Eye Movement as Indicators of Mental Workload to Trigger Adaptive Automation

Tjerk de Greef[1,3], Harmen Lafeber[2], Herre van Oostendorp[2], and Jasper Lindenberg[3]

[1] Delft University of Technology, Man-Machine Interaction Group, Mekelweg 4,
2628 GA Delft, The Netherlands
[2] University of Utrecht, Department of Information and Computing Science, Padualaan 14,
3594 CH Utrecht, The Netherlands
[3] TNO Defence, Security and Safety, Department of Human Factors, Kampweg 5, 3769 ZG
Soesterberg, The Netherlands
t.e.degreef@tudelft.nl, h.lafeber@gmail.com, herre@cs.uu.nl,
jasper.lindenberg@tno.nl

Abstract. This research describes an approach to objective assessment of mental workload, by analyzing differences in pupil diameter and several aspects of eye movement (fixation time, saccade distance, and saccade speed) under different levels of mental workload. In an experiment, these aspects were measured by an eye-tracking device to examine whether these are indeed indicators for mental workload. Pupil diameter and fixation time both show a general significant increase if the mental workload increases while saccade distance and saccade speed do not show any significant differences. This assessment of mental workload could be a trigger for aiding the operator of an information system, in order to meet operational requirements.

Keywords: mental workload, adaptive automation, eye movement, pupil diameter, saccade, fixation time.

1 Introduction

In 1988, in the straight of Hormuz, the USS Vincennes mistakenly shot down a commercial Iranian airplane because it was misidentified as an Iranian F-14 combat airplane. Much is written [1] about contributing factors that lead to this tragic accident and some authors summarize it as human error. This is a typical example of a high-risk professional domain where humans carry large responsibility and where mistakes result in tragic accidents and/or heavy losses. In these information-rich and dynamic environments, a competition for the human's attention is going on between numerous different information items, at times leading to a cognitive overload. This overload originates from the limitations in human attention and constitutes a well-studied bottleneck in human information processing. If the human is getting overloaded, a control mechanism capable of adjusting the balance of work between the human operator and the machine might lower the cognitive burden of the human and in effect optimize the performance of the human machine ensemble. A so-called adaptive system [2] in which the division of labor between human and machine is flexible and

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 219–228, 2009.

responsive to task or human demands, is thought to represent a better solution to the problem of function allocation than the static ones currently in use. Rouse [3] introduced adaptive aiding as an initial type of adaptive automation and stated that adaptive aiding is a human-machine system-design concept that involves using aiding/automation only at those points in time when human performance needs support to meet operational requirements [3]. Whether one uses the terms adaptive automation, dynamic task allocation, dynamic function allocation, or adaptive aiding, they all reflect the dynamic reallocation of work in order to improve human performance or to prevent performance degradation.

As a matter of fact, adaptive automation should scale itself down when things become quieter again and the goal of adaptive automation could be stated as trying to keep the human occupied within a band of 'proper' workload [see 4]. Periods of 'underload' can have equally disastrous consequences as periods of overload due to slipping of attention and loss of situational awareness.

As stated, an adaptive system changes the division of labor between the human and machine as the effectiveness of a human operator is a concern in relation to the task demands. Parasuraman [5], for example, found superior performance when the control of a fault management task (i.e., to monitor an automated system and diagnose the problem in case the system halts) was allocated back to the human for some time. Other studies shift control from the human to the machine in case the human is incapable or indecisive to make a decision as seen in the (no)go decision in case of a engine failure in the takeoff run of an airplane [6]. More recently, adaptive automation is applied to the domain of naval warfare [7] where part of the identification of airplanes or vessels is executed by the system when the human starts to fall behind. In synopsis, a number of studies have shown that the application of adaptive automation enhances performance, reduces workload, improves situational awareness, and maintains skills that are deteriorating as a consequence of too highly automated systems [5-9].

One of the challenging factors in the development of successful adaptive automation concerns the question of when changes in the level of automation must be effectuated. Wilson and Russell [10] define operator psychophysiology as one of five triggering strategies based on an previous division by Parasuraman *et al* [5]. Psychophysiological data from the operator are employed in various studies [9-13] and prove an objective measure. Examples of these measurements are: heartbeat rate, respiratory, facial expressions, perspiration, eye blink rate (see [14] for an overview). Although various studies [9-13] indicate a mental workload effect on psychophysiological characteristics, no single psychophysiological measure can be directly interpreted as such [15]. Variations in psychophysiological measurements, however, can be assigned to a lot of different aspects, with mental workload just being one of them. The main advantages of objective measurements are that they do not have to interrupt the operator in task execution.

One popular type of psychophysiological data measurement involves workload effects on properties and movement of the eye [16-20]. Although a number of studies found [17-19] empirical evidence in the favor of utilizing an increased pupil diameter as an of increased mental workload, not all studies have obtained similar results. Kramer [21], for example, relates the failure to find similar results to factors unrelated to the task that produce larger changes in pupil dilatation such as changes in ambient illumination or screen luminance.

Alternatively, other relations between mental workload and properties of the eye exist. During visual scanning, muscles direct the eye to interesting areas where fixations occur. A fixation is usually defined as a steady focus of the eye for 100 to 200 milliseconds, which provides the visual system with detailed input about the visual stimulus. Simultaneously, parallel processes use peripheral visual information to determine where the next fixation will be located [22]. The movement to another fixation stimulus is defined as a saccade. Tole, Harris, Stephens, and Ephrath [20] found an increase in fixation time when the mental workload increased. However, saccade measurements show no consequent results in relation to mental workload [16, 18].

The previous paragraphs clearly indicate a challenge in utilizing eye properties as indicator for mental workload. One side of the scientific literature shows that pupil diameter, fixation times, saccade distance, and saccade speed can be used as an indicator of mental workload while other studies show counterarguments. We are interested how the properties of the eye respond in various workload conditions in a naval warfare domain. Once successful, these properties of the eye can be used to trigger adaptive automation.

We conducted an experiment where certain properties of eye movement and pupil diameter were measured under different levels of mental workload. Consequently we question whether fixation time, saccade distance, saccade speed, and pupil diameter can be used as objective indicator for mental workload in such a task setting.

As the study evolves around a measure of mental workload, we will manipulate mental workload using a validated model of cognitive task load. Therefore our first hypothesis reads that:

1: three scenarios are generated having a predicted and different mental workload.

Using these differences in mental workload we can perform measurements on properties of the eye. Following experimental effects on pupil diameter and fixation times found in respectively [17-19] and [20], we hypothesize that:

2: if the mental workload of an operator increases, pupil diameter increases, and
3: if the mental workload of an operator increases, fixation time increases.

Furthermore, it is expected that saccade distance will decrease in response to an increase of mental workload due to an effect called tunnel vision (i.e. the loss of peripheral vision with retention of central vision, resulting in a constricted circular tunnel-like field of vision). Also, saccade speed is expected to decrease due to fatigue of the muscles, as evidence for fatigue in pupillary muscles exists [23]. Consequently, we hypothesize that

4: if the mental workload of an operator increases, saccade speed decreases, and
5: if the mental workload of an operator increases, saccade distance decreases.

The next section discusses the method & materials used in this study and section 3 presents the results of the experiment. Section 4 discusses the results and draws conclusions from these results from the perspective of adaptive automation.

2 Research Method

2.1 Participants

Eighteen subjects participated in the experiment and were paid to participate. The test subjects were all university students, with a good knowledge of English. The participant group consisted of ten men and eight women. They had an average age of 25, with a standard deviation of 5.1.

2.2 Experimental Tasks

The subjects were given the role of human operators of (an abstracted version of) a combat management workstation (CMS) aboard naval vessels. The workstation comprised a schematic visual overview of the nearby area of the ship on a computer display, constructed from the data of radar systems. On the workstation the subject could manage all the actions required to achieve mission goals.

More specifically, the goal of the human operator during the scenarios was to monitor, classify, and identify every track (i.e. airplanes and vessels) within a 38 nautical miles range around the ship. Furthermore, in case one of these tracks showed hostile intent (in this simplified case a dive toward the ship), they were mandated to protect the naval vessel and eliminate the track.

To achieve these goals, the subject was required to perform three tasks. First, the classification task gained knowledge of the type of the track and its properties using information from radar and communication with the track, air controller, and/or the coastguard. The subject could communicate with these entities using chat functionality within the CMS. The experiment leader responded to such communications. The second task was the identification process that labeled a track as friendly, neutral, or hostile. The last task was weapon engagement in case of hostile intent as derived from certain behavior. The subject was required to follow a specific procedure to use the weapons.

2.3 Scenarios

We designed three different scenarios, each implying a different cognitive task load. The task loads were under-load, normal load, and an overload achieved by manipulating two of the three cognitive task load factors as defined in Neerincx model [24] of cognitive task load (CTL).

The CTL model is comprised of three factors that have a substantial effect on the cognitive task load. The first factor, percentage time occupied, has been used to assess workload for time-line assessments. The second load factor is the level of information processing that addresses cognitive task demands. The model therefore incorporates the skill-rule-knowledge framework of Rasmussen [26] where the knowledge-based component involves the highest workload. To address the demands of attention shifts, the model distinguishes task-set switching as a third load factor. It represents the fact that a human operator requires time and effort to reorient himself to a different context. These factors present a three-dimensional space in which all human activities can be projected as a combined factor. Specific regions indicate the cognitive demands activities impose on a human operator.

Creating scenarios using the CTL mode has been applied successfully in a number of experimental [27] and realistic [28] settings. We applied the model to implement a certain cognitive load. First, the total number of tracks in a scenario was changed. If many tracks are in the observation range, the percentage of the total time that the human is occupied is high. Second, a larger amount of tracks that show special behavior and more ambiguous properties increases the operator's cognitive workload due to applying more rule and knowledge based reasoning. It forces the human operator to focus attention and to communicate more in order to complete the tasks.

2.4 Variables and Experimental Design

In order to control for intra-individual variability in cognition, we chose to use a within-subject design. In order to limit the potential for individual differences on any experimental condition and to filter out sequence effects, we applied a Latin square design to the combination of independent variables and the sequence of scenarios.

The independent variable was the workload as manipulated in a scenario (see previous section). Furthermore, five dependent variables were measured:

- Mental workload was measured and controlled using an adapted version of a workload watch [25] that signaled the subject every 100 seconds to rate his/her perceived workload on a scale (one to five), by clicking on the corresponding button present on the lower right of the screen. Button 1 indicated low workload, button 3 normal workload and button 5 high workload. The buttons in between indicate intermediate levels of workload.
- Pupil diameter in micrometers during the trials (averaging both eyes).
- Fixation time: the time that fixations lasted within a radius of 40 pixels and a minimum of 100 milliseconds. The fixation times were divided by the total number of fixation to derive the average time a fixation lasts and only the fixations on a track were accounted to get a reliable representation of cognitive processing time.
- Saccade distance: the distance in pixels between one fixation and the next.
- Saccade speed defined as the saccade distance divided by the saccade time.

2.5 Apparatus

The CMS application was run on a computer connected to a 17-inch monitor, with a resolution 1280x1024. The eye-tracking device Tobii X50 was connected to another computer. The experimental leader was situated behind a desk with a third computer running the same scenario as the test person to ensure good communication.

The Tobii X50 recorded the required aspects of eye movement and pupil diameter of the subjects during the task. The Tobii X50 was placed in front and underneath the monitor that the subjects used for the task. This was approximately 60 cm in front of them.

2.6 Procedure

Before the experiment, the subjects were given a clear description of the various tasks to be executed during the scenarios and various test round were offered to the subjects. Before every scenario, a description about the position of the naval ship and its mission was provided. The experiment was conducted in a closed room where the subjects were not disturbed during the task. During the experiment, an experimental leader was situated roughly two meters behind the subject to assist when necessary.

3 Results

For each dependent variable a repeated-measures analysis ANOVA with within-factor scenario was used to analyze the data as the subjects are exposed to each condition in turn. In all cases, an alpha level of .05 was used to determine statistical significance. The data were analyzed using SPSS. For post-hoc analysis, the least significant difference (LSD) test was used and the partial eta square statistics (η_p^2) was adopted to describe the estimated proportion of variance explained by the factors. The partial eta square has the advantage that it is independent on the number of factors.

Verification of the mental workload used data of all but one subject due to a failure in logging (N = 17). The rest of the statistical tests utilized N = 13 because the eye-tracker data from four subjects could not be used due to a technical failure of the eye-tracking device in one of the three scenarios.

3.1 Workload Verification

Repeated-measures ANOVA reveals a significant effect in subjective (indicated) mental workload between the three scenario's (F(2, 33) = 190.632, $p < .001$, $\eta_p^2 = .923$). Least square difference post-hoc analysis reveals that all three means were significantly different ($p < .05$). Compared to the under-load scenario, the perceived mental workload was significantly higher in the normal workload scenario. In turn, the perceived mental workload in the overload scenario was significantly higher again than in the normal-workload scenario (see Fig. 1).

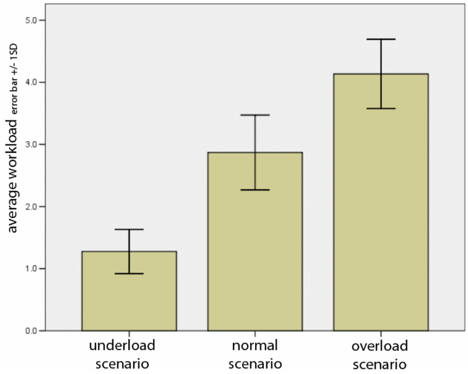

Fig. 1. The subjective workload per scenario as indicated every 100 seconds on a five point scale

3.2 Pupil Diameter

Repeated measures one-way ANOVA reveals that there are significant differences in pupil diameter between the three workload scenario's, F (2, 69) = 3.720 , $p < .005$, $\eta_p^2 = .237$). A post-hoc LSD comparisons revealed that the underload scenario and the overload scenario were significantly different. Compared to the underload scenario (M = 4957.42, SD = 511.164), the average pupil diameter was significantly higher in the overload scenario (M = 5094.90, SD = 547.797). Furthermore the normal scenario (M = 5069.84, SD =152.676) had a significant higher average pupil diameter in comparison with the underload scenario. However, no significant differences were found between normal scenario and the overload scenario.

3.3 Fixation Time

SPSS repeated measures one-way ANOVA displays a significant effect of fixation time on the different types of scenario (F (2, 69) = 13.411 , $p < .05$, $\eta_p^2 = .528$). LSD comparisons revealed that all scenario's were significantly different from each other. Compared to the underload scenario (M = 249.177, SD = 113.192), the average fixation time was significantly higher in the normal scenario (M = 287.992, SD =108.710) and on its turn the normal scenario differed significantly from the overload scenario (M = 334.838, SD = 130.823).

3.4 Saccade Distance and Speed

No significant differences are found with a repeated measures one-way ANOVA in saccade distance between the three scenario's (F (2, 69)= 0.388, p = .683, $\eta_p^2 = .031$). In addition, no significant differences are found with a repeated measures one-way ANOVA in saccade speed between the three scenario's (F (2,69) = 1.768, p = .192, $\eta_p^2 = .128$).

4 Conclusion and Discussion

In high-risk professional domains, like a naval warfare, it is preferred to keep humans in control to handle unanticipated novel situations. However, increased mental workload might decrease human performance resulting in endangering mission goals. Therefore it is important to detect such situations. Various studies suggest that the machine should aid the human operator in those situations to meet operational requirements [3]. However the measurement of such situations is much debated and this research describes an approach to assess mental workload objectively by analyzing differences in pupil diameter, fixation time, saccade distance, and saccade speed under different levels of mental workload. We consequently conducted an experiment to measure these four aspects under various workload conditions to ascertain their utility as an objective mental workload indicator.

The results show that the manipulation of the scenario's worked as expected in that the manipulation of two of the three CTL model variables resulted in significant different subjective mental workload. The results not only confirm hypothesis 1 but also extend the knowledge [27, 28] on the successful application of the CTL model to design scenarios with an intended mental workload.

Although the results reveal a general significant effect of pupil diameter on workload manipulations, the results fail to discriminate between all workload manipulations. We therefore partly accept hypothesis 2 because of the general nature to utilize pupil diameter as an indicator of workload. Consequently we agree with [21] that the pupil diameter responds to many factors with workload being one of them in contrasts other research that found pupil diameter effects [17-19].

Results show, on the other hand, a significant and discriminatory effect of mental workload manipulations on fixation time. This means that we accept hypothesis 3 because fixation time can be used to distinguish between workload conditions. The results comply with research by Tole [20] who found an increase in fixation time when the mental workload increased and [29] found a negative correlation between fixation time and performance. This complies with the findings in this research, given that higher fixation times indicate a higher level of mental workload and mental workload has a negative effect on performance.

Saccade distance and saccade speed show no significant differences when the mental workload increases and we therefore reject hypothesis 4 & 5. These results comply with research [16] that failed to find a relation between saccade measures and mental workload. However, [18] did find a decrease in saccade distance if mental workload increases but stated that many properties of the eye, including saccades, are highly task dependent.

Much research has been done to find a relation between operator psychophysiology and operator workload. As stated before, no single psychophysiological measure can be interpreted as a workload indicator. Therefore, in order to obtain a reliable objective indicator for mental workload, it is necessary to work towards a model that integrates several psychophysiological measurements. The construction of such a model is complex because the reliability and relative importance of the different measurements are hard to define. The results of this research contribute to this model by examining the effects on pupil diameter, fixation time, saccade distance, and saccade speed under different levels of mental workload in a naval warfare setting.

As this research shows that measurements of properties of the eye with an eye-tracking device can provide a valuable addition to the determination of the level of aiding, problems arise when it comes to the practical application of the concept. As indicated throughout the paper, many factors influence workload and properties of the eye. In an experiment, these aspects can be kept as constant as possible. In a practical application, for example with defense tasks on a navy ship, it cannot be expected of the human operator to refrain from drinking caffeine-holding beverages. These aspects make the deduction of mental workload from for example pupil diameter unreliable. However, if a combination of psychophysiological measurements is used and they all indicate a similar operator workload, this indication can be very usable.

Acknowledgements

This study was supported by the MOD-NL DR&D organization under program 'human system task integration' (V206). The department of Human in Command of TNO Defence, Security and Safety kindly provided an internship fee.

References

1. Cannon-Bowers, J.A., Salas, E.: Making decisions under stress: implications for individual and team training. American Psychological Association, Washington (1998)
2. Scerbo, M.: Theoretical perspectives on adaptive automation. In: Parasuraman, R., Mouloua, M. (eds.) Automation and human performance: theory and applications, pp. 37–63. Lawrence Erlbaum Associated Publishers, Mahwah (1996)
3. Rouse, W.B.: Adaptive Aiding for Human/Computer Control. Human Factors 30, 431–443 (1988)
4. Endsley, M., Kiris, E.: The Out-of-the-Loop Performance Problem and Level of Control in Automation. Human Factors, 381–394 (1995)
5. Parasuraman, R., Mouloua, M., Molloy, R.: Effects of adaptive task allocation on monitoring of automated systems. Human Factors 38, 665–679 (1996)
6. Inagaki, T.: Situation-adaptive autonomy for time-critical takeoff decisions. International Journal of Modelling and Simulation 20, 175–180 (2000)
7. Arciszewski, H.F.R., de Greef, T.E., van Delft, J.H.: Adaptive Automation in a Naval Combat Management System. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans (in press)
8. Moray, N., Inagaki, T., Itoh, M.: Adaptive Automation, Trust, and Self-Confidence in Fault Management of Time-Critical Tasks. Journal of experimental psychology, 44–57 (2000)
9. Bailey, N.R., Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Scott, L.A.: Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. Human Factors 48, 693–709 (2006)
10. Wilson, G.F., Russell, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Human Factors 49, 1005–1018 (2007)
11. Byrne, E.A., Parasuraman, R.: Psychophysiology and adaptive automation. Biological Psychology 42, 249–268 (1996)
12. Veltman, J.A., Gaillard, A.W.K.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41, 656–669 (1998)
13. Prinzel, L.J., Freeman, F.G., Scerbo, M.W., Mikulka, P.J., Pope, A.T.: A closed-loop system for examining psychophysiological measures for adaptive task allocation. International Journal of Aviation Psychology 10, 393–410 (2000)
14. Weert, J.C.M.: Ship operator workload assessment tool. Department of mathematics and computer science. Technical University Eindhoven, Eindhoven (2006)
15. Veltman, J.A., Gaillard, A.W.D.: Pilot workload evaluated with subjective and physiological measures. In: Brookhuis, K.A., Weikert, C., Moraal, J., de Waard, D. (eds.) Proceedings of the European Chapter of the Human Factors and Ergonomics Society, Soesterberg, the Netherlands (1996)
16. Brookings, J.B., Wilson, G.F., Swain, C.R.: Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology 42, 361–377 (1996)
17. Iqbal, S.T., Zheng, X.S., Bailey, B.P.: Task-evoked pupillary response to mental workload in human-computer interaction. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, Vienna, Austria, pp. 1477–1480 (2004)
18. van Orden, K.F., Limbert, W., Makeig, S., Jung, T.: Eye activity correlates of workload during a visuospatial memory task. Human Factors 43, 111–121 (2001)
19. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin 91, 276–292 (1982)
20. Tole, J.R., Harris, R.L., Stephens, A.T., Ephrath, A.R.: Visual scanning behavior and mental workload in aircraft pilots. Aviation, Space, and Environmental Medicine 53, 54–61 (1982)

21. Kramer, A.F.: Physiological metrics of mental workload: A review of recent progress. In: Damos, D.L. (ed.) Multiple-task performance, Taylor & Francis, London (1991)
22. Henderson, J.M.: Visual attention and eye movement control during reading and picture viewing. In: Rayner, K. (ed.) Eye movements and visual cognition - Scene perception and reading, pp. 260–283. Springer, New York (1992)
23. Stern, J.A.: The pupil of the eye: what can it tell about mental processes? Human engineering for quality and life 8, 1–2 (1997)
24. Neerincx, M.A.: Cognitive task load design: model, methods and examples. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Design, pp. 283–305. Lawrence Erlbaum Associates, Mahwah (2003)
25. Boer, L.C.: Workload-watch as an element of human engineering testing and Evaluation. In: Eleventh ship control systems symposium, vol. 2, Computational mechanics publications, Southampton, United Kindom (1997)
26. Rasmussen, J.: Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering. North-Holland, Amsterdam (1986)
27. Neerincx, M.A., Besouw, N.J.P.: Cognitive task load: a function of time occupied, level of information processing and task-set switches. Industrial Ergonomics, HCI, and Applied Cognitive Psychology 6, 247–254 (2001)
28. Grootjen, M., Neerincx, M.A., Veltman, J.A.: Cognitive task load in a naval ship control centre: From identification to prediction. Ergonomics 49, 1238–1264 (2006)
29. Goldberg, J.H., Kotval, X.P.: Eye movement-based evaluation of the computer interface. In: Kumar, S. (ed.) Advances in occupational ergonomics and safety, pp. 529–532. ISO press, Amsterdam (1998)