measure affective state and how the user feels about the challenge and are not necessarily descriptive of how the user's skills interact with the difficulty of the task.

Brain-sensing can be used to focus on a user's mental state. Clinical studies using electroencephalogram (EEG) have found significant results that can be used for dynamic adaptations, such as that left frontal lobe activity is correlated with positive emotions and motivation and that frontal theta rhythms are maximized when the challenge was also at its peak [15]. EEG can also find attentional engagement and affective workload in first-person shooters [1]. The key step in creating an affective loop is to identify appropriate classifiers that correlate biosignals and the actual affective state of the users. Fairclough et al. also found significant differences in EEG signals as task load in an n-back task changed and that systolic blood pressure (SBP) increases with task demand, peaks when task demand reaches the maximum challenge, and then decreases once the task is too challenging [15].

While the literature on using EEG as input to adaptive systems is quickly growing, it has historically been extremely sensitive to movement artifacts without careful filtering methods. We investigate the use of fNIRS as a less-explored device in comparison to EEG and complementary in many of its characteristics. In the next section we provide a more detailed description of fNIRS and its benefits for use in DDA systems.

fNIRS and Prefrontal Cortex
fNIRS uses near-infrared light to detect levels of oxygenated (HbO) and deoxygenated hemoglobin (HbR) on the surface of the prefrontal cortex. Light at this wavelength penetrates biological tissue and bone but is absorbed by hemoglobin in the bloodstream, and has similar vascular sensitivity to fMRI [21]. Since neural activity is accompanied by increased oxygen demands in order to metabolize glucose, much like fMRI, fNIRS can detect activation at localized areas of the brain. For a more in depth validation of fNIRS signals in comparison to fMRI, we refer to Strangman et al [38]. fNIRS

detects slow trends of hemodynamic changes, and is thus more appropriate for detecting overall state rather than event-related responses.

Recently, fNIRS has increasingly been leveraged to research users because it is considered to be safe, comfortable, relatively robust to movement artifacts, and can be designed for portability. In addition, it is resilient to head movement, facial movement, ambient noise, heartbeat, and muscle movement [17, 35]. This is critical for complex environments where the user must be able to function freely and normally.

Predictive models have been used to differentiate the fNIRS signal between levels of workload [4, 20, 28, 34], verbal and spatial working memory [19], and game difficulty levels [17]. Furthermore, it has been used to determine periods of cognitive multitasking [36, 37], levels of expertise [7], preference [25, 29], and emotion [40].

fNIRS is a viable technology to monitor the brain activity of UAV pilots in ecologically valid environments. Previous work has shown that mean oxygenation levels increase as the number of vehicles in a simulation increases, but then drop once the operator is tasked with too many elements [2, 7, 22].

Passive Brain-Computer Interfaces
Traditional active and reactive BCIs use physiological signals as input to directly control an application, such as moving a cursor/object, selecting an option, or other intentional actions. In contrast, passive BCIs are based on "reactive states of the user's cognition automatically induced while interacting in the surrounding system" [44]. Passive inputs assess user state and use that to help control interaction without direct or intentional effort from the user. These systems supplement direct input with implicit input, typically derived from physiological sensors attached to the user, in order to trigger commands based on user state. Jacob et al. anticipated the notion of the passive physiological monitoring in 1993 [23], and recent advancements in physiological sensors has allowed them to become feasible. Physiologically-sensed signals represent user state and can provide additional, implicit input to systems. Driven by more efficient monitors and the computational power and algorithms to process large quantities of data in real-time, modern technology can more affordably integrate passive systems and has spawned research into passive biocybernetic adaptation [15].

Adaptations triggered by passive input face two primary challenges: to accurately model the user's cognitive and emotional state and to sensibly adjust the system based on this model. The model must be very cautious since users are often in a different mental state during offline calibration and online feedback [42].

BCI helps solve the first challenge of passive systems by providing user models that tap more directly the source of user state. Cutrell and Tan suggest that the implicit commands inferred from a user's changing brain activity may be the most promising, universal application of BCI [13]. Explicit brain-issued commands suffer disproportionally from errors and have a limited range of input, whereas implicit commands

offer purely additional information without the user's deliberate attention, and the user does not see misclassifactions, nor have to spend additional cognitive resources recovering from these errors.

There are currently few examples of passive, real-time BCI systems that objectively improve user performance in a system. Prinzel et al. built an EEG-based adaptive automation system that alternated between manual and automated modes for a joystic tracking task and auditory pitch counting task [31]. Wilson and Russell designed a similar system based on an EEG engagement index for single-operator single-UAV systems and found success with slowing the UAV down or presenting alert in order to maintain engagement levels [43]. Solovey et al.'s *Brainput* system was able to adapt a scenario where an interactive human-robot system changed its state of autonomy based on whether it detected a particular state of multitasking [37] and Peck et al. demonstrated a passive, implicit adaptive movie recommendation system [28].

FNIRS AS INPUT TO DYNAMIC DIFFICULTY
In this paper, we explore whether we can improve performance by using fNIRS brain sensing as input to a dynamic difficulty system. To investigate this topic, we run a user study on an an task allocation system. As our testbed, we use a simulation designed for UAV path planning [18]. In this simulation, UAV operators are tasked with directing multiple UAVs at once, similar to an air traffic control task. Since this task can either be overwhelming or overly simple, we use fNIRS input to intelligently add or remove UAVs in order to provide an ideal challenge to the user. Mirroring the two primary challenges of DDA systems, we pose the following research questions:

- Can we use fNIRS to identify periods of low and high difficulty as operators perform a navigation task with multiple UAVs?

- Can we use real-time fNIRS data as input to a dynamic difficulty system and improve operator performance?

In the following sections, we describe the UAV simulation task, the details of our Brain-Based Dynamic Difficulty Engine, and the experimental design of our study.
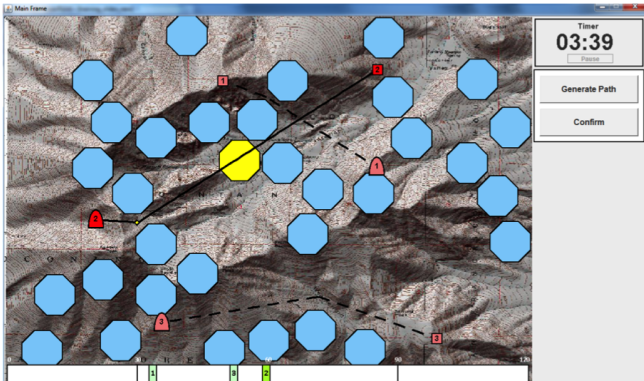

Figure 2. View of UAV operator simulation. Participants guided UAVs to targets while avoiding dynamic obstacles.

UAV SIMULATION TASK
We used a single-operator multiple-UAV system designed by a human automation research laboratory [18]. In the UAV navigation task, participants (or operators) were shown an overhead view of a map with UAVs (displayed in Figure 2) and instructed to guide the UAVs to a sequence of targets as quickly as possible while avoiding obstacles. Participants controlled between three and seven UAVs. Operators were instructed that obstacles, shown as teal octagons, were no-fly zones, and that while UAVs could fly through them, there would be a large penalty for doing so. If entered, obstacles should be exited as soon as possible. Leaving UAVs idle for a long period of time would also result in a penalty, so participants were motivated to balance performing the task quickly and without collisions.

The participants were instructed that they were part of a team of UAV operators and that vehicles would be passed off from their control to other operators, and other operators' vehicles would be passed to them. Thus, participants were prepared for vehicles to appear and disappear during the task. Intermittently, UAVs would be added or removed to modify the operator's current challenge level. To prevent disruption of the user's mental model of the scenario, UAVs were only removed if there were no obstacles in its path, meaning that the UAV should not demand any of the user's attentional resources and thus the user would not be distracted by the change. A pilot study showed that participants found this less disruptive than removing vehicles that needed path modifications, since a UAV could disappear during interaction. Overall, scenarios were hard-coded in order to guarantee consistency in obstacle density and path distance.

Interface Details
The vehicles were shown in pink and were numbered, with each UAV having a corresponding numbered target. When operators clicked on a UAV, that UAV and its target would become red and the user could click anywhere on the map to create a waypoint (indicated by a yellow diamond). When the user pressed "Generate Path," a path was created that included the starting position of the UAV, all of the waypoints, and the target. Once the participant pressed "Confirm," the vehicle would begin to follow the path.

The obstacles intermittently moved, so users could add more waypoints, move existing waypoints, or delete waypoints in the UAV's path. The currently selected path was shown as a solid black line, while other UAVs' paths were shown as dashed lines. When a vehicle's path crossed into an obstacle, the obstacle would turn bright yellow until the obstacle was cleared. Every time the vehicle reached a target, a new target would appear. Along the bottom of the map was a timeline so that operators could see when the UAVs would reach their targets, and a timer was on the top right.

BRAIN-BASED DYNAMIC DIFFICULTY ENGINE
In order to identify user states that correlate with periods of low or high difficulty, we depend on a number of analytical components that we call the *Brain-Based Dynamic Difficulty Engine*. In this section, we describe the process of calibrating

our system to recognize periods of low and high difficulty as the user interacts with an interface. We also describe our adaptation mechanism that attempts to increase or decrease the difficulty of the user's task.
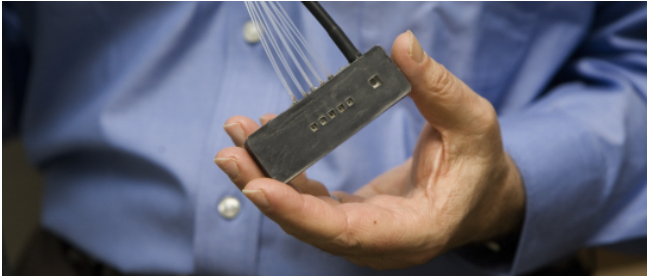

Figure 3. An fNIRS sensor that contains five light sources (only four were used) and one detector.

Equipment
We used a multichannel frequency domain Imagent fNIRS device from ISS Inc. (Champaign, IL) for data acquisition. The system used two probes (shown in Figure 3), one on each side of a subject's forehead, to measure changes in oxygenated and deoxygenated hemoglobin in the prefrontal cortex. Each probe had four sources, each emitting infrared light at two near-infrared wavelengths (830 nm and 690 nm), and one detector. The source-detector distances ranged from 1.5 to 3.5 cm and the system sampled at 11.79 Hz.
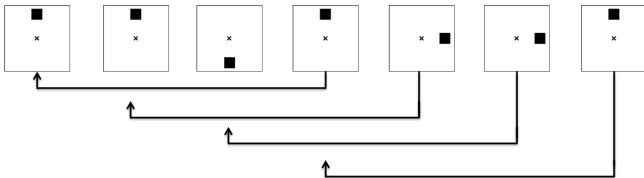

Figure 4. 3-back task. Participants were instructed to respond if the current square appeared at the same location as the one presented three trials back in the sequence.

Calibration Task
In order to identify fNIRS signals that determine periods of high and low periods of difficulty, participants first completed a series of visuospatial n-back trials. The n-back test is widely used and has been established and proven to incite increasing levels of short-term working memory as n increases [5, 27] as well as yield distinguishable fNIRS signals [11, 29]. Participants were shown a screen with a square and an X in the middle, with flashing black boxes (see Figure 4), and were asked to indicate if the currently presented stimulus was the same as the stimulus presented n trials previously. In order to induce levels of low and high difficulty, participants performed 1-back and 3-back trials. Because each participant only has to recall the last stimulus during a 1-back but has to remember the last three stimuli during a 3-back, the 3-back incites significantly higher cognitive load. Each trial consisted of 10 stimuli, each being displayed for 500 ms with an interstimulus interval of 2000 ms. Each trial totalled 25 seconds and was followed by a 15 second break. Participants completed

30 trials and were allowed to take a longer break after every 10 trials.

Filtering the fNIRS Signal
During the n-back task, raw fNIRS data was collected by Boxy (fNIRS acquisition software from ISS, Inc.) and sent to a specialized analysis system which was built in-lab using MATLAB. There, the system calculated the time series of change in light intensity compared to an initial baseline period of 30 seconds for each of our 16 information channels (2 probes x 4 light sources x 2 wavelengths). The signals were filtered for heart rate, respiration, and movement artifacts using a third-degree polynomial filter and low-pass elliptical filter. Further details of these algorithms can be found in [29, 37].

Modeling Difficulty
Previous work that used fNIRS to detect workload in UAV operator tasks took advantage of changes in the oxygenated hemoglobin signal. By calculating the mean and linear regression slope of operators' oxygenated hemoglobin levels in the prefrontal cortex, they were able to see signal differences that correlated with the number of UAVs and task difficulty [3, 7] Thus, for each n-back trial, we constructed an example of 32 features (16 information channels x 2 descriptive features) in order to build our model of user difficulty. We then fed each of these examples into LIBSVM, an integrated software for support vector classification [9], where we used a linear kernel to prevent overfitting the data [11]. In addition, we did a parameter search for the C and γ, the cost and gamma parameters for the machine learning model, in order to find the best model for each individual user. This technique uses the training data to build a support vector machine in order to classify new input and give a probability estimate. Finally, after building the model, we used 10-fold cross-validation to ensure that the model was accurate. For system implementation, see Figure 5. We used an individualized model for each user (Figure 6 is for illustration purposes), and this model was able to predict difficulty level while each user was performing the UAV task.

Classifying New fNIRS Data in Real-Time
After the model was created, we were able to predict user state in real-time by using a sliding window of the same size as our n-back trial length (25 seconds) on new, incoming fNIRS data. For each window of data, we used the same filtering techniques as the training series data and extracted the same mean and slope features.

Predictions and confidence values were sent to the UAV simulation software in real-time, which kept track of recent classifications and calculated its own prediction of user state by averaging the previous 8 seconds that were received from the model. Whenever the confidence average was above 80% for either low or high difficulty levels, a UAV would be added or removed, respectively. We base our confidence average off of results obtained from pilot studies. The adaptation was used to raise or lower the overall difficulty and challenge of the simulation when appropriate. This ensured that we would only make an adaptation if the system was confident in the
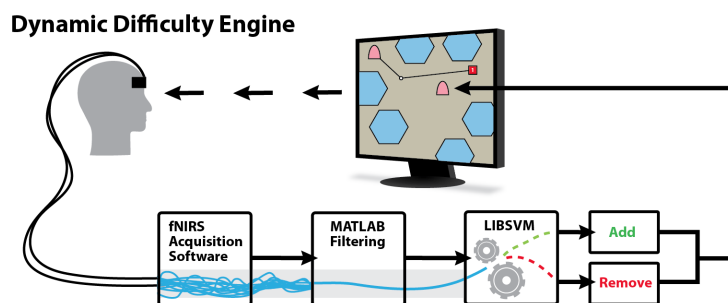
Figure 5. Diagram of our closed-loop dynamic difficulty adaptation engine. Raw signals acquired the fNIRS device are filtered, then used to classify user workload. When we are confident that the user is in a suboptimal state, we appropriately add or remove UAVs in order to provide the right amount of work.

user's state. In the following sections, we compare the success of this adaptation with a non-adaptive condition.

EXPERIMENTAL DESIGN

Training and Calibration
Participants first went through a self-guided slideshow explaining UAV simulation. They then practiced the simulation, spending five minutes controlling three UAVs and five minutes controlling seven UAVs. Our initial pilot tests showed that three UAVs was the minimum to incite workload, and that seven UAVs was the maximum that a single operator could handle without frustration. After this section, they were then described the calibration task – n-back task – and practiced a 1-back and 3-back task. Once the calibration task began, participants performed 30 n-back trials, 15 each of 1-back and 3-back trials, with a 15 second break between trials in order to allow their brain to return to resting state. The data from these trials was used to train our classifier. Although each participant had an individual model, the overall trends are shown in Figure 6.

UAV Scenario - Adaptive vs. Non-Adaptive
We compared our fNIRS-based adaptations to a control condition where UAVs were randomly added and removed. Thus, the independent variable in this experiment is the extent to which the system adapts to the user's state: *adaptive* and *non-adaptive*.

- In the *adaptive condition*, UAVs were added and removed according to brain signals correlating with low workload and high workload. We posit that extended periods of low and high workload cause boredom and anxiety, and lead to periods of disengagement and decreased performance. Thus, when the system was confident that the user was in a state of low or high workload, one UAV would be added or removed, respectively. After a UAV was added or removed, there was a 20 second period where no more vehicles were added or removed. This was done to prevent the user from having to rapidly switch contexts.

- In the *non-adaptive condition*, the simulation did not keep track of user state and intermittently added and removed UAVs (a random interval between 20 and 40 seconds). This

timing was determined based on a series of pilots to correspond as closely as possible to the average number of additions/removals we observed in the adaptive condition.

All participants ran one ten-minute trial of each condition. The order of conditions was counter-balanced across subjects and participants wore the fNIRS device during both trials so they did not notice a difference between the two conditions.

Dependent Measures
In order to evaluate the success of the system, we analyzed the following dependent measures:

- *Successes:* the number of UAVs that reached the target without entering a no-fly zone

- *Failures:* the number of UAVs that entered at least one no-fly zone on the way to the target.

- *Obstacles entered:* the total number of no-fly zones that planes entered during operation. This metric differs from failures because a UAV can enter multiple no-fly zones in a single flight.

- *Distance in obstacles:* the total distance (in pixels) that UAVs travelled in no-fly zones during operation. This measure helps us understand how quickly users were able to recover from these mistakes.

- *Average neglected UAVs:* a neglected UAV is idle waiting for a route or having a no-fly zone in its route and needing redirection. This metric refers to the avg number of neglected UAVs at any given moment during operation.

RESULTS
The study included twelve participants (five male) between the ages of 18 and 26. 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.

Improved Performance: Fewer Failures, Collisions
Across conditions, participants controlled roughly the same number of UAVs over time, averaging 4.41 UAVs in the adaptive fNIRS condition and 4.69 UAVs in the non-adaptive
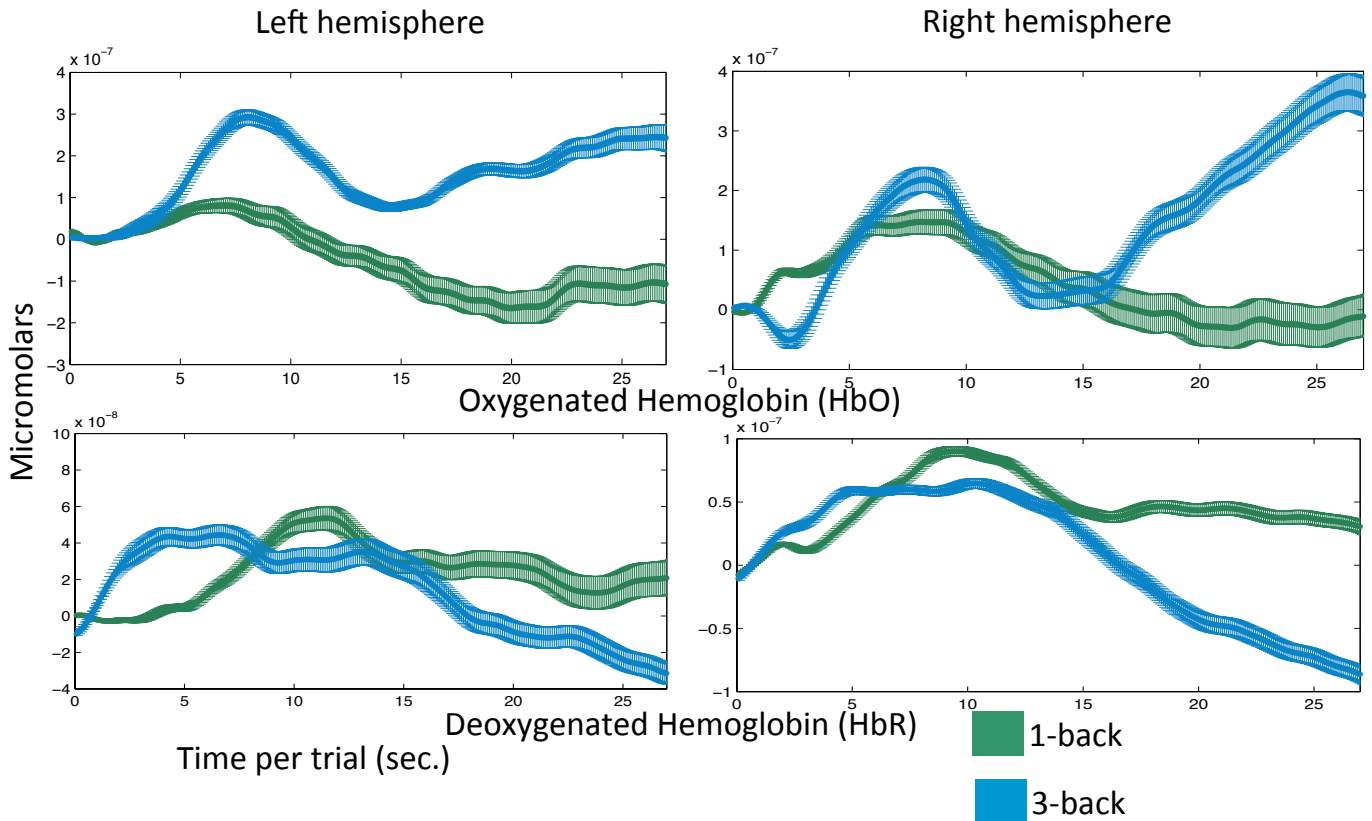
Figure 6. Although we model each participant individually, the fNIRS signal exhibited general trends as participants completed 1-back and 3-back trials. Plotting the mean change in HbO and HbR across all training trials and all participants (15 trials * 12 participants) shows that the 1-back task is correlated with a drop in HbO and increased levels of HbR, while the 3-back task induces an increase in HbO and decrease in HbR. The standard error of the mean is shown with the width of the plots.

condition (no significance in number of UAVs). Participants completed the same number of successful trials across conditions, so there was no significant difference in completion rates. However, their failure rate was significantly higher in the non-adaptive condition. With a paired t-test, we found that participants had significantly fewer failures in adaptive condition ($M = 3.25, SD = 2.14$) than the non-adaptive condition ($M = 5, SD = 1.95$) ($t(11) = 3.17, p < 0.01, Cohen's\ d = 0.92$). In addition, although there was no significant difference in the number of obstacles that appeared in the UAVs' paths across conditions (since the obstacles randomly moved over time), participants entered an average of 4.75 ($SD = 2.77$) no-fly zones in the adaptive condition and 7.42 ($SD = 2.81$) no-fly zones in the non-adaptive condition ($t(11) = 4.14, p < 0.01, Cohen's\ d = 1.2$).

While we did not see a change in number of successful trials, this is to be somewhat expected, since the operators' ability to complete their main objective was not affected by condition. However, the significant decrease in failures as well as variables associated with failures shows that users prevent degrades in performance. One likely explanation is that participants are paying more attention to the simulation and stay more engaged in the task.

Increased Awareness: Quicker Error Correction

We also measured the number of pixels that the UAVs traveled while in no-fly zones, an indicator of how aware the operators were that the UAVs were in obstacles. With a paired t-test, we found that participants traveled significantly less distance in obstacles in the adaptive condition ($M = 101.38, SD = 91.43$) than the control condition ($M = 206.39, SD = 179.1$) ($t(11) = 2.84, p < 0.05, Cohen's\ d = 0.82$). This means that they were more attentive to the appearance of obstacles in the adaptive condition and did a better job of avoiding the obstacles once the obstacles appeared. Participant also neglected UAVs less, having fewer average UAVs that needed attention. Participants had an average of 1.14 ($SD = 0.13$) neglected UAVs in the adaptive condition, with 1.37 ($SD = 0.29$) neglected UAVs in the non-adaptive condition ($t(11) = 2.78, p < 0.05, Cohen's\ d = 0.80$).

DISCUSSION

We presented a working dynamic difficulty system that used brain activity to determine when to perform task mitigations. In the background section of this paper, we described two primary challenges for DDA systems. First, DDA systems must accurately model the interaction between user skills and the challenge of their task. Second, DDA systems must modify the challenge to keep users in a state of flow *without* being disruptive. Mirroring these challenges, we posed research
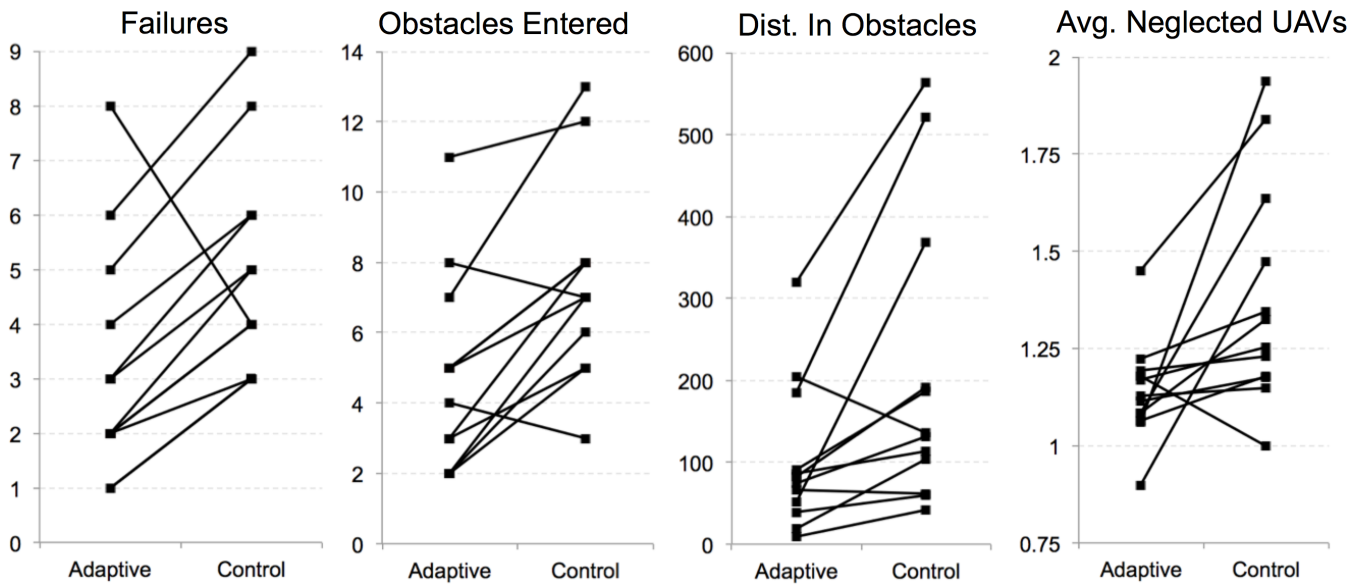
Figure 7. We use a slopegraph to plot the effects of condition on UAV operator performance for each participant. The following four measures showed significantly ($p < 0.05$) better performance in the adaptive condition interface (upward sloping lines), and at least ten of the 12 subjects performed better in the adaptive condition in each of these measures.

questions asking whether using fNIRS as input could address these challenges in a UAV operator task. In this section, we explore the implications of our study as it relates to these questions, beginning with the second of these challenges.

Using fNIRS to Minimize Disruption

It is worth reemphasizing the benefits of using fNIRS for ecologically valid evaluations of user state. Its resistance to movement artifacts in comparison to other many other sensing technologies means that users are able to engage with their task under fewer constraints. Earlier, we mentioned that an important goal of DDA systems was to minimize disruption within the context of adaptation. However, this disruption can also apply to the constraints required by measurement tools. Allowing the user to interact naturally will be important to consider if DDA systems are to be driven by physiological input in the future. While our system still requires users to be tethered by wires and fibers, there is a growing body of research that suggests that fNIRS will become even more lightweight and unobtrusive in the future [2, 41].

Mapping Working Memory to Difficulty

Our results indicate that we were able to model the working memory load of the user in the UAV scenario using fNIRS. Participants who interacted with the adaptive system saw performance boosts in the form of fewer failures and fewer collisions in comparison to a non-adaptive condition. Incorrectly modeling user state would have negatively impacted performance by pushing users into a potentially harmful state such as boredom or overload. Therefore, we were able to model user state sufficiently enough in a real-time environment to improve performance. Although misclassifications are inevitable, one of the challenges of applying adaptations is deciding when to take action. Our system shows that using a

timeseries of confidence classifications can aid the user by only adapting when we are fairly certain of user state.

However, one criticism for using working memory as a proxy for difficulty is that high working memory load may in fact be an indicator of engagement with a task. While we caution against mindlessly mapping working memory to difficulty, the UAV experiment exemplifies a larger classification of tasks where measuring high working memory load can indicate harmful user states. In multitasking scenarios, increasing the amount of mental effort required to accomplish a primary task necessarily decreases the available resources for secondary tasks.

We also believe that the user's increased awareness in the adaptive condition, as demonstrated by distance travelled inside obstacles and neglected UAVs, suggests that our adaptive mechanism successfully preserved the user's mental model of the scenario while modulating their challenge level.

Our system has measurable benefits, with no direct input from the user. This is advantageous because the user may not detect when he or she is disengaging, and will not have time or the extra cognitive capacity to indicate when the workload is overwhelming; continuously inputting information explicitly would be annoying and distracting.

This system is independent of the simulation software package and thus can be applied to any operator scenario. The predictions are sent as text strings (consisting of the prediction and the confidence rates) over TCP/IP, and a we have developed a generic Java class that reads in these predictions and gives moving averages of the predictions. Stripling et al. emphasize the need for a "robust user-friendly closed-loop training environment" that can produce automated methods for adaptive manipulations in complex simulations [39], and this system seeks to do that based on brain signals.

8

CONCLUSION

Dynamic difficulty adjustment is a valuable tool to keep people in an ideal state to learn better and perform better, while also enjoying the experience. As physiological sensors become more prevalent and the signal processing techniques continue to improve, passive systems will become easier to implement, more accurate, and more commonplace.

In this paper, we presented our system, the Brain-Based Dynamic Difficulty Engine, which uses fNIRS brain signals to dynamically adjust the difficulty level of a given task. Our system is real-time and passive, which is ideal for real world conditions such as UAV operations where the operators need to perform tasks and cannot tell the system when difficulty level is inappropriate. We demonstrated that our system can significantly decrease the operator's failure rate by 35% compared to randomly adding and removing over time, holding other factors constant. In addition, the operators exhibited higher level of awareness in dealing with UAVs that had immediate concerns. These results suggest that our system is beneficial to operators, and that fNIRS and other brain sensing techniques for DDA can be used for regulating user state and producing improved performance.

REFERENCES

1. Allison, B., and Polich, J. Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological Psychology 77* (2008), 277–283.

2. Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. Optical brain monitoring for operator training and mental workload assessment. *NeuroImage 59*, 1 (2012), 36–47.

3. Ayaz, H., Shewokis, P. A., Bunce, S., and Onaral, B. An optical brain computer interface for environmental control. In *IEEE EMBC 2011* (2011), 6327–6330.

4. Ayaz, H., Willems, B., Bunce, S., Shewokis, P. A., Izzetoglu, M., Hah, S., Deshmukh, A., and Onaral, B. Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations* (2010), 21–31.

5. Baddeley, A. Working memory: looking back and looking forward. *Nature Reviews Neuroscience 4*, 10 (2003), 829–39.

6. Bederson, B. Interfaces for Staying in the Flow. *Ubiquity* (2004).

7. Bunce, S. C., Izzetoglu, K., Ayaz, H., Shewokis, P., Izzetoglu, M., Pourrezaei, K., and Onaral, B. Implementation of fNIRS for monitoring levels of expertise and mental workload. In *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*. Springer Berlin Heidelberg, 2011, 13–22.

8. Chanel, G., Rebetez, C., Betrancourt, M., and Pun, T. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proc. Entertainment and Media in the Ubiquitous Era*, ACM Press (2008), 13–17.

9. Chang, C., and Lin, C. LIBSVM: a library for support vector machines. *Transactions on Intelligent Systems and Technology* (2011), 1–39.

10. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*. HarperCollins, 1991.

11. Cui, X., Bray, S., and Reiss, A. L. Speeded near infrared spectroscopy (NIRS) response detection. *PLOS ONE 5*, 11 (2010).

12. Cummings, M., and How, J. The Impact of HumanAutomation Collaboration in Decentralized Multiple Unmanned Vehicle Control. *Proc. IEEE 100*, 3 (2012), 660–671.

13. Cutrell, E., and Tan, D. BCI for passive input in HCI. In *Proc. CHI 2008*, ACM Press (2008).

14. Donmez, B., Nehme, C., and Cummings, M. L. Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 40*, 6 (2010), 1180–1190.

15. Fairclough, S., Gilleade, K., Ewing, K. C., and Roberts, J. Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems 6*, 1 (2013), 63–79.

16. Giakoumis, D., Vogiannou, A., Kosunen, I., Moustakas, K., Tzovaras, D., and Hassapis, G. Identifying Psychophysiological Correlates of Boredom and Negative Mood Induced During HCI. *Bio-inspired Human-Machine Interfaces and Healthcare Applications* (2010), 3–12.

17. Girouard, A., Solovey, E. T., and Hirshfield, L. M. Distinguishing Difficulty Levels with Non-invasive Brain Activity Measurements. *INTERACT 2009* (2009), 440–452.

18. Griner, A. Human-RRT Collaboration in UAV Mission Path Planning. *Master's Thesis, Massachusetts Institute of Technology* (2012).

19. Hirshfield, L., Gulotta, R., and Hirshfield, S. This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy. In *Proc. CHI 2011*, ACM Press (2011).

20. Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Jacob, R. J. K., Sassaroli, A., and Fantini, S. Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload with Functional Near Infrared Spectroscopy. In *Proc. CHI 2009*, ACM Press (2009).

21. Huppert, T., and Hoge, R. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage 29*, 2 (2006), 368–382.

22. Izzetoglu, K., Ayaz, H., and Menda, J. Applications of Functional Near Infrared Imaging: Case Study on UAV Ground Controller. In *Foundations of Augmented Cognition*. Springer Berlin Heidelberg, 2011, 608–617.

23. Jacob, R. J. K., Leggett, J. J., Myers, B. A., and Pausch, R. Interaction Styles and Input/Output Devices. *Behaviour and Information Technology 12*, 2 (1993), 69–79.

24. Jegers, K. Pervasive game flow: understanding player enjoyment in pervasive gaming. *Computers in Entertainment (CIE) 5*, 1 (2007).

25. Luu, S., and Chau, T. Decoding subjective preference from single-trial near-infrared spectroscopy signals. *Journal of Neural Engineering 6* (2008).

26. Nacke, L., and Lindley, C. Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, ACM Press (2008), 81–88.

27. Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping 25*, 1 (2005), 46–59.

28. Peck, E., Yuksel, B., Ottley, A., Jacob, R. J. K., and Chang, R. Using fNIRS Brain Sensing to Evaluate Information Visualization Interfaces. In *Proc. CHI 2013*, ACM Press (2013).

29. Peck, E. M., Afergan, D., and Jacob, R. J. K. Investigation of fNIRS brain sensing as input to information filtering systems. *Augmented Human 2013* (2013).

30. Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., and Perry, R. P. Boredom in achievement settings: Exploring controlvalue antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology 102*, 3 (2010), 531–549.

31. Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component. *Human Factors: The Journal of the Human Factors and Ergonomics Society 45*, 4 (2003), 601–613.

32. Qin, H., Rau, P., and Salvendy, G. Effects of different scenarios of game difficulty on player immersion. *Interacting with Computers 22*, 3 (2010), 230–239.

33. Rani, P. Maintaining optimal challenge in computer games through real-time physiological feedback mechanical engineering. *Pulse* (2005).

34. Sassaroli, A., Zheng, F., Hirshfield, L. M., Girouard, A., Solovey, E. T., Jacob, R. J. K., and Fantini, S. Discrimination of Mental Workload Levels in Human Subjects with Functional Near-Infrared Spectroscopy. *Journal of Innovative Optical Health Sciences 1*, 2 (2008), 227–237.

35. Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., Fantini, S., and Jacob, R. J. K. Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines. In *Proc. UIST 2009*, ACM Press (2009), 157–166.

36. Solovey, E. T., Girouard, A., Jacob, R. J., Lalooses, F., Chauncey, K., Weaver, D., Parasi, M., Scheutz, M., Sassaroli, A., Fantini, S., and Schermerhorn, P. Sensing cognitive multitasking for a brain-based adaptive user interface. In *Proc. CHI 2011*, ACM Press (2011).

37. Solovey, E. T., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., and Jacob, R. Brainput: Enhancing Interactive Systems with Streaming fNIRS Brain Input. In *Proc. CHI 2012*, ACM Press (2012).

38. Strangman, G., Culver, J. P., Thompson, J. H., and Boas, D. A. A Quantitative Comparison of Simultaneous BOLD fMRI and NIRS Recordings during Functional Brain Activation. *NeuroImage 17*, 2 (2002), 719–731.

39. Stripling, R., Coyne, J., Cole, A., Afergan, D., Barnes, R. L., Rossi, K. A., Reeves, L. M., and Schmorrow, D. D. Automated SAF adaptation tool (ASAT). In *Foundations of Augmented Cognition*. Springer Berlin Heidelberg, 2007, 346–353.

40. Tai, K., and Chau, T. Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface. *Journal of Neuroengineering and Rehabilitation 6* (2009), 39.

41. Tan, D., and Nijholt, A. Brain-Computer Interfaces and Human-Computer Interaction. In *Brain-Computer Interfaces*. Springer London, 2010, 3–19.

42. Vidaurre, C., Kawanabe, M., von Bünau, P., Blankertz, B., and Müller, K. R. Toward unsupervised adaptation of LDA for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering 58*, 3 (2011), 587–97.

43. Wilson, G. F., and Russell, C. A. Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. *Human Factors: The Journal of the Human Factors and Ergonomics Society 49*, 6 (2007), 1005–1018.

44. Zander, T. O., Kothe, C., Welke, S., and Rötting, M. Utilizing secondary input from passive brain-computer interfaces for enhancing human-machine interaction. In *Foundations of Augmented Cognition, Neuroergonomics and Operational Neuroscience*, Springer Berlin Heidelberg (2009), 759–771.