

EEG and Eye-Tracking Based Measures for Enhanced Training*

Walid Soussou, Michael Rooksby, Charles Forty, James Weatherhead, and Sandra Marshall

Abstract— This paper describes a project whose goal was to establish the feasibility of using unobtrusive cognitive assessment methodologies in order to optimize efficiency and expediency of training. QUASAR, EyeTracking, Inc. (ETI), and Safe Passage International (SPI), teamed to demonstrate correlation between EEG and eye-tracking based cognitive workload, performance assessment and subject expertise on X-Ray screening tasks.

Results indicate significant correlation between cognitive workload metrics based on EEG and eye-tracking measurements recorded during a simulated baggage screening task and subject expertise and error rates in that same task. These results suggest that cognitive monitoring could be useful in improving training efficiency by enabling training paradigms that adapts to increasing expertise.

I. INTRODUCTION

The goal of learning is to transform novices to experts. Training transitions students from a state of work that is deliberate, monitored, and emotional to a state of work that is effortless and natural, but not thoughtless or accidental. In order to evaluate expertise, tests are generally administered to assess performance, such as whether students are able to complete tasks successfully, or whether they can generalize the rules learned. However, such performance-based metrics are currently not able to assess the ease with which the task was completed, a measure that can help distinguish expertise level.

Learning models suggest that monitoring attention load and relating it to performance could help determine expertise levels. Accurate assessment of progression of expertise could improve the efficiency of training paradigms. Physiological signals such as electroencephalography (EEG) and pupillometry have been reported to provide useful measures of cognitive workload. In recent years, Quantum Applied Science & Research (QUASAR) has developed novel dry EEG sensors that can be used in classroom, workstation, and airport contexts. Likewise, eye tracking systems can be mounted near monitors or incorporated into headsets, enabling practical deployment. QUASAR and EyeTracking, Inc. (ETI), a market leader in eye-tracking analysis, have developed gauges based on these deployable sensor technologies for unobtrusively assessing and monitoring attention, cognitive workload, and fatigue.

* Research supported by the Department of Homeland Security. The content does not necessarily reflect the position or policy of the Government and no endorsement should be inferred.

W. Soussou is with Quantum Applied Science and Research (QUASAR), CA 92121 USA. (phone: +1-858-412-1820; fax: +1-858-412-1911; email: walid@quasarusa.com).

M. Rooksby and C. Forty are with Safe-Passage International (SPI), NY 14623 (email: mrooksby@safe-passage.com, cforty@safe-passage.com)

J. Weatherhead and S. Marshall are with Eye-Tracking Inc (ETI), CA 92120 USA (jweatherhead@eyetracking.com, smarshall@eyetracking.com)

It is hypothesized that monitoring cognitive effort during training could be used to elucidate the progression towards expertise. In order to test the hypothesis QUASAR and ETI, teamed with Safe Passage International (SPI), a leading provider of Transportation Security Officers (TSOs) screener training and developer of training software for security applications.

A key portion of the TSOs' role is X-Ray screening, which is a repetitive visual search task that often has a very low probability of encountering a threat, but extremely high consequences if a serious threat is missed. The nature of this task frequently induces fatigue, boredom and distraction. Similar factors are present in other repetitive screening jobs such as radiologists screening medical X-ray or MRI or other imaging, cytologists evaluating microscopic samples, or intelligence analysts searching through satellite imagery. For TSO purposes, training efficiency and effectiveness is of special importance due a very high attrition rate, which therefore requires an effective way to bring new TSOs up to proficiency. Current TSO training constitutes of an initial training of over 100 hours, and continuing training of at least three hours/week.

In this paper, the team describes the use of these cognitive gauges to assess mental workload during performance of a screening task and compare novice and expert screeners. The long-term goal of this effort is the development of a training system for real-time measurements of physiologic responses, in order to provide a customized training experience/environment that is optimally suited to each specific trainee.

II. BACKGROUND

A. Brain-Based Signals in Training Environments

Over the last decade, several researchers have started evaluating or incorporating brain-based signals in training and learning environments. EEG has been used to monitor the progress of trainees through skill levels or identify indices of skill acquisition. One group reported an increase in event-related alpha power that correlated with amount of practice at a shooting task and suggested that it reflected a decrease in cortical activity associated with reduced effort required with expertise. [1] Another group observed lower coherence associated with less cortico-cortical communication in expert marksmen compared to skilled shooters, and attributed this difference to decreased involvement of cognition with expertise. [2] Similar results have been reported with fMRI monitoring.

Eye Tracking research suggests that experts generally demonstrate more focused attention on a task than do novices [3][4]. Thus, it is to be expected that expert screeners will have more deliberate viewing patterns and exhibit fewer lengthy saccades than novice screeners. Moreover, they

should show higher rates of vergence as a result of their focused attention, and develop efficient scanning patterns in much the same way that expert pilots demonstrate efficient scanning of their instrument panel. Novices are expected to show more random scanning behavior, including multiple revisits to objects that are not yet fully identified.

B. QUASAR's Dry EEG Sensors and Signal Quality

Scientists at QUASAR have developed a new revolutionary through-hair hybrid EEG sensor (Figure 1A). [5][6]. These innovative bioelectrodes use a combination of high impedance resistive and capacitive contact to the subject, which produces adequate sensitivity and bandwidth for EEG measurement. The sensors can be applied with light, comfortable pressure and record EEG for practically unlimited durations without a need for skin preparation or electrolytes of any kind.[7] QUASAR has developed a headset based on the sensor technology. (Figure 1B)

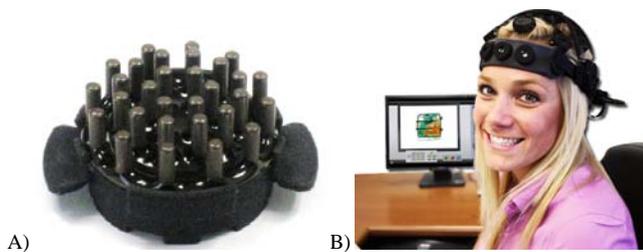


Figure 1. A) Hybrid sensors for through-hair measurement of EEG B) QUASAR headset on subject.

These hybrid sensors have been tested in a variety of contexts and proven to record EEG signals with high fidelity. [5][6] And recently, Air Force researchers evaluated QUASAR's EEG system and reported "*results confirm that the data collected by the new [QUASAR] system is comparable to conventional wet technology.*" [8]

C. EEG Derived Cognitive Gauges

Research during DARPA's Augmented Cognition program suggested that changes in alertness are a principal component of variance in the EEG spectrum and several investigators demonstrated the use of spectral changes for accurate estimation of alertness and cognitive workload, [9] and cognitive fatigue [10]. Furthermore, a number of studies have reported that theta is related to increases in attention, workload, memory load, and working memory performance, and that a large increase in alpha EEG precedes dozing off during a simple visual task. [11]

QUASAR has developed several new algorithms for physiological signal analysis. These efforts include development and implementation of noise and artifact detection, identification and reduction methods, as well as cognitive and physiological state classification. Additionally, we have developed Partial-Least-Square (PLS)-based classification algorithms that enable determination of cognitive and physiological states from EEG and ECG data. [12] Briefly, this learning algorithm extracts spectral features from the EEG signal, trains cognitive models based on the researcher's interests (attention, workload, fatigue ...), then processes EEG data in real time producing cognitive state measures whose output ranges from 1 to 100 representing the relative intensity of the monitored state. The algorithm allows

for expedient subject specific calibration within minutes, or the creation of "normative" models based on EEG features reported in the literature to correlate with the state of interest.

QUASAR has tested and validated this cognitive gauge methodology on several research projects, with an emphasis on states of engagement, workload, and fatigue. All three state models regularly achieve average classifications accuracies >90%, as determined by performance on primary (game), and secondary (auditory N-Back) tasks, primary task difficulty, subjective evaluation (NASA TLX), and time duration since last sleep (for fatigue). [13]

D. Eye-tracking Based Measures of Cognitive State

Several measures of eye tracking data can be combined to assess inattention, fatigue, alertness, and overload. The measures are: Index of Cognitive Activity (ICA, a pupil dilation based metric), fixation data, blink data, and vergence data. As an example, ETI applied this procedure to a fatigue study in which individuals were sleep deprived. It was possible to detect fatigue before the individual began to reach the eyelid closure levels required in PERCLOS (which is based on the percent of time the eye is closed over a full minute). The patented procedures developed by ETI require only a few seconds to detect fatigue and have demonstrated reliability in settings such as automobile driving, visual search, and lunar search and recovery simulations. The general technique is described in [14] One common use of the cognitive state technology is to examine tasks of varying difficulty. These procedures correctly identify easy, moderate, and difficult tasks based only on the results of the eye metric. Typically, 80-95% of all seconds of a task can be correctly classified in terms of task difficulty.

III. EXPERIMENTAL PROCEDURE

A. System Synchronization

1) Temporal alignment

It was critical to time align all data in order to analyze relations between them. Three data sources needed to be synchronized for this project

1. EEG data sampled at 240 Hz and its derived gauges integrating over 2-second epochs.
2. Eye-tracking data sampled at 60 Hz, and its derived gauges and metrics calculated at 2-second epochs in order to conform to the EEG measurements.
3. X-Ray simulator task events

Two methods for synchronizing the various components of the system were used: 1) The eye-tracking system sent the EEG systems sent synchronization triggers. These triggers allowed alignment of the collected data. 2) In addition, in order for all three software systems to share a same computer clock to reference their data collection, we installed them on one laptop with an Intel core i7 processor. All three systems operated normally and did not interfere with each other's performance. Comparing the timestamps of the triggers and event markers, we were able to align all the data files.

2) Spatial Alignment

The boundaries of threat objects in the X-Ray images are defined in SPI's data base as x-y coordinates of polygons. ETI's proprietary software places boundaries around objects

in order to calculate several eye-tracking parameters such as fixation duration, number of revisits or of fixations. In order to determine eye-tracking metrics related to threat or distractor objects, it is therefore important that the spatial coordinates of the objects on the screen are aligned to the spatial coordinates of the eye-tracker. ETI calculated the proper conversion calculation between the two coordinate systems, and verified that identified areas mapped to threat objects on the images.

3) Cognitive State Calibration

QUASAR's EEG based cognitive state gauges use data recorded during calibration tasks to train models that are later used to determine cognitive states. Each task is conducted at the beginning and end of the experiment. For these calibrations, QUASAR uses a public domain battery of psychological tests (Psychology Experiment Building Language) to select the most appropriate task from a set of standardized cognitive tests. For this work, we selected tasks that induce varying levels of cognitive load related to visual discrimination tasks and vigilance. Five tasks were set up to provide baseline conditions:

1. Eyes Open (EO) and Eyes Closed (EC)
2. 3 and 5 grid Matrix Rotation 3 (M3, M5) Task
3. Visual Detection (VD) Task
4. Daydream (DD) Task, where subjects were to asked daydream for 1.5 mins
5. Psychomotor Vigilance (PV) Task

The EEG data collected during these tasks were then used by QUASAR's PLS algorithm to train cognitive workload algorithms individualized for each subject.

B. Experimental Protocol

Two groups of subjects were recruited to conduct the experimental protocol. "Novice" subjects were adults who passed color blindness tests and a ScanX (Leaderwear) software test that is commonly used to identify suitability for X-Ray screening task, and were then recruited to undergo 16 hours of computer-based X-ray baggage screening training developed by collaborator SPI. (Figure 2) "Expert" subjects were experienced TSOs with at least 2 years of experience provided to the project by the sponsor.



Figure 2. Screenshot of SPI X-ray screening computer-based training software. The software emulates X-ray screening equipment functions, including color filter layers.

Subjects underwent the following protocol:

1. Overview of the experiment, introduction to the EEG headset and eye-tracker, and informed consent.
2. 5-minute SPI tutorial followed by 5 minutes of practice

3. Calibration of eye-tracking device for each subject
4. QUASAR researchers placed headsets on subjects and adjusted sensors to ensure proper signal quality as ascertained by impedance monitoring
5. Perform EEG model calibration tasks
6. Perform 2 X-ray screening tests where the objective is to analyze 100 images and identify threats, if any. The test was conducted on SPI's X-ray scanner emulator. A 3rd test was performed after a 15-minute break. There was a 25:75 threat to non-threat ratio in the images.
7. After each test, headset comfort and task difficulty surveys were conducted
8. At the end of the testing the EEG calibration tasks were performed again before taking off headset.

IV. RESULTS

1) Performance on X-Ray Screening Task.

While one Novice subject underperformed compared to the others on the test, the other novice subject performed comparably to the Expert subjects, both in overall score (Figure 3), in response time, by error type, and when breaking down questions by image difficulty and threat condition, as well as subjective surveys.

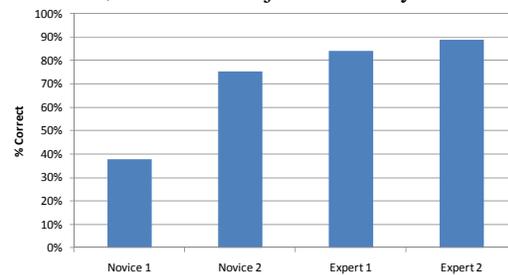


Figure 3. Average performance on the X-Ray Screening task by subject.

2) Cognitive Workload Assessment.

When examining cognitive workload, however, the average workload was significantly higher for the Novice group 72.6% (SEM ± 0.6) and than for the Experts 33.0% (SEM ± 0.8) (Student's T-test $p < 0.05$) for the EEG-based gauges, as well as for the pupillometry based ICA where the average for the Novice group was 0.41, and 0.33 for Experts. Cognitive workload during the calibration tasks was however comparable between the two groups.

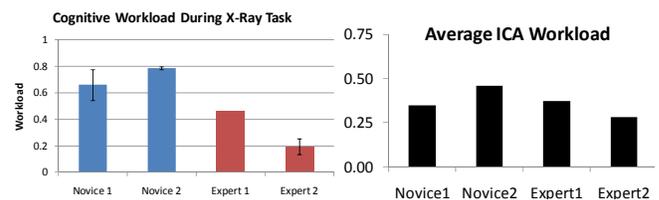


Figure 4. EEG based Cognitive Workload during X-Ray Task (EEG based gauge and Eye-Tracking based ICA)

3) Relationships between Performance and BBM

The relationship between brain-based cognitive gauges and performance on the X-Ray screening task was examined by cross-correlations. First, for each X-Ray screening image

in experimental sets 1 and 2, the following features were calculated for each subject:

1. WKL: EEG-based Cognitive Workload gauge output averaged across the screening time for each image
2. ATT: EEG-based Cognitive Attentiveness gauge output averaged across the screening time for each image
3. ICA: Eye-Tracking based Index of Cognitive Activity output averaged across the screening time for each image
4. EXP: Expertise level of the subject examining the image
5. COR: Correct decision made for the image
6. ERR: Erroneous decision made for the image
7. TNR: True Negative Response (correct no threat)
8. FPR: False Positive Response (incorrect threat)
9. FNR: False Negative Response (incorrect no threat)
10. TPR: True Positive Response (correct threat)
11. DIF: Difficulty level reported for image by SPI experts
12. DUR: Duration of time subject spent examining image

The above 12 features were calculated for each of the test images and were cross correlated. Figure 5 plots the pairwise Pearson's correlation coefficients for these features, and illustrates the relationship between the cognitive gauge output and task performance.

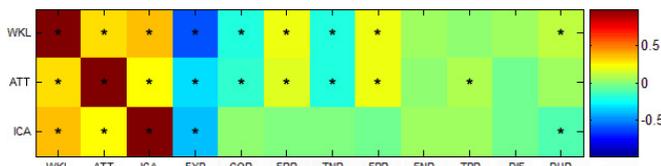


Figure 5. Pearson's correlation coefficients between cognitive workload gauges and performance metrics. Starred squares indicate statistically significant correlations.

First, all three cognitive gauges are significantly correlated to each other, with the EEG-based and eye-tracking-based workload gauges having a 0.35 correlation coefficient, and ATT having correlation coefficients of 0.30 and 0.24 to WKL and ICA respectively.

Second, **all three cognitive gauges are significantly correlated with expertise levels.** The negative indices indicate that lower cognitive workloads are associated with higher expertise levels. This means, that overall, during screening of X-Ray images, an expert will have a lower cognitive workload than a novice.

Third, the EEG-based cognitive gauges are significantly correlated to correct and erroneous decisions. Correct decisions are linked with lower mental efforts while incorrect decisions are associated with higher mental efforts. Specifically, the relationship between cognitive load and error is due to True Negative (TNR) decisions requiring less mental strain than False Positive (FPR) errors. False Negative (FNR) errors are not significantly correlated to any mental gauge, but True Positive Responses (TPR) are significantly correlated with the EEG-based cognitive Attentiveness gauge.

None of the cognitive gauges are significantly correlated to the image difficulty, but both EEG-based and Eye-Tracking-based workload gauges are significantly correlated to the duration of time spent screening an image.

V. CONCLUSIONS

These preliminary yet statistically significant correlations between cognitive workload measures, performance, and expertise level suggest EEG and Eye-Tracking based cognitive metrics can be useful in training environments. New training paradigms could utilize this information to adaptively modify training content with increasing expertise levels, thereby maximizing training efficiency.

REFERENCES

- [1] Kerick SE, Douglass LW, Hatfield BD (2004) Cerebral cortical adaptations associated with visuomotor practice. *Med Sci Sports Exerc* 36:118-129.
- [2] Deeny SP, Hillman CH, Janelle C, Hatfield, B. (2003) Cortico-cortical communication and superior performance in skilled marksmen: An EEG coherence analysis. *Journal of sport and exercise psychology* 25:188-204.
- [3] Ericsson KA, Ward, P.C. (2007) Capturing the naturally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. *Current Directions in Psychological Scienc* 16:346-350.
- [4] Janelle C, Hatfield, B. (2008) Visual Attention and Brain Processes that Underlie Expert Performance: Implications for Sport and Military Psychology. *Mil. Psych.* 20:S39-S69.
- [5] Matthews R, McDonald NJ, Anumula H, Trejo LJ (2006) Novel hybrid sensors for unobtrusive recording of human biopotentials. In: *Proceedings of 2nd Annual Augmented Cognition International Conference*. San Francisco.
- [6] Matthews R, McDonald NJ, Fridman I, Hervieux P, Nielsen T (2005) The invisible electrode – zero prep time, ultra low capacitive sensing. In: *11th International Conference on Human Computer Interaction*. Las Vegas.
- [7] Matthews R, McDonald NJ, Hervieux P, Turner PJ, Steindorf MA (2007) A wearable physiological sensor suite for unobtrusive monitoring of physiological and cognitive state. *Proceedings of the 29th Annual International Conference of the IEEE Engineering Medicine Biological Society* 2007:5276-5281.
- [8] Estep JR, Christensen JC, Monnin JW, Davis IM, Wilson GF (2009) Validation of a Dry Electrode System for EEG. In: *Proceedings of the Human Factors and Ergonomics Society*, pp 1171-1175. San Antonio, Texas.
- [9] Makeig S, Jung TP (1995) Changes in alertness are a principal component of variance in the EEG spectrum. *Neuroreport* 7:213-216.
- [10] Trejo LJ, Kochavi, R., Kubitz, K., Montgomery, L., Rosipal, R., & Matthews, B (2004) Measures and models for estimating and predicting cognitive fatigue. In: *44th Annual Meeting of the Society for Psychophysiological Research*. Santa Fe, NM.
- [11] Torsvall L, Akerstedt T (1988) Extreme sleepiness: quantification of EOG and spectral EEG parameters. *Int J Neurosci* 38:435-441.
- [12] N. J. McDonald and W. Soussou, "QUASAR's QStates Cognitive Gauge Performance in the Cognitive State Assessment Competition 2011," presented at 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts USA, 2011
- [13] Matthews R, Turner P, McDonald NJ, Ermolaev K, McManus T, Shelby R, Steindorf M (2008) Real Time Workload Classification from an Ambulatory Wireless EEG System Using Hybrid EEG Electrodes. In: *30th Annual International IEEE EMBS Conference*. Vancouver.
- [14] Marshall SP (2007) Identifying cognitive state from eye metrics. *Aviat Space Environ Med* 78:B165-175.