Using eye tracking for live measures of workload in a refinery control room process monitoring task

Benjamin Noah, Ling Rothrock
The Pennsylvania State University, University Park, PA

Process monitoring within a refinery control room can be both daunting and listless. The workload of control room operators can fluctuate greatly over a short period of time. It would be ideal if systems could automatically detect these changes in workload so that both the computer system and those supervising are made aware of instances when an operator might be overloaded. While eye tracking has been shown to correlate well with workload, the process industry has yet to study this phenomenon. 48 control room operators from a major refinery participated within this experiment, where task complexity was varied within scenarios by controlling for the number of process deviations within a simulated crude process. It was found that both blink rate and pupil size could be used to track the cognitive load of operators over time. NASA TLX was also measured but it was not as sensitive as pupil size, as it did not capture a difference in scenario difficulty.

INTRODUCTION

Maintaining a steady state process is a primary objective of process control operators. This becomes especially critical for the energy industry as disasters could translate to financial losses, environmental damage, and even death. Nuclear power plants, such as the Chernobyl accident or the recent tsunami-related event in Fukushima Japan, could have more world-wide consequences. While the latter event could be arguably attributed to a natural disaster, the former was more specifically a result of poorly trained personnel. Whatever the cause, it was the design of the facility and the technology used, along with the human operators, that played a major role in each catastrophe. As human factors practitioners, we are interested in the specifics of how human operators interact with their workplace environment, and how we could continuously improve the current status quo. Much research has been done in the areas of improving interface design, team formation, and training (Patrick, James, & Ahmed, 2006). Much less research has been done on adaptive systems within process control. Cognitive-aware systems are those which allow the system to identify human processes, which in response could allow the system to adapt to the human (Bulling & Zander, 2014). Eye tracking technology has the potential to improve the human-machine system within the control room environment by potentially recognizing when a human is under unmanageable stress.

Process Control

During a process disturbance, an operator must quickly and reliably conduct a series of corrective measures to ensure that the process is stabilized. One of the primary methods of monitoring such process are making use of high level overview displays, which present the human operator with an abstract view of the entire process they are responsible for maintaining. This type of display may provide early indicators of unexpected deviations, which often present themselves within qualitative or quantitative indicators on the computer screen. If these go unnoticed, it may lead to unstable conditions upon which the operator is bombarded with warnings in the fashion of changing indicator colors and/or audible alarms. While one such deviation is often manageable, a multitude of them could bring the operator's workload to levels beyond human capability. Furthermore, our knowledge of individual differences tells us that people inherently behave within different capacities - yet systems are designed uniformly. It would be a tremendous benefit, to the operation, for the system to be aware of the current mental state of the human operator. Additionally, supervisors could be informed when an operator may be experiencing high levels of workload.

Workload

This proposal will look at how the previously mentioned eye tracking measures stack up against the commonly used subjective NASA Task Load Index (NASA TLX; Hart & Staveland, 1988). NASA TLX provides a good baseline because it is simply the measure which has been shown to provide a valid indication of workload (Hart, 2006). Other measures of workload (e.g., EEG, ECG, transcranial Doppler sonography, functional near infrared, eye tracking) have been shown to not necessarily correspond well with one another (Matthews, Reinerman-Jones, Barber, & Abich, 2014). However, implementing eye tracking in such an environment with table mounted trackers could be the most feasible without interrupting normal work operations. The other measures, including NASA TLX, require that the operator have a physical device attached to them or that they take time out of their day to respond to workload questionnaires.

Eye Tracking

Various eye tracking measures have been linked to measures of cognitive load. For instance, longer eye fixation durations can be an indicator of more difficulty in interpreting information that is being observed (e.g. Duchowski, 2007; Holmqvist et al., 2011; Just & Carpenter, 1976). Some researchers have used eye blinks to measure attention and tension (Bruneau, Sasse, & McCarthy, 2002). Higher blink rates, for example, could correlate with fatigue (Poole & Ball, 2005). Cognitive workload has also been found to correlate with pupil size, where increasing size indicates higher processing demands (Beatty, 1982; Szulewski, Fernando, Baylis, & Howes, 2014). Saccade velocity has also been found to increase with workload (Bodala, Ke, Mir, Thakor, & Al-Nashash, 2014).

Applying eye tracking within process control simulation experiments is not new. For the nuclear power plant domain, Ha and Seong (2009) measured eye fixations to determine attention on important areas through an fixation-to-importance ratio (FIR). Their primary goal was to use an evaluation method to determine the difficulty in information search. Ikuma et al. (2014) used eye tracking to measure gaze within varying workload simulations of a virtual petrochemical process. Their use of eye tracking, however, was limited to locating where participants looked using areas of interest (AOI). While these sorts of studies provide tools for assisting in the development of good interfaces, they were not used as methods to measure operator performance in real-time. Furthermore, cognitive workload is not behavioral as directing visual attention is.

Outside of process control, eye tracking has been used to infer cognitive workload in areas such as: aviation (Ahlstrom & Friedman-Berg, 2006; Hankins & Wilson, 1998; Wilson, 2002), website analysis (Wang, Yang, Liu, Cao, & Ma, 2014), driving (Hwang, Yoon, Kim, & Kim, 2014; Palinko, Kun, Shyrokov, & Heeman, 2010; Savage, Potter, & Tatler, 2013), word processing and reading (Gwizdka & Zhang, 2015; Just & Carpenter, 1993), medical (Zheng et al., 2012), and robotic control/military applications (Breslow, Gartenberg, McCurry, & Trafton, 2014; Orden, Limbert, Makeig, & Jung, 2001). Most of these studies used eye blink frequency and pupil diameter variation as measures of cognitive workload. Hence, this study investigates these two forms of eye tracking measures. Note that neither of these are direct eye movement measures, which reduces the confounding effects (artifacts) within the data.

METHOD

An experiment was designed to measure the effect of presenting operators with three different visualizations of an overview display (Noah, Kim, Rothrock, & Tharanathan, 2014). The primary objective of that experiment was to determine which display characteristics promote better overall performance and situation awareness in a simulated crude oil process. The main results of that study was that a surface chart display improved performance, while situation awareness did not differ significantly between the three displays. In the same experiment, eye tracking data was also collected for this analysis, as an extension to the study.

While one factor varied the display type at three levels (surface chart, heat map, and visual thesaurus), each participant underwent two experimental trials with each a either low or high scenario complexity. Complexity was balanced so that half of the participants went from low-to-high and the other half went from high-to-low. Finally, the three display type conditions were between-subject.

Subjective workload was measured by using NASA TLX for each, of two, experimental trials. Eye tracking data for both eyes was collected using a head mounted Arrington Research tracker at 30Hz throughout each trial. Measuring workload through eye tracking fits this experiment well as it considers exposure to various visualizations.

The general procedure for the experiment were as follows. First, the participants signed a voluntary consent form. Approval for the study was granted by both the site's management and Abnormal Situation Management (ASM) Consortium, who provided the funds for the project. Next, standardized training was provided and a quiz was administered to ensure proper knowledge acquisition. This included both a questionnaire and a full practice session. The eye tracker was calibrated to each participant at the start of the practice session, and later re-calibrated prior to each experiment trial. Following each experiment trial, participants answered the NASA TLX questionnaire on a computer screen. A debriefing and short usability survey concluded the experiment.

Tasks

The primary task was to monitor and acknowledge any changes in state within the simulated crude process (using both monitors on the left in Figure 1). The only action required was to click on the gauges which changed. The visualizations used were developed to mimic actual overview displays being used in control rooms across the planet. The simulated scenarios were developed along with subject-matter experts so that the process deviations translated to actual progression of similar process within refinery operations. Each task lasted about 20 minutes. The simulated scenarios were set at 9 minutes in duration, with a set of situation awareness questions about half way through and NASA TLX questions at the end. The time to respond to questions was not restricted and varied between participants. The task always started off with 100% normal, steady-state operation. Low complexity scenarios contained an average of 28 abnormal events, and high contained an average of 47.

In addition to the primary task, a secondary task required that participant manage an independent task (right monitor in Figure 1). The secondary task used the Multi-Attribute Task Battery (MATB). The reason for including the secondary task was to ensure a high enough load throughout the experiment. The difficulty level on the secondary task remained constant. See figure 1 for an image of the setup. The primary task made use to two 22" widescreen computer monitors (stacked on top of one another), and the secondary task made use of a 15" square monitor (which was placed immediately to the right).

Figure 1. Experiment setup showing the overview displays (primary task) on the left and MATB (secondary task) on the right.

Participants

In sum, 48 control room operators were recruited from a major oil refinery and they participated in the process monitoring simulation tasks, 16 per display type. The mean age was 37.7. The range for age was 22-59. Experience levels were also recorded in terms of years of experience in the control room. The average was 7.5 years, ranging from 1 to 34. 43 were male, and 5 were female.

RESULTS

The raw data had to be first cleaned and converted into fixations and saccades, from which scanpaths could be plotted. Figure 2 shows a sample of one participants' data. The circles represent fixations, with larger diameter circles corresponding to longer fixations. Overall, the NASA TLX data resulted in a significant difference between trials, $F(1,42)=7.11$, $p=0.011$, with higher workload in the first trial.

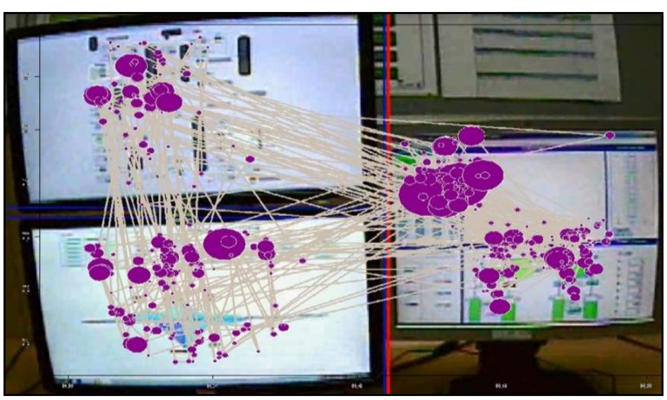

Figure 2. Sample participant scanpath data overlayed on visual scene.

Blinks were extracted from the raw data as artifacts, when the eye tracker could not locate the pupil due to closing of the eye lids. The total number of blinks was calculated within 60 second time blocks. The average number of blinks for participants in these blocks can be seen in Figure 3.

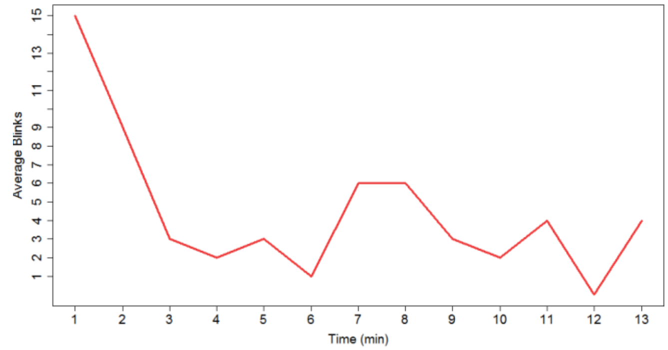

Figure 3. The average number of blinks per participant. Blinks were calculated within 60 second time blocks.

Pupillometry is based on changes in pupil diameter through dilation, and/or constriction of the pupil. The eye tracker recorded estimates of width and height of each pupil at 30Hz. The pupil size was calculated as the area of an ellipse: area = π * width * height. As both eyes behave similarly by reacting to stimuli under the same mechanisms, only one eye can be analyzed. In this analysis, the left eye us chosen for the analysis on pupil size.

The size of pupils with respect to task complexity is shown in Figure 4. The plot shows average pupil size calculated every second, by using 30 consecutive data points. At 30 Hz, 30 data points equal to one second of task duration. The data was fitted in two ways. First, curves were fitted based on local polynomial regression using loess fitting, shown as the solid nonlinear lines. Second, linear lines were also fitted and shown as dotted lines. Analysis of covariance was conducted to determine if the two complexity levels differed. The linear fits for each complexity were statistically different, $F(1,448)=705.94$, $p=0.000$.

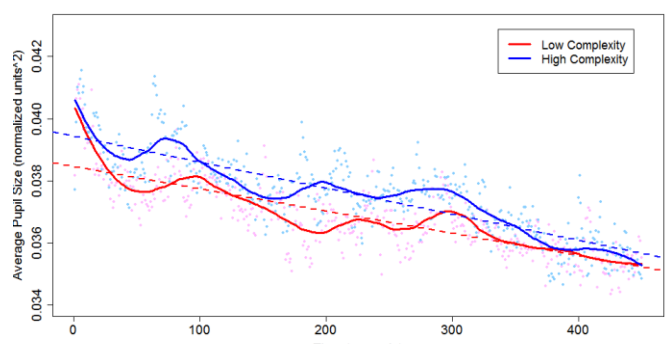

Figure 4. The average pupil size over time for low (red) and high (blue) complexity.

The size of pupils with respect to experimental trial is shown in Figure 5. Once again, this data was averaged every second, in 30 consecutive data points. The solid fitted lines represent the loess fitting curves, and the dotted lines represent linear fits. Analysis of covariance shows that the two trials differed statistically, $F(1,448)=980.36$, $p=0.000$.

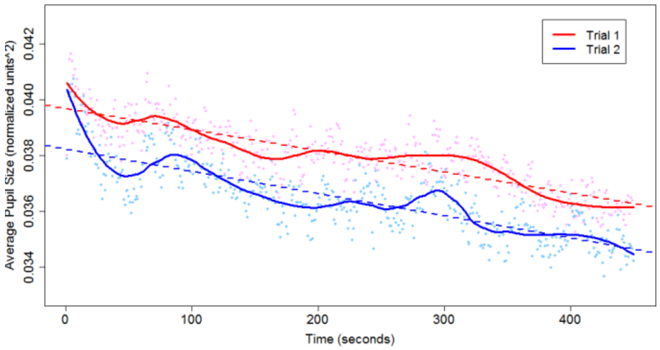

Figure 5. The average pupil size over time for trial 1 (red) and trial 2 (blue).

DISCUSSION

The first interesting finding is that blinks did not increase over time within each experimental scenario (Figure 3). The data showed that participants began each experiment in a high level of stress, perhaps indicating an orientation phase (Hasse & Bruder, 2014). While the scenarios started with no process deviations, more blinks occurred in the first 2-3 minutes, then steadied for the remainder of the time. However, around the 6-minute marker, blinks had a sharp increase. This corresponds to the time when the simulation froze for the situation awareness questions. There were eight questions and it took participants several minutes to respond prior to continuing the simulation. What this indicates is that blinks could potentially provide insight into cognitive load for various stages of process monitoring.

While physiological measures have not been found to correlate well with NASA TLX (Matthews et al., 2014), we found that both NASA TLX and pupil size corresponded to more workload within trial 1 (Figure 5). It is possible that this is evidence of a learning effect. However, performance data did not support a learning effect (Noah et al., 2014). It may be that participants became more familiar over time, but not to the point where it helped them perform better. Another finding was that higher complexity scenarios did show larger pupil sizes throughout the duration of the experiments (Figure 4). This is something that NASA TLX did not capture. Therefore, pupilometry shows to be more sensitive than subjective questionnaires. Finally, it can be seen that both blinks and pupil size data support the interpretation the workload decreased as the experiment progressed.

There are several limitation to this study. While the use of blinks and pupil size help reduce confounding effects, having participants continuously switch visual attention between three computer monitors does present data analysis challenges. Further validation of these findings could be done within the process control domain. Another challenge to pupilometry is that pupil constriction can be attributed to many things. There are several recommended articles on this subject (Beatty & Kahneman, 1966; Granholm, Asarnow, Sarkin, & Dykes, 1996; Granholm & Steinhauer, 2004; Muldner, Christopherson, Atkinson, & Burleson, 2009; Paivio & Simpson, 1966; Pomplun & Sunkara, 2003).

CONCLUSIONS

For a simulated process control task, it was found that measuring pupil size with eye tracking supports the ongoing effort to relate increasing pupil size to higher cognitive load. In addition, the frequency of blinks can potentially indicate an orientation phase and changes in cognitive demands. The main advantage to pupilometry and blink count is the ability to correspond with cognitive workload temporarily. This, in turn, should lend well to integrating eye tracking within adaptive systems for process control rooms of the future.

REFERENCES

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623–636.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292.

Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*(10), 371–372.

Bodala, I. P., Ke, Y., Mir, H., Thakor, N. V., & Al-Nashash, H. (2014). Cognitive workload estimation due to vague visual stimuli using saccadic eye movements. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2993–2996). doi:10.1109/EMBC.2014.6944252

Breslow, L. A., Gartenberg, D., McCurry, J. M., & Trafton, J. G. (2014). Dynamic Operator Overload: A Model for Predicting Workload During Supervisory Control. *IEEE Transactions on Human-Machine Systems*, *44*(1), 30–40.

Bruneau, D., Sasse, M. A., & McCarthy, J. (2002). The eyes never lie: The use of eye tracking data in HCI research (Vol. 2, p. 25). Presented at the Proceedings of the CHI.

Bulling, A., & Zander, T. O. (2014). Cognition-Aware Computing. *IEEE Pervasive Computing*, *13*(3), 80–83.

Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*. Springer Science & Business Media.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461.

Granholm, E., & Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, *52*(1), 1–6.

Gwizdka, J., & Zhang, Y. (2015). Towards Inferring Web Page Relevance – An Eye-Tracking Study. Retrieved from https://www.ideals.illinois.edu/handle/2142/73709

Ha, J. S., & Seong, P. H. (2009). A human–machine interface evaluation method: A difficulty evaluation method in information searching (DEMIS). *Reliability Engineering & System Safety*, *94*(10), 1557–1567.

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, *69*(4), 360–367.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Hasse, C., & Bruder, C. (2014). Eye-tracking measurements and their link to a normative model of monitoring behaviour. *Ergonomics*, (ahead-of-print), 1–13.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Hwang, Y., Yoon, D., Kim, H. S., & Kim, K.-H. (2014). A Validation Study on a Subjective Driving Workload Prediction Tool. *IEEE Transactions on Intelligent Transportation Systems*, *15*(4), 1835–1843.

Ikuma, L. H., Harvey, C., Taylor, C. F., & Handal, C. (2014). A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. *Journal of Loss Prevention in the Process Industries*, *32*, 454–465.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*(2), 310–39.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2014). The Psychometrics of Mental Workload Multiple Measures Are Sensitive but Divergent. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720814539505.

Muldner, K., Christopherson, R., Atkinson, R., & Burleson, W. (2009). Investigating the Utility of Eye-Tracking Information on Affect and Reasoning for User Modeling. In G.-J. Houben, G. McCalla, F. Pianesi, & M. Zancanaro (Eds.), *User Modeling, Adaptation, and Personalization* (pp. 138–149). Springer Berlin Heidelberg. Retrieved from http://link.springer.com.ezaccess.libraries.psu.edu/chapter/10.1007/978-3-642-02247-0_15

Noah, B., Kim, J.-H., Rothrock, L., & Tharanathan, A. (2014). Evaluating Alternate Visualization Techniques for Overview Displays in Process Control. *IIE Transactions on Occupational Ergonomics and Human Factors*, *2*(3-4), 152–168.

Orden, K. F. V., Limbert, W., Makeig, S., & Jung, T.-P. (2001). Eye Activity Correlates of Workload during a Visuospatial Memory Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(1), 111–121.

Paivio, A., & Simpson, H. M. (1966). The effect of word abstractness and pleasantness on pupil size during an imagery task. *Psychonomic Science*, *5*(2), 55–56.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 141–144). New York, NY, USA: ACM. doi:10.1145/1743666.1743701

Patrick, J., James, N., & Ahmed, A. (2006). Human processes of control: tracing the goals and strategies of control room teams. *Ergonomics*, *49*(12-13), 1395–1414.

Pomplun, M., & Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in human-computer interaction. Presented at the Proceedings of the International Conference on HCI.

Poole, A., & Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research: Current status and future. Prospects", Chapter in C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group. *Inc.*

Savage, S. W., Potter, D. D., & Tatler, B. W. (2013). Does preoccupation impair hazard perception? A simultaneous EEG and Eye Tracking study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *17*, 52–62.

Szulewski, A., Fernando, S. M., Baylis, J., & Howes, D. (2014). Increasing Pupil Size Is Associated with Increasing Cognitive Processing Demands: A Pilot Study Using a Mobile Eye-Tracking Device. *Open Journal of Emergency Medicine*, *2014*.

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10.

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology*, *12*(1), 3–18.

Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., & Atkins, M. S. (2012). Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surgical Endoscopy*, *26*(10), 2746–2750.