

Eye Activity Correlates of Workload during a Visuospatial Memory Task

Karl F. Van Orden, Wendy Limbert, and Scott Makeig, Naval Health Research Center, San Diego, California, and Tzyy-Ping Jung, University of California, San Diego, San Diego, California

Changes in six measures of eye activity were assessed as a function of task workload in a target identification memory task. Eleven participants completed four 2-hr blocks of a mock anti-air-warfare task, in which they were required to examine and remember target classifications (friend/enemy) for subsequent prosecution (fire upon/allow to pass), while targets moved steadily toward two centrally located ship icons. Target density served as the task workload variable; between one and nine targets were simultaneously present on the display. For each participant, moving estimates of blink frequency and duration, fixation frequency and dwell time, saccadic extent, and mean pupil diameter, integrated over periods of 10 to 20 s, demonstrated systematic changes as a function of target density. Nonlinear regression analyses found blink frequency, fixation frequency, and pupil diameter to be the most predictive variables relating eye activity to target density. Participant-specific artificial neural network models, developed through training on two or three sessions and subsequently tested on a different session from the same participant, correlated well with actual target density levels (mean $R = 0.66$). Results indicate that moving mean estimation and artificial neural network techniques enable information from multiple eye measures to be combined to produce reliable near-real-time indicators of workload in some visuospatial tasks. Potential applications include the monitoring of visual activity of system operators for indications of visual workload and scanning efficiency.

INTRODUCTION

Assessing and predicting human workload is an important consideration in designing new systems, modifying existing systems, and avoiding task overload in real time through task reallocation or adaptive automation. Previous research has established that truly adaptive systems will require information on the human operator's workload levels in real-time, as it is difficult to predict actual workload based on modeled estimates alone (Byrne & Parasuraman, 1996). Parasuraman, Bahri, Deaton, Morrison, and Barnes (1992) proposed that a combination of three assessment domains (environment, activity, and operator state) can provide estimates

of workload with greater stability than any subset of measures.

Environment (or system state) information refers to knowledge of an operator's task loading. For example, the number of aircraft that must be monitored by an air traffic controller may provide a general indication of workload. Communication activity (monitored on a radio circuit) might reflect the extent to which a set of aircraft requires attention by the controller (i.e., *activity*). Psychophysiological measures (e.g., heart rate variability and electroencephalograph [EEG] spectral measures) provide insight into an operator's psychophysiological state (*operator state*), which in turn may correlate with workload (Kramer, Trejo, & Humphrey, 1996).

Whereas environmental factors can often be assessed via intelligent agents and/or sensors, and operator activity can often be measured directly from system interactions, operator state measures are often more difficult to obtain, process, and relate to task loading and performance. Most psychophysiological measures, including EEG and heart rate, require sensors to be physically attached to the operator. The recent development of high-quality, remotely positioned face- and eye-tracking technology (see Pastoor, Liu, & Renault, 1999) makes possible the unobtrusive acquisition of real-time eye activity information that could contribute to an integrated workload assessment system.

Several eye activity measures have been shown to correlate with visual and/or cognitive demands imposed by tasks. Generally, blink rate and duration decline as a function of greater workload (Brookings, Wilson, & Swain, 1996; Hankins & Wilson, 1998; Veltman & Gaillard, 1998). Data from Wilson, Fullenkamp, and Davis (1994) indicate that blink duration decreased to a greater extent during a taxing visual tracking task with minimal cognitive load than during a more cognitively challenging flight simulation task. Alternatively, higher blink rates and longer blink durations may occur with task-induced increases in saccadic extent caused by higher workload levels, as is the case when a pilot is scanning for information both within and outside a cockpit during landing (Fogarty & Stern, 1989; Hankins & Wilson, 1998). It is possible that the visual system takes advantage of the opportunity to blink during the longer saccades required by some tasks.

Many eye movement parameters (e.g., saccadic extent) are highly task-specific. For flight simulation tasks, visual scanning requirements often change as a function of the flight maneuver being performed by the pilot (Hankins & Wilson, 1998; Itoh, Hayashi, Tsukui, & Saito, 1990; Katoh, 1997). Several visual search studies (e.g., Van Orden, Nugent, Lafleur, & Moncho, 1999; Zelinsky, Rajesh, Hayhoe, & Ballard, 1997) showed that more effortful search, as indicated by poorer performance accuracy and longer search times, is associated with a greater frequency of fixations and no change in dwell time. However, Callan (1998) reported that the

frequency of long fixations (exceeding 500 ms) correlated with the number of flight rule errors committed during a flight simulation task. The occurrence of long fixations was attributed to increases in cognitive processing load during periods in which pilots were experiencing greater task difficulty.

Pupil diameter, although also affected by changes in illumination, stimulus characteristics, and accommodative behaviors, has been shown to generally increase with higher cognitive processing levels (Bucks & Walrath, 1992; Beatty & Wagoner, 1978). Pupil changes can be dynamic, as during comprehension of discrete sentences (Just & Carpenter, 1993), or sustained, as is the case during digit span recall (Granholm, Asarnow, Sarkin, & Dykes, 1996).

Although they have demonstrated general eye activity changes with shifts in task demand, previous studies have not attempted to combine multiple eye measures or utilize sufficiently short (~1-min) integration times; both of these would be required in a system that would contribute to real-time workload estimation. A recent study by Van Orden, Jung, and Makeig (2000) on eye activity correlates of fatigue successfully estimated performance in a sustained visual tracking task by combining 60-s moving-window estimates of eye activity. Here a similar signal-processing approach is used to analyze eye data during a task in which memory and visual activity demands are varied over time, and the accuracy of models combining eye measures in predicting moment-to-moment fluctuations in task demand is assessed.

METHOD

Participants

Eleven paid volunteers (five females and six males; 20 to 54 years old, mean age 31.6 years) completed the study. Participants were recruited from the laboratory staff and from a local college. None of the participants wore corrective lenses or was familiar with the specific purposes of the study.

Task Description

A mock air warfare task (similar to the target/threat identification task originally devel-

oped by Ackerman and Kanfer, 1994) was run on an 80486 computer (Datel Computers, San Diego, CA). The goal of the task was to allow “friendly” targets to pass and to fire on “enemy” targets that approached two ships on a display (see Figure 1). Targets traversed the distance from display edge to ship inner ring in 5 to 7 s. The participant was required to classify each white circular target symbol as a friend or an enemy and to destroy enemy targets when they had moved between two range rings surrounding each ship.

Identifying a target was accomplished by first selecting a target symbol using a trackball-controlled cursor, then eliciting an identification tag next to the symbol, and, finally, changing the symbol’s color to red. A list of names and classification categories, permanently displayed at the upper left corner of the display (e.g., “TRON – ENEMY,” “TACO – FRIEND,” etc.), was used by the participant to classify each target. Pressing one of two classification buttons on a control panel cleared the identification tag and changed the target’s color to yellow (regardless of whether it was a friend or an enemy) to indicate that it had been classified. Thus participants were required to remember the classifi-

cation of each target as it approached the ship icons. To prevent participants from memorizing the identification list (upper left panel), the identification of each target name was changed at random intervals of about 20–30 s, but only when no targets were present on the display.

Firing on a target required selecting the target and pressing a “fire” button on the control panel. If the inbound target was an enemy, it would then disappear from the screen. If the fired-on target was a friend, the symbol would turn to red with a white cross at its center and a buzzer would sound in the participant’s headphones to indicate that an error had been committed. Once a target crossed the inner range ring, it had effectively reached a ship. If the target was a friend, it would simply disappear from the display. If it was an enemy, an explosion icon would appear over the ship, accompanied by an explosion sound in the participant’s headphones, indicating that a defensive error had occurred. The ship icon would reset to normal after approximately 2 s. Participants could choose to repeat the target identification query process; however these actions took several seconds – time the participant could not afford when multiple targets were present.

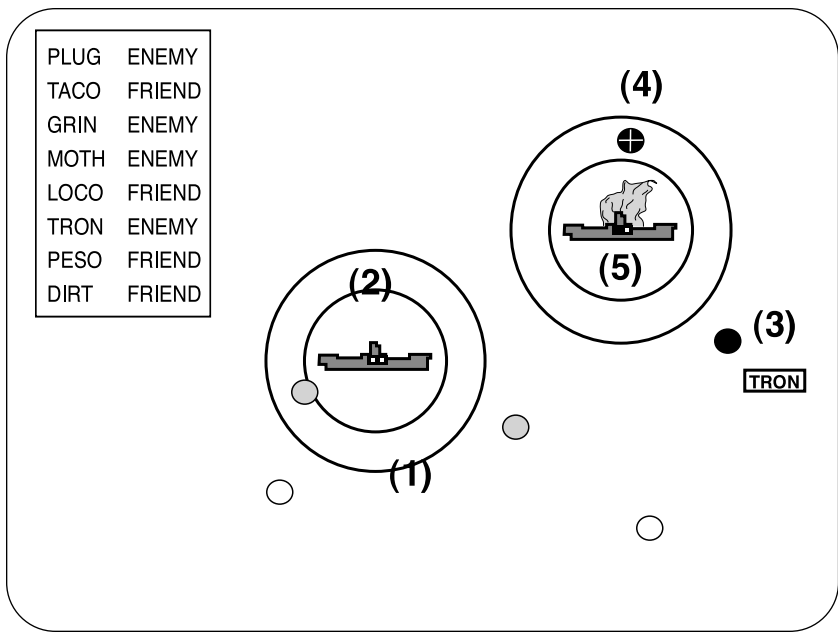


Figure 1. Task display. (1) = outer range ring; (2) = inner range ring; (3) = hooked target (being queried); (4) = friendly target that has been shot; and (5) = destroyed ship (which would reset after 1–2 s).

Workload Modulation

Task workload was varied by controlling target density on the display. At low target densities, one target appeared at random intervals (25 to 75 s), resulting in an average of one to two targets on the display at any time, including intervals of no target activity at all. At higher target densities, one target appeared every 5–7 s during a period of 1–2 min, resulting in an average of seven to nine targets on the display. Higher-density periods could also extend for periods of 5–7 min. At onsets of higher target density periods, the number of targets on the display increased steadily for approximately 1 min, at the end of which the rate of target presentation equaled the rate at which they were processed by the participant. The number of targets declined in a similar manner during transitions from higher to lower density levels.

Eye Tracking

Visual activity was monitored using an Applied Sciences Laboratory SU4000 eye-tracking system (Applied Science Laboratories, Bedford, MA). Participants were positioned in a chin-and-headrest, and eye activity was obtained with a remote camera system (colinearly aligned near-infrared optics) connected to the eye-tracking system. The system calculated the location and diameter of the pupil reflection and the location of the corneal reflection at a sampling rate of 60 Hz.

Procedure

Participants were tested on two days. On the first day, they completed three 30-min task-training sessions. Prior to beginning training, participants were fit with electroencephalographic (EEG) electrodes; those data are not reported in this paper.

Participants next completed the first of four 2-hr testing sessions. In each testing session, higher target density levels were presented within each of six 20-min segments, according to a randomized schedule. The three remaining testing sessions were completed on a different day.

Scoring

The task- and eye-time series data were well suited to an analysis approach previously used to study eye (Van Orden et al., 1999) and EEG

(Jung, Makeig, Stensmo, & Sejnowski, 1997; Makeig & Inlow, 1993) correlates of fatigue. In the present study, we attempted to use moving windows of the smallest duration that would enable adequate sampling of eye activity events and remain sensitive to changes in workload or the onset of errors. To estimate target density, the number of targets present on the display was assessed every 2 s. A 20-s window moving through the data in 2-s steps was used to derive local estimates of target density.

Blinks were defined as partial or complete eye closures lasting a minimum of 83.3 ms. Partial eye closures were calculated as horizontal pupil diameters of 35% or less of its mean value during a 240-s baseline period at the outset of the trial. The 83-ms minimum closure duration criterion prevented brief signal losses from being counted as blinks. Moving estimates of mean blink duration and blink frequency were then calculated using a 20-s square-window moved through the data in 2-s steps.

Point-of-regard (POR) data were used to calculate spatial locations and dwell times of eye fixations using a standard space-by-time boundary algorithm that examined successive POR locations in relation to a moving-mean centroid (see Van Orden et al., 2000, for details). Moving estimates of fixation frequency (fixations/min) and total fixation dwell time were derived using 10-s square windows moved through the data in 2-s steps. Moving estimates of mean pupil diameter (excluding closures and signal losses) were calculated similarly, using a window of 2-s duration. For each participant, the moving estimates of blink frequency and duration, fixation frequency and duration, and mean pupil size were merged with target density data prior to subsequent analyses.

RESULTS

Task performance data across all participants are shown as a function of target density in Figure 2. A repeated-measures analysis of variance (ANOVA) procedure for target density and session number indicated significantly higher task error rates at higher target densities, $F(8, 360) = 13.97, p < .001$. Tukey HSD post hoc analyses were conducted to test for

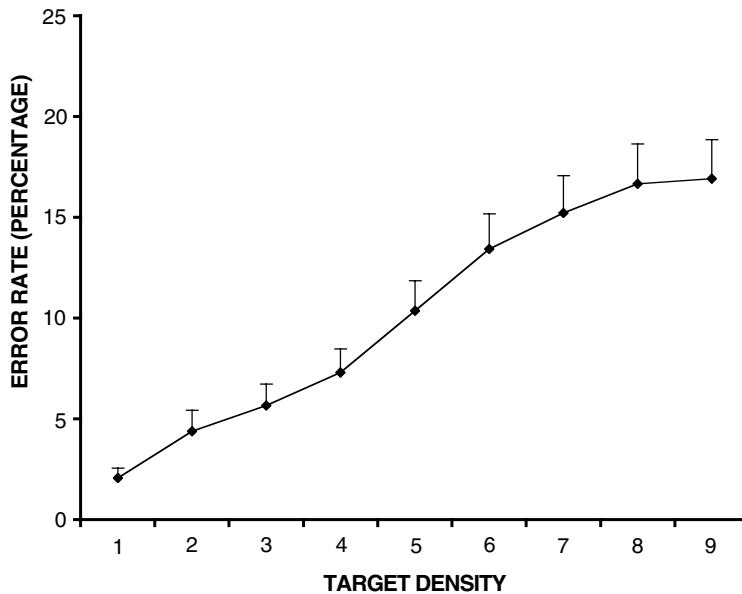


Figure 2. Mean percentage task error as a function of target density for all participants. Error bars represent one standard error of the mean.

differences among the means and indicated that error rates increased significantly ($p < .05$) when target density increased by three density levels or greater (e.g., density level one to four, two to five, etc).

Next, changes in eye activity as a function of target density were examined. Data from periods in which there were no targets on the display were excluded from all analyses. A significant main effect resulted from an ANOVA on blink duration, $F(8, 360) = 7.2$, $p < .001$. Subsequent post hoc testing indicated that the mean blink duration at target density of one item was significantly larger than means at densities of four or more targets, and the mean blink duration for target density of two items was significantly larger than means at target densities of five items or more ($p < .05$; see Figure 3). A similar pattern was found for mean blink frequency, $F(8, 360) = 13.00$, $p < .001$. The Tukey HSD test indicated that the mean blink frequencies for one- and two-target periods were higher than means at target densities of four or more, and blink frequency during three-target periods was significantly higher than mean frequencies at target densities of six or more ($p < .05$; see Figure 4).

An ANOVA and subsequent post hoc tests on fixation frequency data, $F(8, 360) = 6.37$, p

$< .001$, found a similar pattern of results as was observed for blink duration. The ANOVA on saccadic extent, $F(8, 360) = 3.15$, $p < .005$, indicated significant differences at only the most extreme target density means ($p < .05$). Mean fixation frequency and saccadic extent data are presented in Figures 5 and 6. Mean fixation duration did not vary as a function of target density, $F(8, 360) = 0.15$, $p < .5$. Although frequency of long fixations was significantly related to target density by ANOVA, $F(8, 360) = 2.37$, $p < .05$, post hoc testing failed to reveal any significant differences among the means.

Finally, pupil diameter was found to change as a function of target density, $F(8, 360) = 2.14$, $p < .05$. Pupil diameter means for target densities one and nine were significantly different ($p < .01$). Relative pupil diameter data are shown in Figure 7. (Actual diameter is dependent on distance of eye-tracking optics from the eye. Generally, 110 units is approximately 4 mm, and the slope of the function relating general units to actual diameter is nearly linear – 10 units represent about 1-mm change in diameter.) Session number was not a significant variable in any of these analyses.

Given the presence of target-density-related changes for several eye activity measures, a non-linear regression approach was used to combine

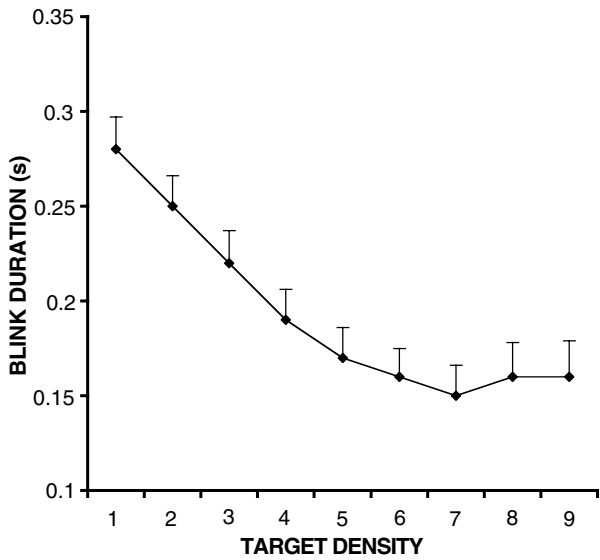


Figure 3. Mean blink duration as a function of target density. Means are based on four experimental sessions from each of 11 participants. Error bars represent 1 standard error of the mean.

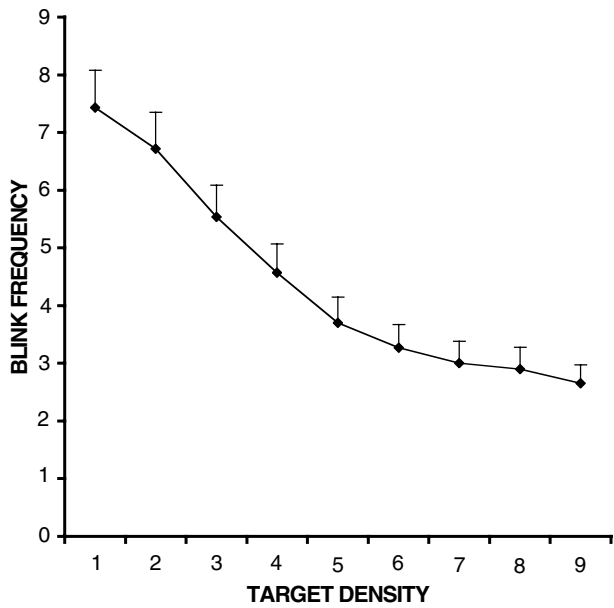


Figure 4. Mean blink frequency per 20-s epoch as a function of target density.

eye measures into a group regression model for estimating target density. In some sessions, several participants (e.g., Participant 2 and the final sessions for Participants 1–5 and 7) performed at near-chance levels of performance when target densities exceeded six items (error rates in excess of 40%). Because these sessions tended to occur near the end of the experiment, we suspect the participants reduced their effort (i.e.,

gave up) at these workload levels. These sessions were excluded from subsequent analyses.

Using the data from remaining participants, a standardized across-subjects (general) model of target density was developed using a stepwise procedure:

$$\text{Target Density} = -0.45 \times \Delta\text{BF} + 0.22 \times \Delta\text{FF} + .12 \times \Delta\text{PD}, \tag{1}$$

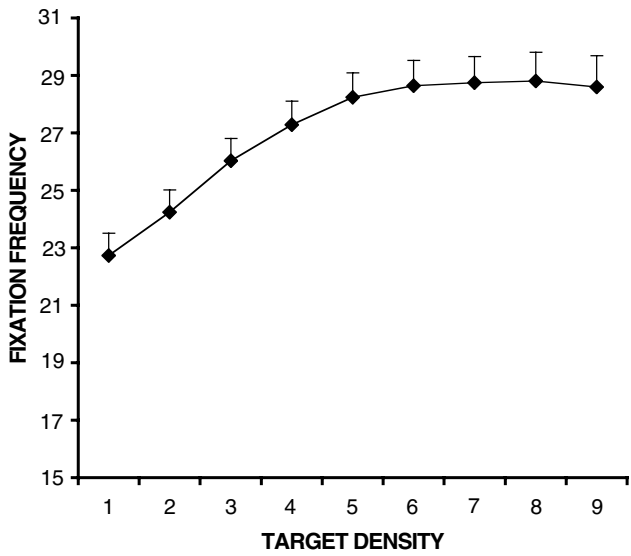


Figure 5. Mean fixation frequency per 10-s epoch as a function of display target density.

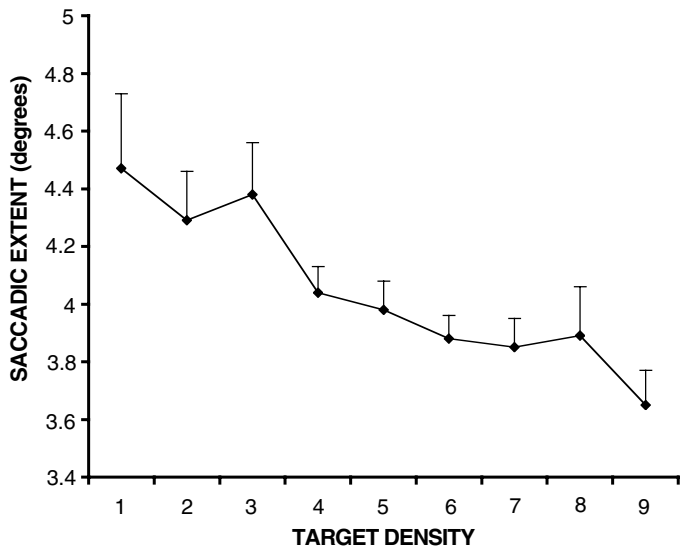


Figure 6. Saccadic extent in degrees of visual angle as a function of display target density.

where ΔBF = deviation from baseline blink frequency, ΔFF = deviation from baseline fixation frequency, and ΔPD = deviation from baseline pupil diameter.

The correlation coefficient (R) of general model-derived estimates of target density with actual levels was 0.55. Blink frequency accounted for 23.2% of the variance in the data, fixation frequency for 6.1%, and pupil diameter for 1.3%. Components of regression models developed for individual participants demon-

strated considerable variability and contained squared and cross-product predictor variables. The mean performance of these models was $R = 0.62$ (range 0.39–0.79). To mitigate the excessive statistical power resulting from more than 3000 data points in each time series, a $p < .01$ cutoff statistic was derived (from surrogate correlations within the present data set) and a correlation coefficient of $R = 0.36$ was established, as necessary, to indicate significant departure from zero. (See Van Orden et al., 1999,

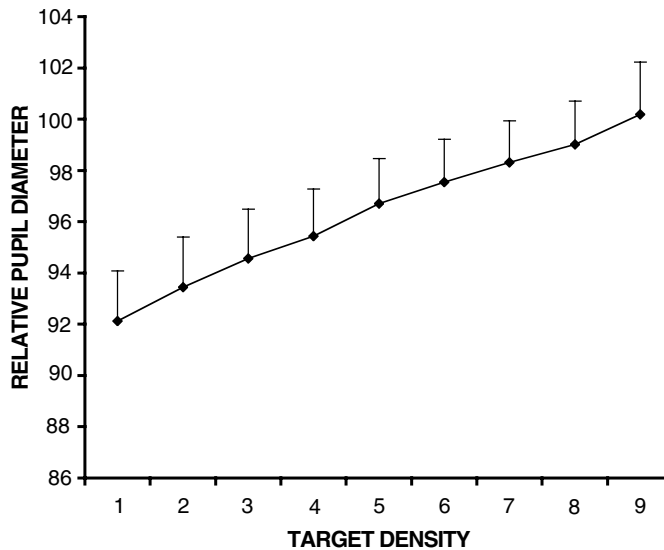


Figure 7. Mean relative pupil diameter as a function of target density.

for details concerning the calculation of a similar cutoff.)

The presence of complex nonlinear and combinatorial components within individual participant regression models indicated that artificial neural network (ANN) models would be useful for developing eye-based estimates of target density. For each participant, data from each session served as a test data set for ANN models developed using the remaining sessions (cross-session validation). Excluding the data from the test session, we shuffled two-thirds of the data and then used them to train a three-layer perceptron network employing back propagation algorithm to estimate target density. Data from the remaining one-third of the data set were used to validate the need for further network training.

Training was halted when the mean estimation error for the validation data stopped decreasing. Five different nets were trained on these data using different initial node weights. The median (third) best-performing neural net from the within-session validation data was then used to produce target density estimates from eye activity data for the participant's remaining test session. Multiple training and selection of the median performing ANN is a common practice to avoid selecting a network that has prematurely terminated its recursive training or has overfit the training data. Within-

session correlation coefficients of estimated to actual target density ranged from 0.27 to 0.88 (mean $R = 0.74$), and cross-session coefficients ranged from 0.06 to 0.85 (mean $R = 0.66$).

Mean within-session root mean square (RMS) error was 1.73 targets, and mean cross-session RMS was 2.02 targets. The minimum cross-validation correlation (0.06) was clearly an outlier; the next-lowest correlation coefficient was 0.33. The remaining 32 of 34 between-session correlation coefficients were above the 99% cutoff ($R = 0.36$). ANN-derived correlations of estimated versus actual target density for all sessions are presented in Table 1. ANN-estimated and actual target density data from a well-modeled cross-session data set ($R = 0.79$; rank: 7th of 34 sessions) are presented in Figure 8.

DISCUSSION

Eye activity measures obtained in the present experiment demonstrated systematic changes as a function of target density and replicated patterns found in previous eye activity reports. Under high target density conditions, blink frequency, fixation frequency, and saccadic extent demonstrated patterns indicative of higher visual processing load (Brookings et al., 1996; Hankins & Wilson, 1998; Veltman & Gaillard, 1998). Pupil diameter changes were consistent with greater cognitive load, as measured when

TABLE 1: Within- and Between-Session Neural Network Correlation Coefficients (*R*)

Participant	Session	Within-Session	Between-Session
1	A	0.78	0.65
	B	0.77	0.69
	C	0.88	0.85
3	A	0.48	0.33
	B	0.57	0.47
4	A	0.81	0.70
	B	0.64	0.52
	C	0.27	0.06
5	A	0.74	0.67
	B	0.65	0.57
	C	0.59	0.50
6	A	0.79	0.72
	B	0.78	0.72
	C	0.79	0.76
	D	0.83	0.77
7	A	0.68	0.51
	B	0.71	0.50
	C	0.63	0.56
8	A	0.75	0.66
	B	0.76	0.56
	C	0.76	0.74
	D	0.82	0.64
9	A	0.78	0.71
	B	0.84	0.79
	C	0.87	0.85
	D	0.81	0.79
10	A	0.79	0.78
	B	0.76	0.72
	C	0.76	0.68
	D	0.73	0.72
11	A	0.88	0.75
	B	0.88	0.83
	C	0.85	0.82
	D	0.87	0.83
Mean		0.74	0.66

participants were required to remember the classifications of a greater number of inbound targets (Beatty & Wagoner, 1978; Just & Carpenter, 1993). More important, moving-mean estimation methods combined with individualized ANN models were successful in producing reliable estimates of actual target density. These methods may represent a feasible approach to the design and development of real-time psychophysiological-based workload assessment systems.

Figure 8 illustrates both successes and limitations of the present approach. The responsiveness of the ANN model to rapidly fluctuating target density levels reflects the utility of the moving-mean estimation procedure with brief parameter integration times. The relatively good accuracy can be attributed to the capability of ANNs to model complex data sets, although continued training over extended periods might be expected to further improve the performance of these models. ANNs specifically designed

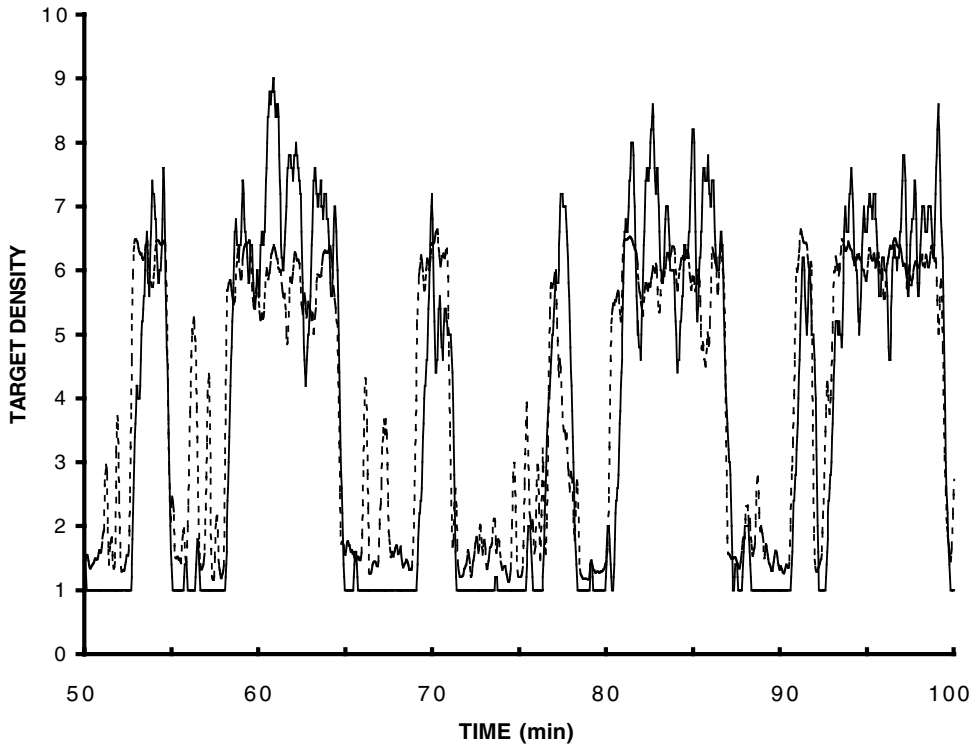


Figure 8. Plot of the neural-network-derived workload estimate (dashed line) superimposed on actual target density (solid line) during 50 min of Participant 9's second session ($R = 0.79$, RMS estimation error 1.36 items). The neural network was trained on data from this participant's three other experimental sessions.

for classification of different workload states, as would likely be used in a real-world application, might also demonstrate improved performance. The spikes in estimated target density that appear at low actual target density levels (e.g., near 57 and 66 min in Figure 8) likely represent moments of elevated task-independent visual activity – such as when a participant under low workload momentarily examines his or her surroundings. These behaviors would produce incongruities in the data that could be expected to occur irrespective of modeling accuracy.

Changes in eye activity with increases in task demand will likely be dependent on the nature of the task; thus a useful workload estimation system would couple psychophysiological measures with system state information (Parasuraman et al., 1992). For example, pilot psychophysiological data would be most useful if the specific flight maneuver (e. g., landing, level flight) was known by the system, so that workload within a maneuver could be more

closely estimated. Furthermore, multiple psychophysiological algorithms may be required to simultaneously monitor different operator states. In the present study, excessive variance in eye parameters in the absence of on-screen targets required modeling of workload for densities of one target or greater. At very low workload levels, real-world systems would benefit from both task loading information and, possibly, an algorithm to monitor eye activity correlates of drowsiness (see Van Orden et al., 2000) to more accurately characterize operator state.

To the extent possible, behavioral monitoring is a necessary component of any workload estimation approach. Several participants in the present study demonstrated behavior consistent with marginal effort, more frequently near the conclusion of the protocol. Although there have been attempts to correlate EEG signals to cognitive effort (Gevins et al., 1998), this dimension is not readily modeled using eye measures. Nevertheless, performance consequences should make such a condition easily

identifiable in real-world systems. For the majority of participants who attempted to maintain good performance in the present study, combined eye activity measures could be used to accurately model task load.

In addition to providing unobtrusive operator state data, eye data may provide valuable information about which stimuli an operator is (or is not) attending to during a task. In dynamic multitask environments, integrated workload assessment could be useful for triggering adaptive automation or other workload management strategies.

ACKNOWLEDGMENTS

This research was supported by a grant from the Office of Naval Research (ONR.WR.30030-6429). The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the U.S. government. The authors thank Echo Leaver and Shawn Wing for their assistance with the study. This work is not subject to U.S. copyright restrictions.

REFERENCES

- Ackerman, P. L., & Kanfer, R. (1994). *Improving problem-solving and decision-making skills under stress: Prediction and training* (Final Report for Contract N00014-91-J-4159). Minneapolis: University of Minnesota.
- Backs, R. W., & Walrath, L. C. (1992). Eye movements and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, 23, 243-254.
- Beatty, J., & Wagoner, B. L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science*, 199, 1216-1218.
- Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42, 361-377.
- Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, 42, 249-268.
- Callan, D. J. (1998). Eye movement relationships to excessive performance error in aviation. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 1152-1156). Santa Monica, CA: Human Factors and Ergonomics Society.
- Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: Their relationship to higher cognitive processes. *International Journal of Psychophysiology*, 8, 35-42.
- Gevens, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors*, 40, 79-91.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 3, 457-461.
- 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, 360-367.
- Itoh, Y., Hayashi, Y., Tsukui, I., & Saito, S. (1990). The ergonomic evaluation of eye movements and mental workload in aircraft pilots. *Ergonomics*, 33, 719-733.
- Jung, T.-P., Makeig, S., Stensmo, M., & Sejnowski, T. J. (1997). Estimating alertness from the EEG power spectrum. *IEEE Transactions on Biomedical Engineering*, 44, 60-69.
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47, 310-339.
- Katoh, Z. (1997). Saccade amplitude as a discriminator of flight types. *Aviation, Space and Environmental Medicine*, 68, 205-208.
- Kramer, A. F., Trejo, L. J., & Humphrey, D. G. (1996). Psychophysiological measures of workload: Potential applications to adaptively automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance* (pp. 137-162). Mahwah, NJ: Erlbaum.
- Makeig, S., & Inlow, M. (1993). Lapses in alertness: Coherence of fluctuations in performance and the EEG spectrum. *Electroencephalography and Clinical Neurophysiology*, 86, 23-35.
- Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems* (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center.
- Pastoor, S., Liu, J., & Renault, S. (1999). An experimental multimedia system allowing 3-D visualization and eye-controlled interaction without user-worn devices. *IEEE Transactions on Multimedia*, 1, 41-52.
- Van Orden, K. F., Jung, T.-P., & Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, 52, 221-240.
- Van Orden, K. F., Nugent, W., Lafleur, B., & Moncho, S. (1999). *Assessment of variable coded symbology using visual search performance and eye fixation measures* (NHRC Technical Report No. 99-4). San Diego, CA: Naval Health Research Center.
- Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41, 656-669.
- Wilson, G. F., Fullenkamp, B. S., & Davis, I. (1994). Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, Space and Environmental Medicine*, 65, 100-105.
- Zelinsky, G. J., Rajesh, P. N. R., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, 8, 448-453.

Karl F. Van Orden is a research psychologist at the Space and Naval Warfare Systems Center in San Diego, California. He received a Ph.D. in biological psychology from Syracuse University in 1988.

Wendy Limbert is a Ph.D. student at the University of California at Santa Cruz. She received her M.A. in psychology from San Diego State University in 1998.

Scott Makeig is a research psychologist at the Naval Health Research Center, San Diego, California. He has joint appointments with the University of California, San Diego, and the Salk Institute for Biological Studies. He received a Ph.D. in music psychobiology from the University of California, San Diego, in 1985.

Tzzy-Ping Jung is a research faculty member at the Institute for Neural Computation, University of California, San Diego. He received a Ph.D. in electrical engineering from Ohio State University in 1993.

Date received: July 30, 1999

Date accepted: June 19, 2000