ITEA Journal 2010; 31: 275–279

Augmenting Test and Evaluation Assessments Using Eye-Tracking and Electroencephalography

Anthony Ries, Ph.D., and Jean Vettel, Ph.D.
U.S. Army Research Laboratory,
Human Research and Engineering Directorate,
Translational Neuroscience Branch, Aberdeen Proving Ground, Maryland

Tools that provide continuous, objective measurements of human-system interactions can augment measures obtained through subjective assessments and/or expert observation by providing near–real time performance metrics. Two tools for the Test and Evaluation (T&E) community will be discussed: eye-tracking applications that are viable for use in T&E today and electroencephalography-based metrics that hold promise for the future.

Key words: Continuous measurements; Electroencephalography (EEG); eye-tracking; mental state; subjective assessments.

System evaluation would be much easier if testers could recruit the ideal operator: someone who never gets fatigued, who maintains a consistent level of concentration on the task, and who can accurately recall their moment-by-moment experience of task difficulty during the testing session. This would ensure that any performance decrements during testing were not a result of the operator being tired, not concentrating, or simply not remembering what happened at a particular point in time. Unfortunately, these operators are hard to find! Instead, evaluators must try and measure the operator's mental state in order to assess how that state affects performance during system interaction, or alternatively, how the system interaction influences mental state, which in turn affects performance. Currently, to evaluate operator mental state, evaluators must rely on self-assessment questionnaires, such as the National Aeronautics and Space Administration Task Load Index (NASA-TLX) (Hart and Staveland 1988), that interrupt the operator at discrete times throughout the testing session to provide an introspective assessment. Not only does the interruption break mental concentration on the task, but self-reports are not sensitive to fluctuations of cognitive state within a task; rather they provide an average subjective estimate over a length of time.

Self-assessment measures are not ideal because they lack objective means to measure particular mental state changes, what influenced the state change, and what happened because of the state change. Consider a scenario where the operator must perform several tasks on a new system, and a self-assessment questionnaire is administered at different intervals throughout the study (i.e., discrete measurements). The operator's fatigue level may fluctuate over the course of the study; however, since individuals must rely on memory to recall past events and are not always accurate in their self-assessment reports, discrete measurements can lead to inaccuracies when trying to capture dynamic mental state fluctuations during the test. The lack of a continuous measure can lead to errors in system evaluation, attributing the decrement in user performance to system design rather than attributing it to changes in the operator's mental state (i.e., level of fatigue). One solution to the problem of tracking performance changes over time is to simply take measurements more frequently; however, this comes with the serious disadvantages of breaking continuity of the task, not to mention the introduction of additional task complexity created by the demands of completing multiple surveys as well as performing the task itself. A more efficient solution to this problem would be the use of tools that permit continuous measurement of task performance, eliminating interruptions for self-assessments.

This article discusses how two measurement tools, eye-tracking and Electroencephalography (EEG), provide continuous measurements related to operator performance without creating task disruption. Eye-tracking provides numerous measures of user eye

Report Documentation Page					*Form Approved OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE **JUN 2010**		2. REPORT TYPE			3. DATES COVERED **00-00-2010 to 00-00-2010**
4. TITLE AND SUBTITLE **Augmenting Test and Evaluation Assessments Using Eye-Tracking and Electroencephalography**					5a. CONTRACT NUMBER
					5b. GRANT NUMBER
					5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)					5d. PROJECT NUMBER
					5e. TASK NUMBER
					5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Research Laboratory,Human Research and Engineering Directorate,Translational Neuroscience Branch,Aberdeen Proving Ground,MD,21005**					8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					10. SPONSOR/MONITOR’S ACRONYM(S)
					11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

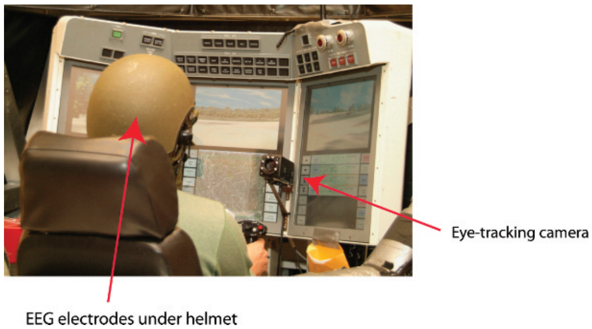

Figure 1. Integration of eye-tracking and Electroencephalography (EEG) tools during operator-system interaction. Eye-tracking and EEG provide continuous, objective measurements of operator performance.

movements and visual scanning patterns, while EEG provides a measure of electrical activity in the brain that can be linked to many complex behaviors and operator mental states. These two tools complement one another and may eventually be used effectively together in T&E environments. For example, *Figure 1* shows an operator in a futuristic crew station, wearing a helmet containing EEG electrodes to record brain activity. In addition, a camera mounted within the driving simulation continuously records operator eye measurements. Both of these tools can capture dynamic changes in the operator's mental state throughout the testing session, providing information regarding how the operator interacts with the system interface, without interrupting operator task performance. In this article, potential applications of these two tools to T&E environments will be discussed, with an emphasis on eye-tracking applications that are ready for use today, and EEG applications that provide promise for the future.

Eye-tracking in T&E

Eye-tracking is the process of measuring either the point of gaze (where the operator is looking), or the motion of the eye relative to the head. There are a number of methods for measuring eye movement, including the use of video images from which the eye position is extracted. Advances in computer and video technology have lead to the development of eye-tracking systems that are portable and simple to use (Babcock, Lipps, and Pelz 2002). Eye-tracking offers the evaluator an objective and unobtrusive means to continuously measure human performance in a diverse set of environments and field settings. After a quick calibration, eye-tracking can provide several continuous user measures that can be linked to operator mental state. These measures include blink rate, frequency of eye movement, pupil dilation, the amount of eyelid closure over time (described as Percentage Eye Closure or PERCLOS), and the length of time spent looking at a particular location. Although all of these continuous measures offer advantages, we will highlight just a few of them here to demonstrate how these measures can augment a variety of T&E scenarios today.

Figure 2 shows an example of how an eye-tracker could be used to plot the gaze path of an operator using a crew station computer interface with multiple display screens. In this example, the operator scanned several displays showing urban environments and system status, searching for images of people who could pose a threat to the security of their vehicle. The operator used the touch-screen interface on the center console to complete a threat report anytime a threatening person was seen. The red circles indicate where the operator fixed their gaze on the screen, with the numbers inside the circles indicating the order in which the operator's eyes traveled across the multiple display screens. The size of the red circles describes the relative amount of the time spent looking at each location, with larger circles representing longer gaze fixation times. As can be seen, the operator searched from left to right, starting first with the exit points on the corner building at the left-most display, continuing to the right across the center, and ending at a point between the right-most display and the status display. We can see that the operator spent time looking at the threat report console (circle 3) before scanning the other two buildings (circles 4 and 8). These continuous measures show operator scanning patterns used to perform the task, as well as which system interface features are used most frequently.

Eye-tracking measures have been used successfully to reveal differences between novice and expert operators. Ottati, Hickox, and Richter (1999) compared eye movement patterns of novice and experienced pilots in a flight simulator that required the use of electronic maps for navigation. Traditionally, pilots use printed maps for this task. However, in Ottati's task, the pilots were required to identify critical terrain navigation landmarks from electronic instrumentation. The eye-tracking data revealed search path differences between novices and experts. Unlike the experienced pilots, novice pilots gazed longer and more frequently outside the cockpit instead of at the cockpit instrumentation. In addition, when the novice pilots did look at the mapping instrumentation, they gazed much longer than the experts, suggesting that they struggled to identify the critical landmarks on the electronic map. These differences, revealed by eye-tracking measurements, can be useful for developing training procedures to teach novice operators to use expert-like strategies in reading electronic maps.

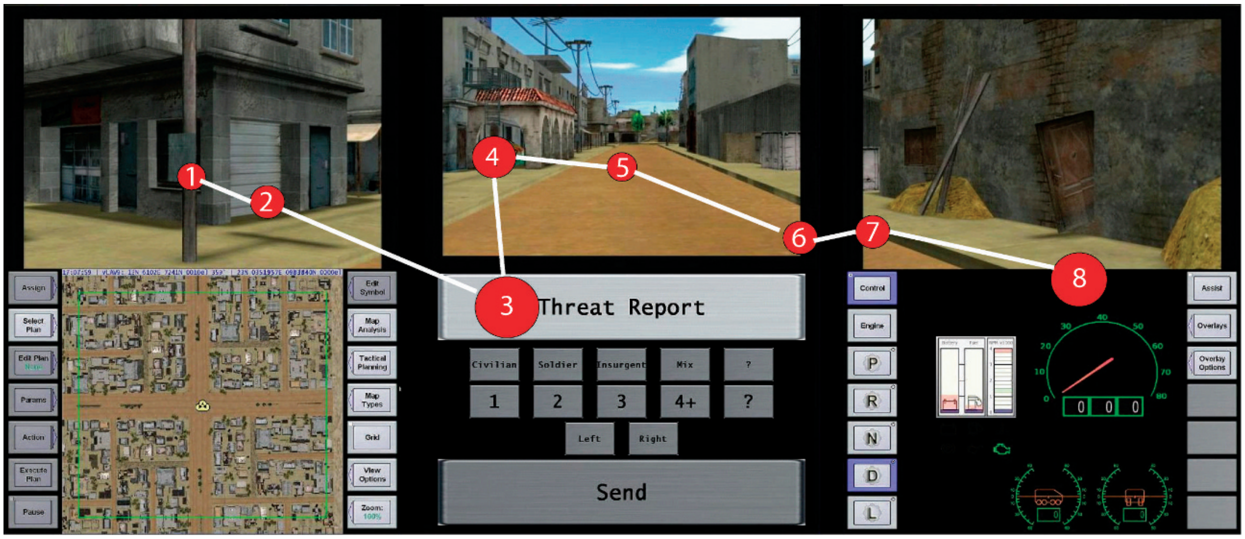

Figure 2. Eye movements superimposed on a crew station computer interface. Numbers indicate the order of eye fixations, and the size of each circle represents the total fixation duration.

Eye-tracking measures such as PERCLOS can help characterize dynamic changes in the operator's mental state owing to factors such as fatigue. In a study by Dinges et al. (1998), operators were intentionally sleep deprived for 42 hours and were then required to monitor a system for target events and to report when the events occurred. The operators completed questionnaires to monitor how sleepy they felt, and an eye-tracking system was used to monitor their visual search. Results of the study indicated that PERCLOS reliably predicted when operators were fatigued more effectively than the questionnaires, and the measure revealed points in time in which operators experienced changes in their level of fatigue. Performance decrements can be linked to the operator's state by capturing these dynamic changes.

This short review only touches the surface of the current-day capabilities of incorporating eye-tracking measures into T&E assessments. The examples presented here highlight potential applications of continuous measures to assess and/or compare how operators interact with system interfaces, as well as monitor dynamic changes in the operator's fatigue level. However, many other applications have been identified in several commercial applications, including the automobile and aviation industries. With their portability and ease-of-use, eye-tracking systems can easily be incorporated into T&E environments to identify additional assessment capabilities.

EEG in T&E

EEG provides a measurement of electrical activity in the brain using recordings from electrodes on the scalp. These measurements have been linked to dynamic changes in behavior and factors related to an operator's mental state. Traditional EEG systems have been very bulky, entailing set-up time to attach electrodes to the scalp with gel, and requiring the operator to minimize any head or body movement while physically tethered to the EEG recording devices. Traditional EEG systems are also highly susceptible to electrical artifacts from nearby equipment and other non-brain sources of electrical activity. System calibration must be performed for each operator before each testing session because the day-to-day variability of operator EEG measurements can be high (East, Bauer, and Lanning 2002). These attributes make traditional EEG systems impractical for current T&E environments; however, in the past few years, EEG systems have been developed that minimize these limitations. These newer systems are light-weight, often incorporating the electrodes into a hat or helmet, and are designed for use in real-world, operational environments. They contain advanced amplification and wireless transmission technology to minimize the impact of electrical interference. One such system is shown in *Figure 3*. As technology develops, these newer EEG systems will likely evolve to be as portable and easy-to-use as the eye-tracking systems in use today.

Results from EEG-based measurements collected in controlled environments show potential for application in operational settings. For example, Pope, Bogart, and Bartolome (1995) utilized a real-time index of operator task engagement, based on the power of EEG spectra, in an adaptive system to mitigate the effects of fatigue. EEG was collected while operators performed several

Figure 3. An example of an electroencephalography system designed to be worn under a helmet. A wireless base station receives and decodes the neural signals from several meters away, allowing the operator to be fully mobile. While this device shows promise in controlled environments, there is still work to be done to increase reliability in operational settings.

tasks including monitoring, resource management, and compensatory tracking. The tracking task was either automated by the system or manually controlled by the operator, and the system dynamically switched between the two modes based on the changes in the EEG-based index in order to minimize the effects of fatigue.

Similarly, Wilson, Lambert, and Russell (2000) also used an EEG-based measure in an automated system that adapted its functioning based upon changes in operator mental state. As in the Pope, Bogart, and Bartolome (1995) study discussed above, operators were required to perform multiple tasks. When the operator could not complete all assigned tasks, the system assisted with the monitoring task, requiring operators to solely focus on the remaining two tasks—resource management and tracking. The system switched into this assist mode only when the EEG-based measure of operator state indicated that they were overloaded and struggling to complete all tasks simultaneously (referred to as "high workload" by the authors). Of importance, by using the EEG measure to dynamically adapt the system, the number of errors decreased significantly in both the resource management task (33 percent) and in the tracking task (44 percent).

Other EEG findings obtained in controlled, laboratory settings show promise for future application in T&E environments. For instance, EEG-based measures have revealed neural signatures that are sensitive to error detection. The error-related negativity (ERN) produces a distinct pattern of brain activity when an individual makes an incorrect response or error. Monitoring operator errors in near–real time would allow testers the ability to identify when certain errors occur during system interaction as well as the ability to adapt system components based on errors committed. The ERN is not only sensitive to when an operator makes a known error, but it is also seen when an operator witnesses someone else making an error (van Schie et al. 2004). Preliminary applications of the ERN are currently being employed during the testing of brain-computer interfaces (Ferrez and Millan 2008). The ability to detect operator errors in real time could substantially augment T&E assessment capabilities.

Another laboratory-based finding demonstrates the utility of EEG for detecting predefined targets or identifying sudden changes in the environment without relying on overt responses from the operator. Using EEG in this way may provide testers with the ability to detect anomalies or unexpected changes that occur during system interaction. One instance of using this approach comes from image classification where an operator must detect a target of interest embedded in a series of rapidly presented images. Systems have shown their ability to accurately identify particular images for further analysis based on EEG measurements alone even though the images were shown for 100 milliseconds (Gerson, Parra, and Sajda 2006). This approach proved to be highly accurate in discriminating between target and nontarget images, and much faster than simply relying on self-reports from the operator performing the target-scanning task. The potential of EEG to rapidly identify what information is critical to the task at hand could greatly enhance how operator performance and system evaluation are analyzed in T&E environments.

These examples highlight the potential of EEG-based measures to index factors related to operator mental state. The examples described here are just a subset of the laboratory-based EEG findings that may have direct relevant applications to T&E. EEG-based measures can greatly augment the types of continuous operator assessments currently available with eye-tracking, and the combined use of both eye-tracking and EEG measurements may improve real-time assessment of operator mental states across many tasks, systems, and environments above and beyond using each method alone.

Conclusions

Eye-tracking and EEG are two tools that provide continuous measurements of operator performance, and both provide powerful analysis tools to the T&E community. Eye-tracking systems are simple to use, portable, and provide measurements related to operator mental state, such as blink rate and PERCLOS, that can be applied in a T&E setting today. Although they are not ready for use in all field settings, EEG systems have the potential to provide additional measures related to operator mental state that may be of use to testers and evaluators. Studies using EEG have successfully adapted systems and improved operator performance, and other more preliminary studies using laboratory-based EEG measures such as the ERN show how real-time assessment of operator performance can be augmented in applied settings. As additional research is conducted, EEG-based measures obtained in controlled settings can be applied to the complex and dynamic environments of T&E. Future research should explore the combined use of eye-tracking and EEG systems to create a broad-based measure of changes in operator performance, expanding assessment capabilities to a diverse set of tasks and environments. ❑

DR. ANTHONY RIES is a research psychologist at Aberdeen Proving Ground in Aberdeen, Maryland. He is currently serving as deputy manager of the "High-Definition Cognition in Operational Environments" U.S. Army Technology Objective (Research), which is focused on enabling assessment of soldier performance in operationally relevant environments using behavioral and physiological measures. Dr. Ries received a bachelor of science degree in psychology from Northwest Missouri State University in 2000 and master of arts and doctor of philosophy degrees in cognitive psychology from the University of North Carolina at Chapel Hill in 2003 and 2007, respectively. E-mail: Anthony.ries@us.army.mil

DR. JEAN VETTEL is employed by the U.S. Army Research Laboratory in the Translational Neuroscience Branch at Aberdeen Proving Ground in Aberdeen, Maryland. Her research focuses on improving soldier–system performance by conducting neuroscience research in complex, operational settings. She received a bachelor of science degree from Carnegie Mellon University in 2000 and master of science and doctor of philosophy degrees from Brown University in 2008 and 2009, respectively. E-mail: Jean.vettel@us.army.mil

References

Babcock, J., M. Lipps, and J. B. Pelz. 2002. How people look at pictures before, during, and after image capture: Buswell revisited. In *Proceedings of SPIE, Human Vision and Electronic Imaging*, 4662:34–47. Bellingham, WA: SPIE.

Dinges, D. F., M. Mallis, G. Maislin, and J. W. Powell. 1998. Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. Department of Transportation Highway Safety, pub. 808 762.

East, J. A., K. W. Bauer, and J. W. Lanning. 2002. Feature selection for predicting pilot mental workload: A feasibility study. *International Journal of Smart Engineering System Design* 4: 183–193.

Ferrez, P. W., and J. del R. Millan. 2008. Erro-related EEG potentials generated during simulated brain-computer interaction. IEEE Trans Biomedical Engineering, March, 923–929.

Gerson, A. D., L. C. Parra, and P. Sajda. 2006. Cortically-coupled computer vision for rapid image search. IEEE Trans Neural Systems and Rehabilitation Engineering, June 1–6.

Hart, S. G., and L. E. Staveland. 1988. Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In *Human mental workload*, ed. P. A. Hancock and N. Meshkati. Amsterdam, The Netherlands: Elsevier.

Ottati, W. L., J. C. Hickox, and J. Richter. 1999. Eye scan patterns of experienced and novice pilots during Visual Flight Rules (VFR) navigation. In *Proceedings of the Human Factors and Ergonomics Society, 43rd Annual Meeting*, Houston, Texas. Santa Monica, CA: Human Factors and Ergonomic Society.

Pope, A. T., E. H. Bogart, and D. Bartolome. 1995. Biocybernetic system evaluates indices of operator engagement. *Biological Psychology* 40: 187–196.

van Schie, H., R. Mars, M. Coles, and H. Bekkering. 2004. Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience* 7: 549–554.