Performance using low-cost gaze-control for simulated flight tasks

Ulrika Ohlander[1], Oscar Linger[2], Veronica Hägg[2], Linn Nilsson[2], Åsa Holmqvist[2], Sandra Durefors[2], Jens Alfredson[1,2], & Erik Prytz[2]
[1]Saab AB, [2]Linköping University
Sweden

In the current study, interaction using gaze control was compared to computer mouse using the MATB-II (Multi-Attribute Task Battery) environment. The study had two aims; the first was to explore the utility of low-cost technologies in a rapid prototyping and testing environment for aviation. The second aim was to use such an environment to compare a novel interaction device (a low-cost gaze control device) to a familiar interaction device (computer mouse). *Method*: Thirty participants performed two scenarios with each interaction device. The software MATB-II provided simulated flight tasks and recorded performance. Mental workload was assessed by the NASA Task Load Index (TLX) questionnaire after each scenario. *Results*: The results showed that gaze control resulted in significantly higher overall mental workload than computer mouse. Performance was better with mouse in two of the four tasks. *Conclusions*: Concerning the first aim, the study demonstrated the value of low-cost technology for initial user testing before using more expensive high-fidelity environments. Concerning the second aim, the computer mouse resulted in better performance and lower mental workload. This may either be due to higher user familiarity with computer mouse interaction or to limitations of the gaze control equipment and insufficient adjustments of the interface design to optimize for gaze control.

Introduction

Fighter pilots in air combat are exposed to both high G-loads as well as high workload. Pilots often need both hands to control the aircraft during fast-paced scenarios. Moreover, during flight with high G-loads it is difficult for the pilot to move their arms and hands to interact with the cockpit interface. Today, the main solution to these situations is HOTAS control, Hands on Throttle and Stick, which give the pilot access to control devices on the joystick and throttle as well as the possibility to interact with tactical systems. To a certain extent, voice control is also available as a solution today. As the flow of information from the system to the pilot is constantly increasing, and the tactical systems gets more complex, the need for the pilot to interact with the system is also increasing. This difference in bandwidth between the computer and the human creates a demand to utilise all possible means of interaction between the pilot and the system, and to investigate and explore new possible enabling technical solutions (Jacob, 1991). The technology for interacting

In D. de Waard, A. Toffetti, R. Wiczorek, A. Sonderegger, S. Röttger, P. Bouchner, T. Franke, S. Fairclough, M. Noordzij, and K. Brookhuis (Eds.) (2017). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

with graphical user interfaces using eye gaze is rapidly improving and eye tracking has now become a feasible interaction method to explore further. Using eye movements to control a pointing device is typically seen as a simple and intuitive task that can be performed with high speed (Drewes, 2010; Penkar, Lutteroth, & Weber, 2013). In a simulator study, O´Connel, Castor, Pousette & Krantz (2012) demonstrated that target designation using eye tracking was beneficial. However, additional research is needed to further explore the benefits and limitations of eye gaze as an interaction method in flight decks. The current study was conducted to investigate the impact of eye gaze as input device on workload and performance in a simulated flight scenario in MATB-II (Multi-Attribute Task Battery).

Simulations of complex environments such as flight decks are often complicated and costly to perform. MATB-II is intended to provide the same type of workload as for a pilot in a real airplane using only a desktop PC and gaming hardware for controls, which means a significantly lower cost. The purpose of this study was to examine gaze-control using a commercially available eye tracker, the Eye Tribe, in this type of low-cost simulated environment. The objectives were to investigate the interaction and measure workload when using eye-gaze in a simulated flight deck, and at the same time explore the possibilities to achieve a low-cost testing environment with MATB-II and commercial gaming controls.

In this study, eye-tracking performance was compared to ordinary computer mouse control. Using a mouse is obviously not how a pilot interacts with cockpit systems, but the participants in the study can be assumed to have a high familiarity with mouse control. Thus, the comparison can be regarded as being between the new method, eye-gaze, and a commonly used controlling device for the target group (Rupp, Oppold, & McConnell, 2013).

Mental Workload and Performance

Gopher and Donchin (1986) define mental workload as a function of the mental capacity that is necessary for information and decision processing, and the capacity available in the current situation. A task that is difficult for one person can be considered easy for another. One explanation of this is experience, which reduces the mental workload (Wickens, Hollands, Banburry, & Parasuraman, 2012). Mental workload is often measured by using subjective methods (Charlton, 1996). Subjective measurements have several practical advantages, such as easy implementation and providing detailed data for perceived load (Eggemeier & Wilson, 1991). The NASA Task Load Index (Hart & Staveland, 1988) is one of the most commonly used subjective methods for measuring workload. Mental workload can affect human performance, such that workload that exceeds the operator’s coping capacity can lead to sharp performance degradation (Hancock & Warm, 1989). However, performance can be protected against the increase in workload through, for example, increased effort up to a certain limit (Hancock & Warm, 1989; Hockey, 1997). Svensson, Angelborg-Thanderz, Sjoberg, and Olsson (1997) showed that task difficulty is positively correlated to workload, and that workload negatively correlated with fighter pilot performance in a high-fidelity simulated flight deck.

Method

A repeated measures design was used with the independent variables being the control method and task difficulty, and the dependent variables being workload and performance.

Participants

Thirty participants (19 male, 11 female) were recruited to participate in this study. They ranged in age from 19 to 43 years with a mean age of 24.2 (SD = 5.8). All reported normal or corrected-to-normal vision with no color vision deficiencies. All reported that they normally used their right hand for computer mouse interaction. None had prior experience with MATB-II.

Apparatus

MATB-II (NASA, 2014) is a computer program that simulates tasks in cockpits in order to evaluate operator performance and workload (Comstock Jr & Arnegard, 1992). The original MATB has been modified to MATB-II and the major changes include adaption to later versions of Windows operating systems and new graphical interfaces (Santiago-Espada, Myer, Latorella, & Comstock Jr, 2011). There are four different tasks for the user: System Monitoring Task, Tracking Task, Communications Task and Resource Management Task. The program is mainly developed to simulate tasks that correspond to a pilot’s mental workload during flight, but can be used for other purposes such as measuring general multitasking capacity (Aricò et al., 2014; Rupp et al., 2013).

Modifications in the MATB-II interface were made to allow tasks to be performed using gaze control. As the eyes constantly make small flickering movements (saccades) there can be problems using eye control to target small buttons in graphical user interfaces with high precision (Drewes, 2010; Yamato, Monden, Matsumoto, Inoue, & Torii, 2000). According to Drewes (2010), it is difficult to achieve the same precision pointing with the eyes, as is possible with a computer mouse and it might therefore not be appropriate to replace a computer mouse with gaze control without a customized user interface. It has been suggested that the interface should be adjusted by enlarging buttons to make the eye control work effectively (Yamato et al., 2000). In this study, all the elements that the user could click on were enlarged. One part of the Communications Task was removed because the level of precision required was not manageable with gaze control. Two of the, authors who were not involved in implementing the modifications, were asked to test the interface with the eye-tracker to verify that the new interface could be used with gaze control. The same modified interface was used for all test cases regardless of interaction method.

The eyes are naturally used to search for information. The approach used in this study was that the main use for eye-gaze should be in visual search and selection tasks, while tasks like acknowledging and activating should be done using other modalities (Kumar, 2007). For the gaze control condition, the space bar on the keyboard was therefore used as a select or confirm button. This was also done to avoid the “Midas-touch” problem, i.e. that the user would inadvertently interact with everything he/she looks at. MATB-II records several performance measures from the participants. The measures used in this study were: response time and correct responses in percent from the System Monitoring Task, deviation from the centre point in the Tracking Task, and the deviation from target fuel units (2500) in the Resource Management Task. Perceptual sensitivity, the discriminability index d’ (Green & Swets, 1974), was calculated based on the hits, misses, correct rejections and false alarms from the Communication Task.

Gaze control was implemented using an Eye Tribe device (Eye Tribe). This device consists of an eye-tracking bar mounted below a computer screen. The eye-tracker’s update frequency is 30-75 Hz, and the stated accuracy is 0.5-1° visual angle.

Procedure

The participants first completed an informed consent form and a demographics questionnaire. Next, an instruction video (ca. 8 minutes long) was used to introduce the participants to the different tasks of MATB-II. After the video, the instructor explained the NASA TLX form and provided a guide to use when completing the questionnaire later. The participants completed two training scenarios in MATB-II, the first time using the mouse to interact with the program and the second time using the gaze control equipment. The Eye Tribe equipment was calibrated and re-calibrated to the user before each gaze control scenario. A joystick was used for the tracking task in both the mouse and the gaze control conditions. The participants were encouraged to ask clarifying questions during the practice scenarios. After the practice scenarios, the participants completed four different scenarios:

- Gaze control with low task difficulty (GC-L).
- Gaze control with high task difficulty (GC-H).
- Mouse control with low task difficulty (M-L).
- Mouse control with high task difficulty (M-H).

The length of each scenario was five minutes. Task difficulty was manipulated by the length of the response windows and number of task interactions. The more difficult conditions requirered faster responses and more input from the participant. The response time for the scenarios was 20 s for the low task difficulty and 15 s for the high task difficulty. The order of the scenarios was blocked by control condition and counterbalanced across participants. The participants completed a NASA TLX questionnaire at the end of each scenario.

Results

Three participants failed to answer the NASA TLX within the 30-second time limit. This resulted in mental workload data from 27 participants and performance data from 30 participants.

Workload

The experienced mental workload was significantly higher with gaze control than with computer mouse, $F(1, 26) = 102.935$, $p < .001$ partial $\eta^2 = 0.798$ (see Figure 1). There was no significant difference between low and high task difficulty, $F(1, 26) = .294$, $p = .592$, partial $\eta^2 = 0.011$, and no significant task difficulty and control mode interaction, $F(1, 26) = .263$, $p = .613$ partial $\eta^2 = 0.010$.

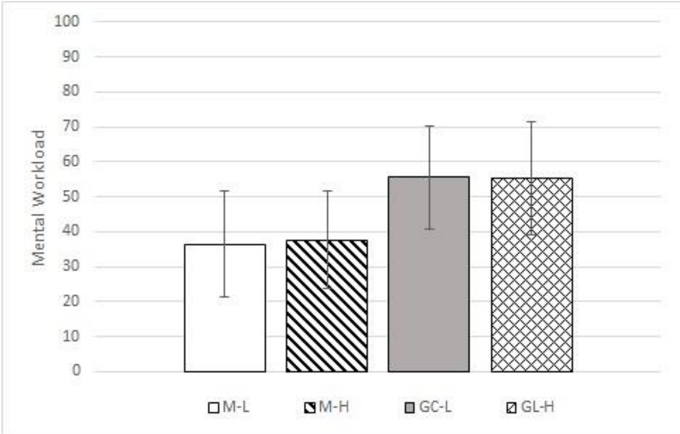

Figure 1. Mental workload as assessed with the total score of the NASA TLX for the different scenarios. M=Mouse control, GC= Gaze control, L = Low, H= High task difficulty

Performance

There was no significant difference between gaze control and computer mouse interaction for d'-scores in the communication task, $F(1, 29) = .214$, $p = .647$, partial $\eta^2 = 0.007$ (see Figure 2). The participants did however perform significantly better in the high task difficulty scenarios than the low task difficulty scenarios, $F(1, 29) = 63.076$, $p < .001$, partial $\eta^2 = 0.685$. There was no significant interaction between control mode and task difficulty, $F(1, 29) = 3.52$, $p = .558$, partial $\eta^2 = 0.012$.

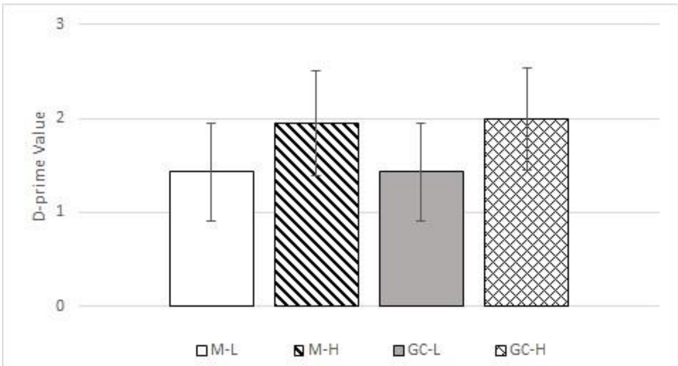

Figure 2. d' in the communications task. M=Mouse control, GC= Gaze control, L = Low, H= High task difficulty

There was no significant difference between gaze control and computer mouse interaction for fuel deviation in the resource management task, $F(1, 29) = .704$, $p = .408$, partial $\eta^2 = 0.024$, nor between high and low task difficulty, $F(1, 29) = 1.120$, $p = .299$, partial $\eta^2 = 0.037$, nor a control mode by task difficulty interaction, $F(1, 29) = .021$, $p = .886$ partial $\eta^2 = 0.001$.

For the system monitoring task the percentage of correct responses was significantly lower with gaze control than mouse control, $F(1, 29) = 11.768$, $p = .002$, partial $\eta^2 = 0.289$ (see Figure 3). However, there was no effect of task difficulty, $F(1, 29) = .494$, $p = .488$ partial $\eta^2 = 0.017$, nor a control mode by task difficulty interaction, $F(1, 29) = 0.019$, $p = .893$ partial $\eta^2 = 0.001$.

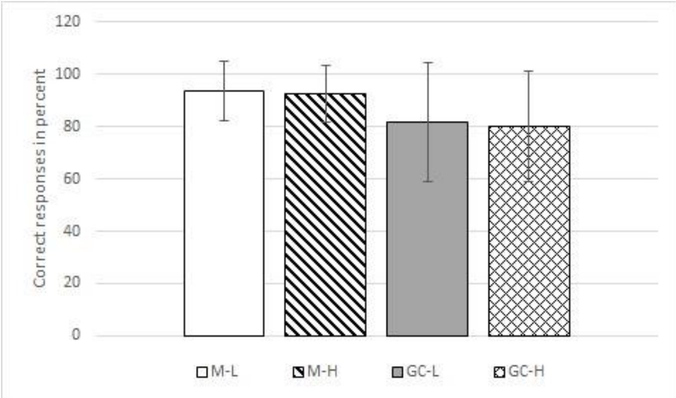

Figure 3. Correct responses in percent for the system monitoring task. M=Mouse control, GC= Gaze control, L = Low, H= High task difficulty

The response time for the system monitoring task was significantly higher for gaze control compared to mouse control, $F(1, 29) = 29.85$, $p < .001$, partial $\eta^2 = 0.507$ (see Figure 4). The response time was also significantly lower in the high task difficulty scenarios as compared to the low task difficulty scenarios, $F(1, 29) = 15.035$, $p = .001$, partial $\eta^2 = 0.341$. There was no interaction between control mode and task difficulty for the response time, $F(1, 29) = 1.967$, $p = .171$ partial $\eta^2 = 0.064$.

For the tracking task there was a significantly higher deviation from the centre for the gaze control mode than the mouse control mode, $F(1, 29) = 37.693$, $p < .001$, partial $\eta^2 = 0.565$ (see Figure 5). There was no effect of task difficulty, $F(1, 29) = 4.111$, $p = .052$ partial $\eta^2 = 0.124$, nor a control mode by task difficulty interaction, $F(1, 29) = .366$, $p = .550$ partial $\eta^2 = 0.012$.

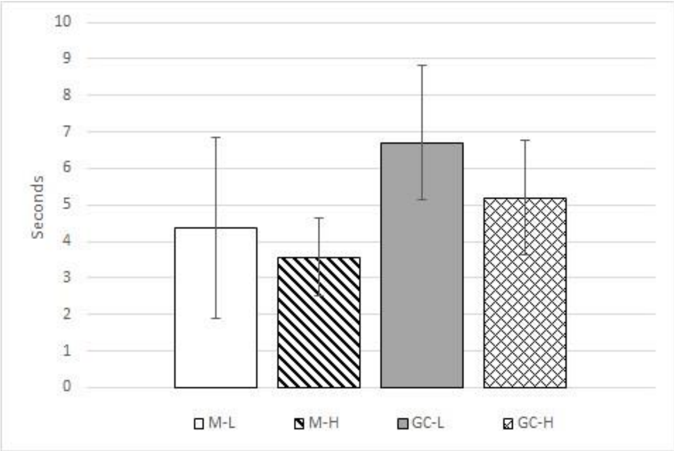

Figure 4. Response time in seconds for the system monitoring task. M=Mouse control, GC= Gaze control, L = Low, H= High task difficulty

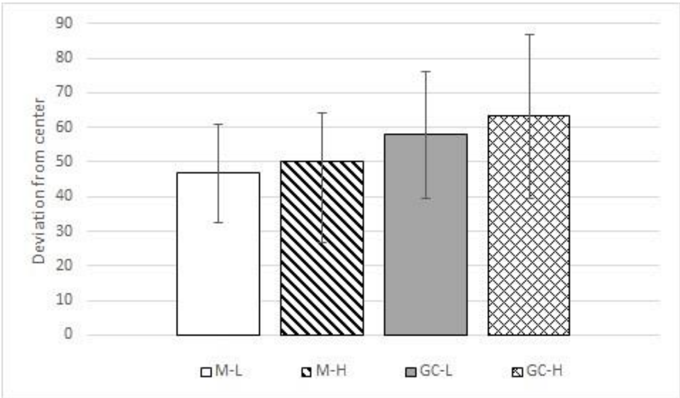

Figure 5. Deviation from centre for the tracking task. M=Mouse control, GC= Gaze control, L = Low, H= High task difficulty

Discussion

The results show that gaze control was associated with higher mental workload in all scenarios and decreased performance in two out of four scenarios in comparison with mouse control. These results indicate that gaze control was inappropriate for the tasks used in the current study. One reason could be that the eyes were used for both information retrieval and control, whereas the manual input was used only for control. Gaze control could therefore have led to resource conflicts (Wickens, 1980). Another reason could be that the interface was not optimally adapted for gaze control. Some adjustments with regard to the size of user control areas (e.g., buttons) were made but these were not subject to rigorous validation.

The participants did not rate the high task load scenarios as producing significantly higher workload, and performance was better for the high task load scenarios for the auditory and system monitoring task. This could indicate that the experimental manipulation of taskload did not work as intended and did not increase the taskload appropriately. An alternative hypothesis as to why the response times were longer in the low task load scenarios than in the high task load scenarios could be a speed accuracy trade off (Wickens et al., 2012). When the demand for multitasking was lowered in the easy scenarios the participants could have focused more on the tasks demanding static input (i.e. resource management task and tracking task), as well as prioritizing precision over speed.

Regarding the objective to explore the use of a low cost environment for testing and research, the study shows promising results. The tested equipment and software turned out to be useful and a first step towards building a simulation environment for early evaluation of new concepts.

Future studies

Multiple resource theory (Wickens, 1980) describes how using different modalities may lower mental workload, and future studies could therefore be conducted to investigate the possibility of controlling some elements in MATB-II using gaze control and others using mouse control. This could explore the possibility of utilizing the advantages of gaze control complementary to the mouse control. Another modality to consider is vocal control, were voice commands could be used to perform clicks. This usage of a combination of different modalities might provide alternative action possibilities for the pilot in difficult situations.

The participants were all highly experienced in using computer mouse, which was expected, and none had previously used gaze control. Future studies could explore the effects of training and experience in using gaze control in simulated flight deck settings.

For a more complete view of the participant’s mental workload, physiological measures could be added. This more quantitative approach may provide different insights than the subjective NASA TLX method that was used in this study. Some physiological measures of interest are pupil size and heart rate variability.

Conclusions

Gaze control was associated with higher mental workload in all scenarios and lower performance in two out of four scenarios. The different levels of task load affected task performance such that the participants performed better in the high task load scenarios. Overall, the results showed that the participants performed better and reported lower mental workload using a computer mouse as compared to gaze control in the simulated flight deck.

References

Aricò, P., Borghini, G., Graziani, I., Taya, F., Sun, Y., Bezerianos, A., Thakor, N., Cincotti, F., & Babiloni, F. (2014). Towards a multimodal bioelectrical framework for the online mental workload evaluation. In *36th Annual International Conference of the IEEE Engineering in Biology and Society*, Chicago, USA.

Charlton, S.G. (1996). Mental workload test and evaluation. In T.G.O´Brien & S.G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 181-197). Mahwah, NJ: Lawrence Erlbaum and Associates.

Comstock Jr, J. R., & Arnegard, R. J. (1992). MATB - Multi-Attribute Task Battery for human operator workload and strategic behavior research *Technical Memorandum 104174*. Hampton, Virginia: NASA, Langley Research Center.

Drewes, H. (2010). *Eye gaze tracking for human computer interaction.* (Dissertation), LMU, München. (No. 11591)

Eggemeier, F.T., & Wilson, G.F. (1991). Performance-based and subjective assessment of workoad in multi-task environments. In D.L. Damos (Ed.), *Multiple-task performance* (pp. 217-278). London: Taylor & Francis.

Eye Tribe. (2016). The eye tribe tracker. Retrieved 2 February, 2016, from https://theEyeTribe.com/products/

Gopher, D., & Donchin, E. (1986). Workload-An Examination of the Concept. In K. R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2). New York: Wiley.

Green, D.M., & Swets, J.A. (1974). *Signal detection theory and psychophysics*. Oxford, UK: Robert E Krieger.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *31*, 519-537.

Hart, S.G., & Staveland, L.G. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North Holland Press.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload; a cognitive-energetical framework. *Biological Psychology, 45*, 73-93.

Jacob, R.J.K. (1991). The use of eye-movements in Human-Computer Interaction Techniques: What you Look at is What You Get. *ACM Transactions in Information Systems*, *9*, 152-169.

Kumar, M. (2007). *Gaze-enhanced User Interface Design.* (Dissertation), Stanford University, Stanford.

NASA. (2014). MATB-II. *http://matb.larc.nasa.gov/*. Retrieved 2015-05-13

O´Connell, S.D., Castor, M., Pousette, J., & Krantz, M. (2012). *Eye tracking-based target designation in simulated close range air support.* In *the Human Factors and Ergonomics Society 56th Annual Meeting*, Boston, MA.

Penkar, A.M., Lutteroth, C., & Weber, G. (2013). *Eyes only: navigating hypertext with gaze.* Paper presented at the Human-Computer Interaction INTERACT, Cape Town, South Africa.

Rupp, M.A., Oppold, P., & McConnell, D.S. (2013). Comparing the performance, workload, and usability of a gamepad and joystick in a complex task. In *the Human Factors and Ergonomics Society 57th Annual Meeting*, San Diego, CA.

Santiago-Espada, Y., Myer, R. R., Latorella, K.A., & Comstock Jr, J.R. (2011). The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User´s Guide. Hampton, VA: NASA, Langley Research Center.

Svensson, E., Angelborg-Thanderz, M., Sjoberg, L., & Olsson, S. (1997). Information complexity-mental workload and performance in combat aircraft. *Ergonomics*, *40*, 362-380.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239-257). Hillsdale, NJ: Lawrence Erlbaum.

Wickens, C.D., Hollands, J.G., Banburry, S., & Parasuraman, R. (2012). *Engineering Psychology and Human Performance* (4th ed.). New York, NY: Pearson Education Inc.

Yamato, M., Monden, A., Matsumoto, K., Inoue, K., & Torii, K. (2000). Button selection for general GUIs using eye and hand together. In *The 5th International Working Conference on Advanced Visual Interfaces*, Palermo, Italy.