Ongoing Efforts towards Developing a Physiologically Driven Training System

Joseph Coyne[1], Ciara Sibley[1], and Carryl Baldwin[2]

[1] Naval Research Laboratory,
4555 Overlook Ave SW, Washington, DC, USA
[2] George Mason University,
Fairfax, VA, USA
{Joseph.Coyne,Ciara.Sibley}@nrl.navy.mil, cbalwi4@gmu.edu

Abstract. There have been a number of successes of real-time application of physiological measures in operational environments such as with the control of remotely piloted vehicles (RPV). More recently, similar techniques have been investigated within the context of improving learning. A major challenge of the learning environment is that an individual's ability to perform the task, and thus their workload experienced during the task, are constantly changing. Cognitive Load Theory provides insight into how workload interacts with learning. One aspect of this theory is that as information is learned it reduces working memory demands. This paper discusses results from an RPV training study investigating the effects of workload and learning on pupil diameter. Specifically, pupil diameter decreased overtime as the task difficulty was held constant, and increased as new information was presented. The results of these studies are discussed in terms of how they can be used in a physiologically driven adaptive training system.

Keywords: Augmented Cognition, Pupil Diameter, Training, Workload.

1 Introduction

The introduction of the terms Neuroergonomics [1] and Augmented Cognition in recent years has signaled a renewed interest in applying measures of the brain to improving performance at work. Neuroergonomics refers to an interdisciplinary area of research where findings and techniques from neuroscience are used to better understand how the brain functions at work. The goal of this field is to create a better work environment which is informed by neuroscience. Augmented Cognition also has an emphasis on applying neuroscience to work but has been more specifically focused on using neurophysiological signals as inputs for closed loop systems. The goal of these closed loop systems is to detect when individuals are overloaded or underloaded and adjust their environment accordingly. The ability to create a close looped system, however requires that we have metrics that are sensitive to changes in cognitive load in near real-time. There have been a number of different types of sensors used with EEG being among the most popular due to its high temporal resolution.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 404–412, 2011.

1.1 EEG and Workload

EEG has been one of the most extensively used real-time physiological measures of mental workload and several techniques have been emerging which show promise. Pope et al [2] developed a function utilizing a simple algorithmic formula based upon Beta power divided by Alpha plus Theta Power. This formula called the Engagement index has been applied to an adaptive closed loop simulated piloting task [3] called the MAT-B and was demonstrated to improve performance in a vigilance task. The formula does not require any individual calibration which likely reduces its sensitivity but increases its ease of use.

Advanced Brain Monitoring's (ABM) B-Alert system [4] uses a proprietary method of classifying both workload and engagement based upon discriminant function analysis. The workload index is reported to track processes that generally considered to be executive functioning whereas engagement is associated with more attentional and sensory processing resources. The system applies its classification to 1 second segments of EEG data. The ABM system requires EEG data be collected from participants on a simple vigilance task over 15 minutes to establish individualized coefficients for the DFA. The metrics have been demonstrated to successfully track workload in digit span and mental arithmetic tasks [4], however others have found it to be insensitive to more complex spatial processing tasks [5].

Ongoing research at Wright Patterson Air Force Base has applied Artificial Neural Networks (ANN) to classify high and low workload in a simulated UAV task [6, 7]. The ANN incorporates EEG and a number of other physiological signals including ECG and Pupillometry. Once the ANN has been trained using physiological data collected on segments of high and low workload on the task it can accurately assess high and low workload on untrained segments with 85-90% accuracy. However, the ANN's accuracy in distinguishing high and low workload drops when used on different days and different tasks which restricts its potential applicability to the use in operational settings.

Additionally there has been some recent success in using single trial Event Related Potentials (ERP's) [8]. ERPs are typically averaged across a number of trials and components such as P300 (a peak in amplitude occurring around 300ms after a stimulus has been presented) have been found to distinguish low and high workload. Moving to a single trial analysis increases the amount of noise but also increases the potential applicability of ERP as a real-time operational metrics of workload. However single trial ERP is still developing and its application requires precise millisecond timing between the task environment and EEG system. Although the approaches to classifying EEG data vary they have definitely shown promise as a means of assessing mental workload in real-time. However there are additional time costs to using EEG in an operational environment such as set up time and time to train the classification algorithms which may limit its operational utility. Other physiological metrics such as pupillometry also show promise as a potential real-time metric of cognitive load and require less set up time.

1.2 Pupillometry and Workload

Although not a direct measure of brain activity, pupil diameter has a long history of being tied to different cognitive processes. An association has been found between increased pupil dilation and activation of the middle frontal gyrus, which has been associated with central executive and high demand functions [9]. Pupil diameter has been shown to steadily increase as workload or working memory demands increase in a large variety of different simple and complex tasks [5, 10-13]. Averaged pupil diameter has been found to be sensitive to multiple levels of workload through increasing levels of difficulty of the task. There is still some conflicting evidence however about what happens to pupil diameter during overloaded conditions, as some evidence suggests pupil diameter levels off [12] while others have actually found pupil diameter to drop when task demands exceed available resources [10].

Additionally, pupil diameter has been found to be linked with fatigue by pupil size decreasing over the course of experimental sessions [11] as well as with motivation, with individuals demonstrating larger pupil diameters when an incentive to perform is provided [13]. While pupil diameter is typically averaged and used as a post hoc assessment of workload, there have been several investigators looking at applying pupil diameter to real time or over smaller time windows [14]. Marshall's index of cognitive activity [14] is a real time gauge of workload based upon applying a proprietary wavelet analysis to pupil diameter. Other researchers [5] have used average pupil diameter and maximum pupil diameter over short time windows.

Pupillometry has the potential to be a valuable index of mental workload in an operational setting. Although pupil diameter also varies with other more tonic psychological phenomenon such as fatigue and incentive, there is still evidence to suggest that it can still detect more short term changes in workload while these other phenomena are occurring [5, 13]. Technology for eye tracking systems has improved such that they are now completely off the head and require less than a minute to calibrate to an individual. These advancements in eye tracking systems make it more viable to investigate questions about pupillometry under various conditions.

1.3 Closed Loop Physiological Systems

Real-time adaptive automation involves the allocation of responsibilities or aiding to a human-machine system during a task and based on an input metric (EEG, pupillometry, performance) that is being analyzed in real-time. In recent years, researchers at Wright Patterson Air Force Base have had success using artificial neural networks (ANN) to classify operator state with accuracies up to 85% for an individual [6, 7].The researchers fed EEG and eye tracking data in real-time into an ANN and showed a 50% improvement in performance on a RPV target identification task when using adaptive automation in the task to slow down vehicle speed when workload was classified as high, and speeding it up when the ANN classified workload as low. The research serves as one of the best examples of how real-time physiological assessment and monitoring can be used to improve performance in an operational environment. The success has fueled interest in applying physiological sensors as real-time inputs to other closed loop systems and moving it from the operational domain to a training domain.

1.4 Applying Physiology to Learning

The goal within the operational environment is to identify periods of time where the operator is overloaded and then provide automation or aiding to reduce workload. The training environment however is dynamic and as trainees acquire skills, they utilize different brain regions and demonstrate reduced workload while performing the same task [15]. Cognitive Load Theory (CLT) [16] is a theory of learning which states that learning is essentially, processing and organizing information in working memory and storing it in long term memory. The organization of information in working memory is a cognitively demanding process and the largest bottleneck in the learning process. Overloading an individual's working memory capacity results in an inability to transfer information to long term memory and thus increased time to learn the task. Once information is learned and stored in long term memory the demands on working memory (cognitive load) are greatly reduced. The goal of adaptive training would therefore be to identify when an individual has learned a specific level of a task or information and then present them with additional information to learn.

Presently there are no systems that monitor task knowledge in real-time via physiology and then adapt training material. One approach to measuring skill mastery is being performed by researchers at EGI, where they are investigating the use of single trial event related potentials in specific brain regions during a language learning task. The present research addresses an alternative approach, measuring working memory load or workload in real-time with pupillometry, as an individual learns a task.

As an individual acquires knowledge (i.e., stores it in long term memory), they become less reliant on working memory and eventually the task becomes automatic. Therefore an individual who is learning a task should demonstrate both higher performance and lower workload once they have mastered the task. The present experiment trained individuals on a simulated RPV intelligence, surveillance and reconnaissance (ISR) task. Trainees had to calculate the vehicle's direction of movement based upon the UAV heading and the apparent target direction of travel on the simulated video feed. Skill level in the task was manipulated by increasing the amount of mental rotation necessary to calculate direction of travel.

2 Method

2.1 Participants

Thirteen participants (18-30 years, M = 22, SD = 4.10) from the George Mason University volunteered to engage in a UAV training simulation experiment in exchange for course credit. Four participants' data were excluded due to problems with missing data.

2.2 Materials

Virtual Battlespace 2 (VBS2) was used to construct simulated UAV video files that were played back in a separate application created for this experiment. VBS2 is a high-fidelity, three-dimensional virtual training system used for experimental and military training exercises. In addition, the Tobii X120 off the head unit was used to collect eye tracking data. The unit sat in front of the participant and just below the surface of the monitor running the simulation. The system recorded both eyes at 60 samples per second. Neuroscan was used to collect EEG data at 500 samples per second. EEG data however is not considered in this paper.

2.3 UAV Desktop Simulation

After receiving a brief PowerPoint training about the task, participants engaged in a UAV desktop simulation in which he or she was trained to report information on moving vehicles as seen from a UAV (see Image 1). Participants were asked to identify and report heading information about the target vehicles crossing the screen. At the beginning of each experimental block, examples of each target to identify were presented and participants were expected to learn to recognize each by name (ID task). An example of one of the vehicles is shown in Image 2. For each experimental trial, participants were given the heading of the UAV and were asked to estimate the heading of the vehicle on the ground (heading task). 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 and the identity of the target, participants were then asked to rate their perceived mental effort in calculating the target heading and identifying the target.

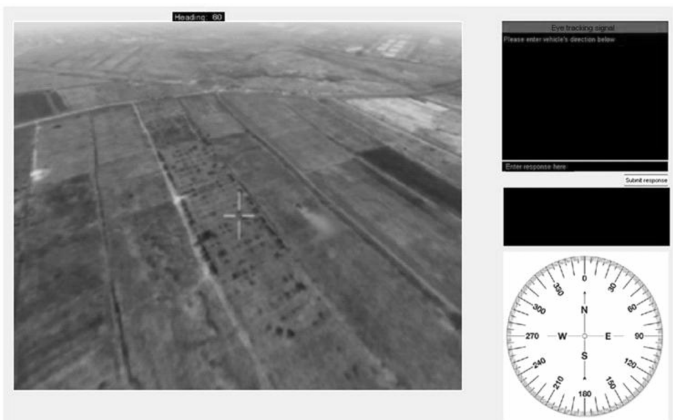

Fig. 1. Interface for the experiment

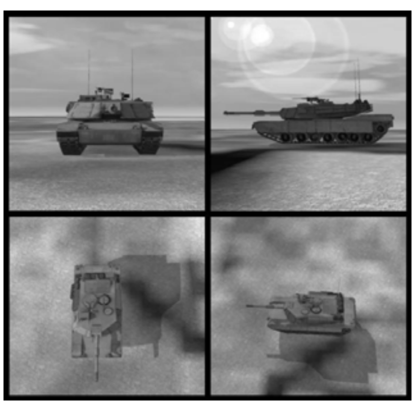

Fig. 2. Picture of an M1A1 that participants were expected to recognize by name

The difficulty of the task progressed over three blocks of trials. Only one vehicle was shown on the screen at a time and a total of twenty vehicles were shown within each block. Difficulty was manipulated by varying the UAV heading as well as the possible target heading. Additionally, the ID task difficulty was manipulated by increasing the number of vehicles the participant learned for each block from two to four to six. 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. During this level, participants were only expected to learn two vehicles.

The most difficult level (block three) showed the UAV heading at any 30 degree increment and changed after each target, and the target heading could also be any 30 degree increment. During this final block, participants were presented with six vehicles and were expected to recognize all six of them by name.

2.4 The Experiment

Written informed consent was obtained from all participants and the participants were then introduced to the experimental tasks. This experiment took place over one day with a duration of approximately two to three hours, including EEG preparation, eye tracking calibration, training and experimental trials. After being prepped for EEG, participants reviewed a PowerPoint training on the task and then began the experiment. All participants completed blocks one, two and three in the same order from easiest to hardest.

3 Results

The researchers considered one of the most cognitively demanding parts of the simulation to be during the target heading calculation. Therefore, this analysis primarily focused on the pupillometry data during this task. This was done in order to compare pupil dilation during the high demand parts of the task to other less demanding parts. Additionally, in order to investigate the affect of learning on pupil dilation, the first and last three trials of each difficulty block were analyzed and compared to each other. We hypothesized that as the participant began to learn the material, it would become less challenging throughout that difficulty block, and that his/her pupil size would get closer to baseline levels towards the end of the block.

3.1 Maximum Pupil Size and Difference Scores

To best capture the period of mental effort during the heading calculation task, we took the maximum pupil size (an average of the top five pupil sizes) during each part of the task. Figure one depicts the maximum pupil sizes during the first and last three trials of each block during the ID and heading task. According to performance data, participants focused their efforts on the heading task, which is confirmed in the pupillometry since pupil size is greater during the heading task than it is during the ID task; both of which are within seconds of each other.

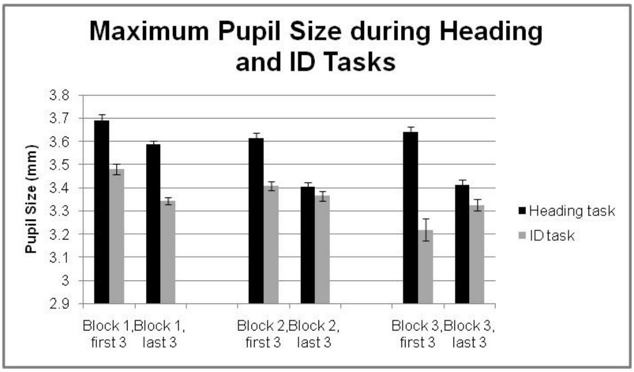

Fig. 3. Pupil dilation is sensitive to mental effort and is highest during the most demanding part of the task

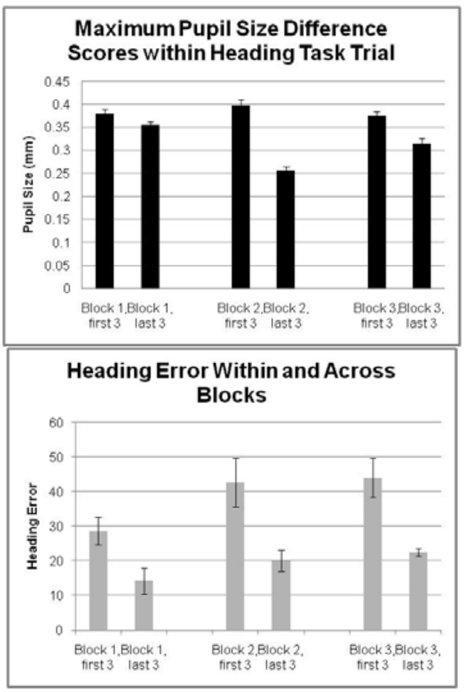

Fig. 4. Maximum pupil difference size is highest at the beginning of the new block of difficulty, and then attenuates with learning (left figure). This trend follows with the performance data (right figure) showing participants' error reducing by the end of each block.

Difference scores were then calculated for pupil dilation by calculating each individual's average pupil size during each individual trial and subtracting that from the maximum pupil size data. This was done in order to show change from the average and reduce factors caused by individual variability. Positive scores mean that the pupil size was greater than average, while negative scores indicate pupil size was smaller than average. Figure 2 (left) shows not only that pupil size is far above average (average being zero), but also that that pupil size attenuates towards the end of each block, but then jumps back up at the beginning of the next difficulty block. This pattern also corresponds with performance data (figure 2, right) where the participants' performance becomes significantly better towards the end of each block.

A paired-samples t-test was conducted to compare the pupil size for the first three compared to the last three trials across each block. There was a significant difference in the pupil sizes in the first part of the block ($M=0.38$, $SD= 0.16$) compared to the second part of the block ($M=0.31$, $SD= 0.14$), $t(45)= 2.40$, $p= 0.01$.

4 Discussion and Implications

The results of the present study demonstrate that pupil diameter is sensitive to both phasic and tonic changes of workload. Phasic, short term sensitivity is evidenced by large increases in pupil diameter as individuals moved from the search phase (easiest portion of the task) to the mental calculation phase (hardest phase). Pupil diameter was also sensitive to tonic changes in workload as evidenced by the gradual decrease in pupil diameter from the beginning of the block to the end of the block. The fluctuations of pupil diameter during the heading task between and across blocks matches what was expected based upon Cognitive Load Theory. That is, within each block of difficulty pupil diameter significantly decreased when comparing the beginning of the block to the end of the block. These changes in workload also corresponded with increases in performance from the beginning of the block to the end of the block. The combined increase in performance and decrease in workload suggests that information was transferred to long term memory and the burden on working memory was reduced i.e., information was being learned. Additionally, as new information was presented e.g., from block 1 to block 2 pupil diameter once again significantly increased.

Although the present study was not a closed loop adaptive training, it did provide evidence to suggest that pupil diameter could be used to drive such a system. Unlike the operational environment, developing a closed loop training system has many unique challenges beyond the simple identification of sensitive metrics; although even within metric selection, learning presents some unique challenges. For example, how do you train an ANN when workload fluctuates as a function of learning? Identifying high and low workload may need to be done with a different task which may impact the sensitivity of the ANN. This is in contrast to the operational environment, where we simply identify periods of overload and turn some additional automation on, or identify periods of underload and turn some automation off. Changes in learning appear to be more gradual and present challenges that need to be investigated. For example, determining what threshold level suggests an individual is ready to learn new material is something that still needs to be identified. This threshold may also change as the goal of learning shifts from acceptable performance to retention. Future studies will investigate these and other questions with the intention of eventually closing the loop in a training environment.

Acknowledgements. This work was funded by the Office of Naval Research's Human Performance Training and Education Program Grant number: N0001410WX20539.

References

1. Parasuraman, R.: Neuroergonomics: Research and Practice. Theoretical Issues in Ergonomics Science 4, 5–20 (2003)
2. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology 40, 187–195 (1995)
3. Freeman, F.G., Mikulka, P.J., Prinzel, L.J., Scerbo, M.W.: Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. Biological Psyhology 50, 61–76 (1999)
4. Berka, C., Levendowski, D., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviation, Space, and Environmental Medicine 78, B231–B244 (2007)
5. Sibley, C., Coyne, J.T., Cole, A., Gibson, G., Baldwin, C.L., Roberts, D., Barrow, J.: Adaptive training in an Unmanned Aerial Vehicle: Examination of several candidate real-time metrics. In: Karwowski, W., Salvendy, G. (eds.) Applied Human Factors and Ergonomics, Taylor & Francis, Boca Raton (2010)
6. Wilson, G.F., Russell, C.A.: Real-Time Assessment of Mental Workload Using Physiological Measures and Artifiicial Neural Networks. Human Factors 45, 635–643 (2003)
7. Wilson, G.F., Russell, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Human Factors 49, 1005–1018 (2007)
8. De Lucia, M., Michel, C.M., Murray, M.M.: Comparing ICA-based and single-trial topographic ERP analyses. Brain Topography 23, 119–127 (2010)
9. Spinks, J.A., Zhang, J.X., Fox, P.T., Gao, J.H., Tan, L.H.: More workload on the central executive of working memory, less attention capture by novel visual distracters: Evidence from an fMRI Study. NeuroImage 23, 517–524 (2004)
10. Poock, G.K.: Information processing vs. pupil diameter. Perceptual & Motor Skills 37, 1000–1002 (1973)
11. Khaneman, D., Peavler, W.S.: Incentive effects and pupillary changes in association with learning. Journal of Experimental Psychology 79, 187–196 (1969)
12. Peavler, W.S., McLaughlin, J.P.: Pupil size, information overload, and performance differences. Psychophysiology 11 (1974)
13. Heitz, R.P., Schrock, J.C., Payne, T.W., Engle, R.W.: Effects of incentive on working memory capacity: Behavioral and pupillometric data. Psychophysiology 45, 119–129 (2008)
14. Marshall, S.P.: Identifying cognitive state from eye metrics. Aviation, Space, and Environmental Medicine 78, B165–B175 (2007)
15. Luu, P., Shane, M., Pratt, N.L., Tucker, D.M.: Corticolimbic mechanisms in the control of trial and error learning. Brain Research 1247, 100–113 (2009)
16. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science: A Multidisciplinary Journal 12, 257–285 (1988)