Adaptive Training in an Unmanned Aerial Vehicle: Examination of Several Candidate Real-time Metrics

Ciara Sibley[2], Anna Cole[2], Gregory Gibson[1], Daniel Roberts[3], Jane Barrow[3], Carryl Baldwin[3], Joseph Coyne[1]

[1]Naval Research Laboratory

[2]Strategic Analysis

[3]George Mason University

ABSTRACT

The present study examined the sensitivity of several candidate metrics of real-time workload within the spatial component of an unmanned aerial vehicle (UAV) task. Advanced Brain Monitoring's (ABM) wireless B-Alert system was used to collect participant's EEG workload and engagement data. Eye tracking data was also collected. The UAV simulation required participants to report heading information of moving vehicles, as seen from the UAV. There were four blocks of difficulty, over which a significant performance decrement was shown. Additionally, participants rated their workload significantly higher and pupil diameter significantly increased across blocks of increasing difficulty, as well as within each block during periods of highest mental demand. ABM's workload and engagement metrics however did not show a significant change over or within blocks. The results showed that pupil diameter shows promise as a correlate of mental workload.

Keywords: Mental Workload, Training and Simulations, Augmented Cognition, Adaptive Training

INTRODUCTION

Augmented Cognition emphasizes the use of a closed-loop system using real-time physiological assessment to improve human performance (Schmorrow & Stanney, 2008). In a training environment closed-loop systems could reduce the time required to train an individual by keeping workload at an optimal level for learning (Coyne, Baldwin, Cole, Sibley, & Roberts, 2009). Several metrics such as pupil diameter and electroencephalographic (EEG) have been shown to vary predictably with increases and decreases in workload. Monitoring these different metrics allows training to be optimized in computer based training (CBT) environments. Ultimately, this research will impact the way CBT is conducted by establishing the foundation for adaptive automation through monitoring neural resources.

EEG and eye tracking metrics have been extensively investigated as a means of assessing cognitive workload. For example, (Berka et al., 2007) developed a mental workload metric based on an individual's EEG signal that tracks task demand in mental arithmetic and digit span tasks. Other researchers have focused on eye tracking metrics and found changes in pupil diameter, fixation duration, and blink frequency to be predictive of various levels of cognitive demand in a task (Tsai, Viirre, Strychacz, Chase, & Jung, 2007; Van Orden, Limbert, Makeig, & Jung, 2001; Veltman & Gaillard, 1996). Additionally, many researchers have had success using artificial neural networks (ANN) to accurately classify different operator states for individuals (Wilson, 2005; Wilson & Russell, 2003) and improve performance with the aid of adaptive automation (Wilson & Russell, 2007).

Recent advances in eye tracking and EEG technologies have made utilizing closed-loop systems based on physiological measures more feasible. For example, accurate and unobtrusive off-the-head eye trackers now allow and account for head movements and can collect and process data in real-time. Furthermore, technologies like wireless EEG caps and dry, no-prep electrodes have recently been developed (Christensen, Estepp, Wilson, & Davis, 2009); both of which reduce the prep time normally required. Both EEG and eye tracking data can also now be collected and run using affordable personal computers that are capable of processing and storing large amounts of data. These and other similar advances have made it viable to utilize this type of technology in a CBT environment.

The ultimate goal of this multi-year effort is to build an automated training environment where objective physiological metrics along with subjective workload ratings and quantifiable performance measures can be used to classify an individual's workload and guide desktop training simulations. The purpose of the current study, reported here, was to examine neurophysiological markers of workload in a simulated UAV task at varying levels of difficulty.

METHOD

PARTICIPANTS

All participants (N= 15) were volunteers recruited from the Naval Research Laboratory. None of the participants had any prior experience with UAV simulators. Two were dropped from the study: one was due to second day attrition and the other because of partial dropped eye tracking data. Therefore, thirteen participant's eye tracking and performance data were analyzed and only the last nine participant's electroencephalographic (EEG) data were analyzed due to a hard drive error that caused four participant's data to be lost.

MATERIALS

Advanced Brain Monitoring's (ABM) wireless B-Alert system was used to collect participant's EEG data. The system uses a wireless six channel head cap that transmits data via Bluetooth to a PC running ABM's B-Alert software. ABM's classification algorithms assessed raw EEG and provided a second by second workload and engagement metric on a scale of 0-1. In addition, the Tobii X120 off-the-head eye tracker was used to collect pupil diameter and gaze position data. The unit was placed in front of the participant and just below the surface of the monitor running the simulation. The system recorded both eyes at 120 samples per second.

Virtual Battlespace 2 (VBS2) by Bohemia Interactive, Australia was used to construct the UAV simulation scenarios. VBS2 is a high-fidelity, 3-D virtual training system used for experimental and military training exercises. One Windows PC ran the UAV scenario, while a second PC recorded the eye tracking data, and a third recorded the EEG data. All computers were time synched using network time protocol in order to ensure accurate post-hoc data analysis.

TASKS AND PROCEDURES

UAV DESKTOP SIMULATION

After receiving a brief PowerPoint training about the task, participants engaged in a UAV desktop simulation created from videos using VBS2 where they were trained to report information on enemy targets as seen from a UAV. A continuous video stream from the UAV was shown on the monitor (Image 1) and participants were asked to report heading information about the target vehicles crossing the screen. Participants were given the heading of the UAV and were required to estimate the heading of the vehicle on the ground. A graphical depiction of a compass facing due north with 30 degree increments was provided to the participant for reference. After entering the target heading estimation, participants were then asked to rate their mental effort in calculating the target heading.

The difficulty of the task progressed over four blocks of trials. Only one vehicle was shown on the screen at a time and a total of sixteen vehicles were shown within each block. Difficulty was manipulated by varying the UAV heading as well as the possible target heading. For example, the easiest level (block one) showed the UAV heading at only 0 degrees and the target's heading could be either 0, 90, 180 or 270 degrees. The most difficult level (block four) showed the UAV heading at various 30 degree increments, which changed after every two targets, and the target heading could be any 30 degree increment.

Since this simulation is ultimately intended to help train a UAV operator, the order of difficulty levels were not randomized. On the first day, participants only completed one block, referred to as the baseline block, which was the equivalent difficulty level of block four. On the second day of the experiment, participants progressed through the task from block one to block four. This was done in order to assess learning, by comparing performance on the baseline block and block four. Each block took approximately eight minutes to complete.

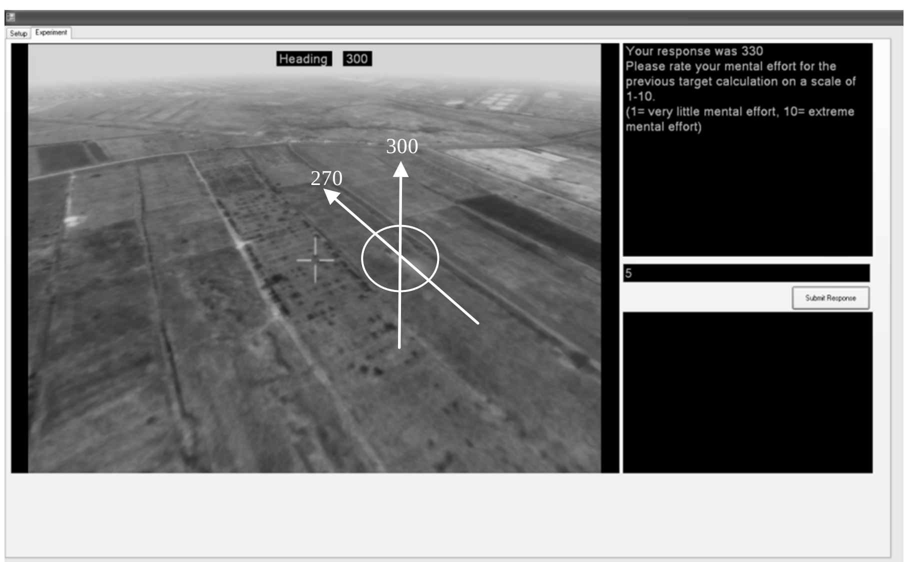

IMAGE1. Screenshot of the UAV simulation. Note the dust trail of an enemy vehicle just to the right of center. Based on the given UAV heading of 300, the participant would correctly report this vehicle heading as approximately 270°.

THE EXPERIMENT

All participants took part in two, one hour sessions over two days. At the beginning of each day, participants were prepped for EEG recording with ABM's six electrode wireless headset. Both EEG and eye tracking data were collected while participants were engaging with the UAV simulation.

On the first day, participants completed ABM's thirty-minute vigilance task. This task was developed by ABM as a means to filter out noise and uniquely fit classification algorithms to a participant in order to assess various levels of cognitive state. The vigilance task and software are part of ABM's real-time EEG classification system. After completing that task, any subsequent EEG data was run through ABM's classification algorithm to provide an individual's workload and engagement in real-time. After this process, participants reviewed a PowerPoint presentation that contained an overview of the tasks and training on how to complete the heading determination task. Participants were given a brief practice on the task and they next completed the experimental baseline block.

On the second day, participants were prepped for EEG and the experimenter briefly reviewed the task instructions. Following the instructions, participants began the UAV simulation while participant performance, EEG, and eye tracking data were collected, along with subjective mental effort ratings. All participants proceeded from blocks one through four with targets appearing at the exact same time, in the same order.

RESULTS

BEHAVIORAL PERFORMANCE

Analysis of performance data for blocks one through four confirmed effective manipulation of difficulty among levels within the UAV simulation. A significant difference existed among blocks one through four in heading error, $F(3, 36) = 16.52$, $p = .000$, $\bullet^{2} = .75$, subjective workload ratings, $F(3, 36) = 43.47$, $p = .000$, $\bullet^{2} = .78$, and for errors of omission, $F(3, 36) = 4.50$, $p = .006$, $\bullet^{2} = .29$. Heading error was computed by dividing the error from correct heading answer by 180 degrees; subjective ratings were on a scale of one to seven; and errors of omission were averaged over the entire block. See Figure 1 for a depiction of these effects.

While a statistically significant difference does not exist between heading error on the baseline block (M= 0.16, SD= 0.08) and block four (M= 0.13, SD= 0.07), the average error did decrease slightly and errors of omission decreased from 1.69 on the baseline block to 0.92 on block four.

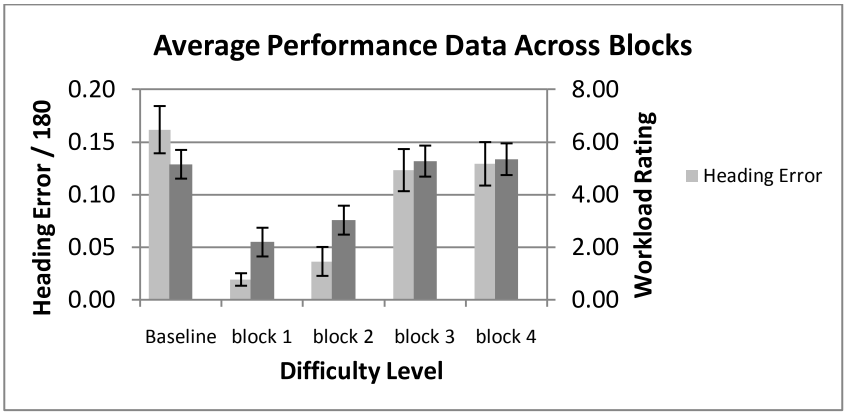

FIGURE 1. Average heading error and subjective workload ratings across all blocks

ABM'S WORKLOAD AND ENGAGEMENT INDICES

Preliminary analysis of the ABM workload and engagement metrics showed almost identical levels of workload and engagement when the metrics were averaged within each block and then compared across block levels. Thus, we further investigated the metrics by averaging each classification over the three seconds preceding participant response for each target heading. This time was chosen because it should correspond with when the participant is calculating the target heading, and thus is most cognitively loaded. Still, results revealed no significant difference in the ABM's workload metric across blocks one through four, $F(3, 24) = 1.62$, $p = .211$. Similarly, no significant difference existed in ABM's engagement metric across blocks one through four, $F(3, 24) = 1.41$, $p = .265$. See Figure 2.

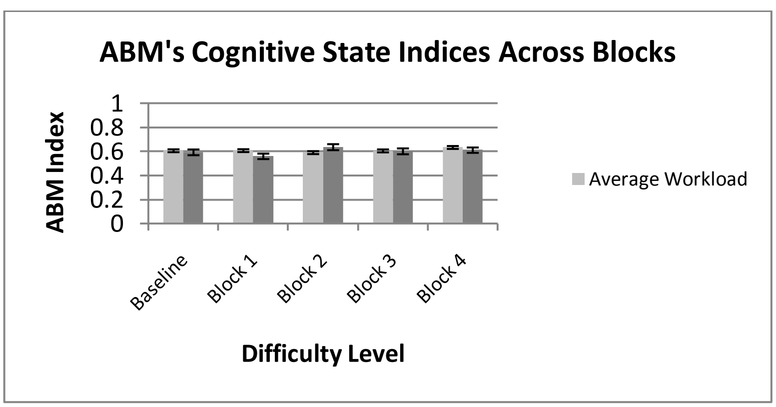

FIGURE 2. ABM's engagement and workload indices averaged three seconds prior to participants providing their heading response

PUPILLOMETRY

Pupil dilation was investigated as a measure of mental workload, and consequently pupil diameter was averaged within each entire block and compared across difficulty levels. Analysis revealed significant differences in pupil diameter among blocks one through four for the left eye ($F(3, 36) = 6.9$, $p = .005$, $\bullet^2 = .37$) as well as the right eye ($F(3, 36) = 6.9$, $p = .008$, $\bullet^2 = .37$), as shown in Figure 3. Analysis of the pupil diameter between the baseline and block four yielded some interesting results. A significant difference between the baseline (M= 3.34, SD= .46) and block four (M= 3.25, SD = .39) did exist for the left eye, $F(1, 12) = 5.18$, $p = .042$, $\bullet^2 = .30$. However no significant differences existed between the baseline (M= 3.34, SD= .36) and block four (M= 3.31, SD= .40) for the right eye, $F(1, 12) = 0.17$, $p = .688$, $\bullet^2 = .02$.

Further investigation of pupil dilation also prompted averaging pupil size over the immediate seconds preceding participant response for each target heading. Increments of one, three, and ten seconds were investigated and all yielded similar results. In particular, pupil diameter across blocks one through four was significantly larger one second preceding heading response when averaged across the whole block for the left eye, $F(1, 12) = 64.96$, $p = .000$, $\bullet^2 = .84$, and the right eye, $F(1, 12) = 88.11$, $p = .000$, $\bullet^2 = .88$. This suggests that pupil dilation is sensitive to phasic changes in workload over a small amount of time and confirms pupil dilation as a highly promising correlate of workload. See Figure 4 for a comparison of the different average time increments.

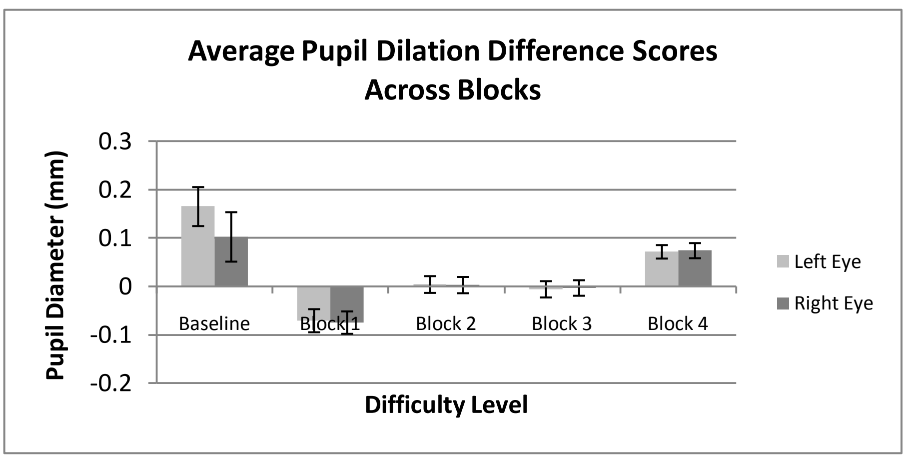

FIGURE 3. Average of all participants' pupil size difference from his or her average pupil size for each block

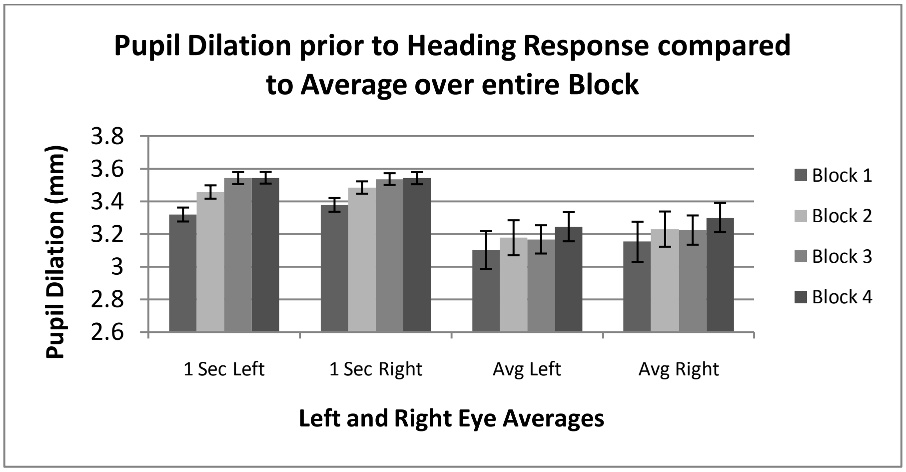

FIGURE 4. Average pupil dilation averaged one second prior to heading response compared to pupil dilation averaged over the entire block

DISCUSSION

Analysis of the performance data and subjective workload ratings indicated that the various levels of difficulty were successfully manipulated across the task. Subjective workload ratings, errors of omission, and heading error all increase in accordance with increasing levels of difficulty (i.e. from block one to block four). While a significant difference does not exist between heading error for the baseline block and block four, about half as many errors of omission occurred on block four (i.e. a failure to respond due to time pressure or simply not knowing the answer). Thus, heading errors on block four could be influenced by fewer omissions, and therefore be slightly higher than if the omission rate were the same between blocks four and the baseline block.

Comparison of performance between the baseline and block four are of interest as a means of assessing the UAV simulation as a potential training simulation. Due to potentially a lack of power and other factors, pre (baseline block) and post (block four) test training effects weren't statistically different. However, the total time allotted to training was only about forty minutes over both days, since each block took about eight minutes to complete. Hence, with more time to train an individual at each level in a real world training simulation, one would expect to see smaller heading errors and errors of omission by the end of training, compared to the baseline trial. In addition, one would expect ratings of workload to be significantly lower on the post test than the pre test.

Neither ABM's workload index nor engagement index were sensitive to changes in this task across difficulty levels. Changes were also not apparent when the index was calculated three seconds prior to heading response, when workload and engagement should have been highest within the block. On account of these findings, future studies will not be using ABM's cognitive state classification algorithms, but instead will investigate the use of artificial neural networks as a means of assessing workload in a UAV training simulation.

The most promising results of this study were systematic changes in pupil dilation as a function of difficulty level. The initial analysis of pupil diameter was performed by averaging an individual's pupil diameter over each eight minute block. Simply comparing average block pupil dilations yielded significant differences in pupil size across blocks (see Figure 3). Further investigation showed that average pupil size one second prior to submitting heading response was significantly higher compared to pupil size during the rest of the block (see Figure 4). This pre-response computation was also calculated at three and ten seconds preceding response, and yielded similar effects; indicating that this effect was likely not due to some kind of response initiation. Therefore, pupil dilation is not only sensitive to changes in workload over large periods of time, but also is sensitive within the demands of a task. These results substantiate the robustness of pupil dilation as a means of assessing cognitive load.

One surprising result was the large difference between average left and right pupil diameter for the baseline block, that actually yielded differing results when comparing dilation between the baseline block and block four. Left eye data is consistent with research that suggests differences in workload across difficulty levels should diminish with practice (Berka et al., 2004). However, data from the right eye would suggest that this is not the case. At present, further investigation is necessary before any firm conclusions can be drawn.

Future studies are planned to investigate how measures of workload change with practice within a difficulty level. In particular, other eye tracking metrics, such as blink frequency/duration, fixation frequency/duration, and divergence will be investigated. Analysis of blink data were not possible for this study, due to the inability to reliably differentiate lost eye tracking data from blinks using the Tobii eye tracking system. Future studies will use EOG to solve this problem. Fixation data and nearest neighbors analyses also were not possible to analyze because of too much error in the Tobii calibration. This problem has been resolved with new software that will be incorporated into future studies.

Another area of interest will be collecting physiological data when a participant is overloaded. We intend to increase the difficulty level of the hardest block in order to purposely overload the participant. Additionally, fewer blocks will be necessary since it is difficult to distinguish four distinct levels in the performance data. Three levels with more trials in each level will be used in follow up studies.

Overall, these findings show promise for using pupil diameter as a means of assessing workload. More data collection is necessary to investigate other eye tracking and EEG correlates. Using spectral analysis of the EEG recordings may prove more sensitive than the ABM engagement index explored in the present study. Ultimately, with the combination of performance, subjective ratings, eye tracking data, and EEG, we are confident that we will be able to successfully predict user workload and eventually perform mitigations within a closed loop system.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research's Human Performance and Education Program.

REFRENCES

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., & Olmstead, R. E. (2004). Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. *International Journal of Human-Computer Interaction, 17*(2), 151 - 170.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation Space and Environmental Medicine, 78*(5 II).

Christensen, J. C., Estepp, J. R., Wilson, G. F., & Davis, I. M. (2009). A Demonstration of a Dry/No Preparation Electrode System for EEG. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 53*, 1652-1653.

Coyne, J. T., Baldwin, C., Cole, A., Sibley, C., & Roberts, D. (2009). *Applying Real Time Physiological Measures of Cognitive Load to Improve Training*. Paper presented at the Proceedings of the 5th International Conference on Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience: Held as Part of HCI International 2009.

Schmorrow, D., & Stanney, K. M. (2008). *Augmented cognition: A practitioner's guide*. Santa Monica, CA: Human Factors and Ergonomics Society.

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation Space and Environmental Medicine, 78*(5 II).

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T.-P. (2001). Eye Activity Correlates of Workload during a Visuospatial Memory Task. *Human Factors, 43*(1), 111.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology, 42*(3), 323-342.

Wilson, G. F. (2005). *Operator functional state assessment for adaptive automation implementation*. Paper presented at the Proceedings of SPIE - The International Society for Optical Engineering.

Wilson, G. F., & Russell, C. A. (2003). Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors, 45*(4), 635-643.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors, 49*(6), 1005-1018.