Challenges and Limits of Gaze-including Interaction

Sandra Trösterer
Chair of Human-Machine Systems, TU Berlin
Franklinstr. 28-29, 10587 Berlin, Germany
str@mms.tu-berlin.de
+49 30 314-79522

Jeronimo Dzaack
Chair of Human-Machine Systems, TU Berlin
Franklinstr. 28-29, 10587 Berlin, Germany
jdz@mms.tu-berlin.de
+49 30 314-79519

ABSTRACT

Interacting with computing systems can be demanding for the user due to factors such as the complexity of the task or time pressure. So far it is unclear how new interaction techniques that include gaze as input modality are affected. In this paper we present the results of a study in which we investigated the influence of mental workload and visual distraction on gaze-including interaction. Gaze-based interaction with two different dwell times, a combination of gaze and key presses, and for control reasons mouse interaction were investigated. We found that gaze-including interaction reaches its limit in situations of increased workload, especially with visual distraction added.

Author Keywords

Gaze-based interaction, gaze-including interaction, mental workload, visual distraction.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION AND THEORETICAL BACKGROUND

Using gaze data to control computing system gains in importance in modern human-computer interaction (HCI) as it offers a way of contactless and fast interaction. Nevertheless we have to keep in mind that in a working context the requirements that are put on the user can be challenging. Time pressure, task complexity or the need for multitasking might stress the user. Therefore, designing gaze-including interaction in that context is a challenge, especially because the gaze has special features that might hinder interaction efficiency. One general problem is the so-called “Midas touch” problem, i.e. the gaze is always online, and therefore we have to be aware that everywhere we look, we might activate a command [4]. Another problem lies in the fact that, according to the theory of guided search, our visual search is influenced by task-related knowledge (top down) but also salient features of our environment (bottom up, [7]), i.e. our gaze is much more distraction-prone compared to a mouse movement. Additionally the kind of task, task know-how and task difficulty can lead to changes in eye movement behavior (see [6] for an overview). Regarding mental workload, i.e. the total amount of mental activity imposed on working memory at an instance in time [1], e.g. [5] found that imposed cognitive workload leads to a more data-driven behavior, i.e. one reacts even more to salient features. Now the question is, if our gaze is that sensitive, what will happen if we use it to interact with a computer system in situations of increased workload? How will this affect the performance and efficiency? Does a gaze-including kind of interaction reach its limits in such a context? In order to gain more insight and to derive design notes, we conducted an explorative study investigating the influence of mental workload and visual distraction on different kinds of interaction.

2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction.
February 13, 2011, Palo Alto, California, USA.

EXPERIMENTAL STUDY

Method

Participants

39 subjects (19 female, 20 male) between the ages of 21 and 38 years (mean age 26, s=3.7) participated in the study. All subjects were European and PC-users who spend approximately 4.5 hours a day working with a PC and mouse. 20 subjects had already experience with eye tracking, 9 subjects even with gaze-based interaction.

Experimental design

The study was realized as a 2x3x4 design, where the kind of workload (*between*), the level of workload (*within*) and the kind of interaction (*within*) have been varied. Basically subjects had to perform a simple searching task: Four capital letters (one target *X* and three *O*s as distractors) were presented at the corners of the computer screen and four buttons were displayed at the center of the screen. The goal was to find the target letter *X* in one of the corners of the display and to confirm its position with the corresponding button on the screen. If the target *X* was e.g. in the lower left corner, subjects had to press the lower left button. This had to be done within a time interval of 4000 ms. After a button-press or a time-out, a fixation cross was presented for 1000 ms before the next trial (see Figure 1). In every trial the *X* appeared randomly at another corner.

For one group of subjects workload was externally induced, i.e. subjects had to perform a (1) secondary acoustic task simultaneously with the simple searching task. The acoustic task was an adapted acoustic version of the Sternberg task, i.e. five random numbers between 1 and 9 were presented orally to the subject, followed by a sound and one number. Subjects then had to indicate orally whether this number was part of the previously mentioned numbers or not. A second sound indicated the beginning of the next trial. The acoustic task was independent of the searching task, i.e. the tasks were not clocked.

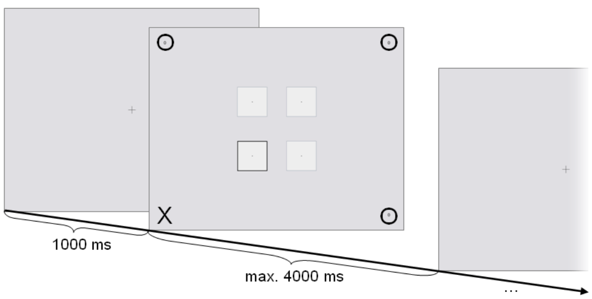

Figure 1. Searching task (*X* and *O*s are enlarged for better visibility and corresponding button is marked)

For the other group of subjects workload was induced by a (2) higher complexity of the searching task. Therefore, the fixation cross was substituted by a small arrow showing a rotation to a certain direction (right or left) and a certain amount of rotation (quarter, half, three quarters, full). Subjects had to keep in mind that rotation as it indicated the position of the correct button based on the corresponding button, i.e. if the *X* was in the lower left corner, but a quarter right rotation was shown before, the correct button to press was the upper left button.

The level of induced workload was varied in three stages. In the (1) low workload (LW) condition subjects had to perform the simple searching task (in both groups). In the (2) higher workload (HW) condition subjects had to perform the searching task with either the secondary acoustic task (group 1) or the higher task complexity (group 2). In the (3) highest workload (HTW) condition the tasks were the same as in the HW condition, but visual distraction was added, i.e. every 200ms the color of one of the four buttons randomly changed to red.

Subjects had to perform the tasks with four different kinds of interaction: gaze-based interaction with (1) 500 ms and (2) 250 ms dwell time respectively, a (3) combination of gaze to point and a key press to activate the buttons on the screen, and (4) mouse interaction. The order was permutated for each subject. Within a certain kind of interaction the order in which subjects had to perform the tasks regarding the level of workload was permutated as well. For every subject reaction times, errors (false responses, time outs), subjective experienced stress indicated on the German "Skala zur Erfassung subjektiv erlebter Anstrengung" (SEA-Skala; [3]), i.e. "scale for the acquisition of subjective experienced stress", and the gaze data were recorded. Additionally the answers to a structured interview at the end of the experimental session were transcribed.

Materials and apparatus

The experiment was realized on a PC with a standard mouse and a 19" monitor with 1280x1024 pixel resolution. The displayed buttons on the monitor had a size of 150x150 pixels, letters were 17 pixels in height and 14 pixels in width. The button size was chosen this large to allow an easy gaze-interaction and is based on empirical values of pretests. The small size of the letters was chosen to evoke visual search and avoid pop-out effects. An iView-X RED (SensoMotoric Instruments) eye-tracking unit was mounted under that monitor and was controlled by a further laptop (see Figure 2). Gaze data was tracked with 50Hz resolution and was processed in real-time using iCOMMIC (integrated COntroller for MultiModal InteraCtion; [2]), in order to transform the data to corresponding pointing and manipulation commands. Each kind of interaction was set up in this framework. For the mouse interaction the cursor was displayed in the usual way, for the gaze-including interaction forms the cursor was set half-transparent in order to reduce visual distraction by the cursor itself. For the combined condition a wireless number pad (Logitech N305, 2.4 GHz wireless connection) was used to perform the key press.

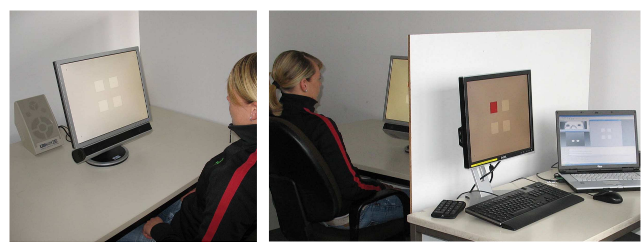

Figure 2. Experimental setup

A java applet was developed to present the secondary acoustic task in a standardized way. This applet was controlled on a further laptop with two sound speakers attached that were positioned behind the monitor in order to guarantee best sound. The demographic questionnaire, written instructions, and the SEA-scale were given to the subjects in paper form.

Procedure

Subjects were welcomed and seated at a desk to fill out a demographic questionnaire. Afterwards, they were informed that the aim of the study is to investigate different kinds of interaction. Instructions for the simple search task were given to the subject in a written form. They were instructed to perform the task as fast and as accurate as possible. After calibration of the eye-tracker, the subject had the possibility to train the task using the mouse. Therefore, 20 trials were presented to the subject. After they finished the training, subjects were given the instructions for the second task (HW) and the third task

(HTW). Subjects then had the possibility to train these tasks too with 20 trials each. After the training, subjects were calibrated again. In the main part of the study each subject fulfilled each task (40 trials) with each kind of interaction, i.e. 12 conditions. After each task subjects were asked to assess their mental stress using the SEA-scale. While filling out the SEA-scale, the experimenter configured the system for the next task and/or the next kind of interaction. After each kind of interaction, the calibration of the eye was reassessed and if necessary a new calibration was performed in order to ensure the most accurate gaze. Finally, participants were interviewed and asked what kind of interaction they preferred the most and which interaction they liked the least. Overall, the experiment lasted one hour. At the end of the experiment, subjects were compensated with 10 Euros and thanked for their participation.

Results

In the following section we present selected results regarding the most important aspects of our study. Data was preprocessed using a routine programmed in Java. Further statistical analysis was conducted with PASW Statistics 18.

Subjective experienced stress

In order to control whether workload was really induced to the subjects, we first had a look at the subjective experienced stress. A 3-factorial ANOVA with repeated measures on two factors was conducted, including the variables kind and level of workload and the kind of interaction. We found a significant interaction for the kind and level of induced workload ($p<.05$). As can be seen in Figure 3 there is no difference in the subjective experienced stress for the LW condition.

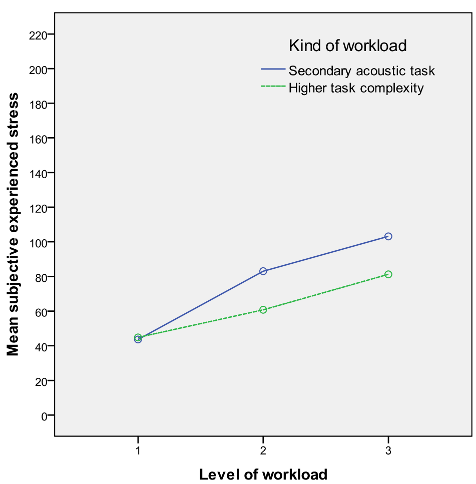

Figure 3. Mean subjective experienced stress for kind and level of workload (1=LW, 2=HW, 3= HTW)

This indicates that the groups had the same abilities to perform the task, as both groups had to perform the simple searching task on that level. For the HW and HTW level we find a significant difference between the two kinds of workload ($p<.05$). The task with the additional acoustic task is experienced more stressful (mean SEA-value 93) than the task with the higher complexity (71; $p<.05$). Generally we found a significant main effect for the level of workload ($p<.001$). In the LW condition the mean SEA-value is 44, in the HW condition it is 72, and 92 in the HTW condition (all $p<.001$). Therefore, we can conclude that workload was successfully induced. Furthermore we found a highly significant main effect ($p<.001$) for the kind of interaction. Mouse interaction is experienced significantly less stressful than all other kinds of interaction (all $p<.001$). The mean SEA-value was 55 for the mouse, 69 for the combined, 73 for the 500ms dwell and 80 for the 250ms dwell condition. We could not find a significant difference in experienced stress between the gaze-including kinds of interaction.

Reaction times for correct responses

Did the workload affect the performance of the subjects regarding their reaction times and what is the impact of the different kinds of interaction? To answer these questions we computed a 3-factorial ANOVA with repeated measures on two factors as well.

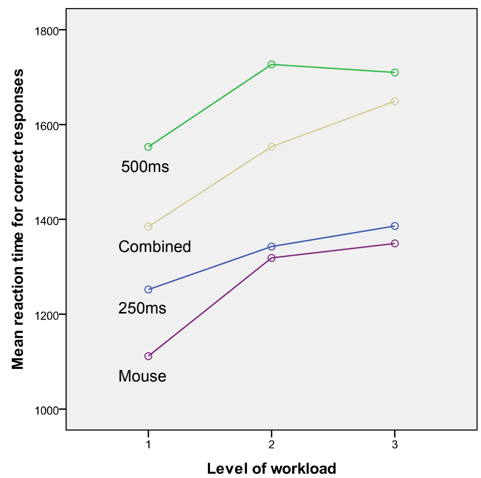

Figure 4. Mean reaction time for level of workload (1=LW, 2=HW, 3=HTW) and kind of interaction

We found a highly significant interaction for the level of workload and the kind of interaction ($p<.001$) and a highly significant main effect for the level of workload ($p<.001$). Generally, in the LW condition mean reaction times for correct responses are lower (1325ms) than in the HW condition (1485ms), but as can be seen in Figure 4 from the HW to HTW condition there is a further increase in reaction time just in the combined condition. No significant differences could be found when comparing the HW and HTW level for all other kinds of interaction. Additionally we found a highly significant main effect for the kind of interaction ($p<.001$). Mean reaction times for mouse and 250ms dwell gaze-based interaction do not significantly differ, whereas reaction times for the combined condition and the 500ms dwell gaze-based interaction are each significantly higher ($p<.001$).

False response errors

Regarding the total number of false response errors, we found, that the number is significantly lower in the mouse (67 errors) and combined condition (77) compared to the 250ms (637) and 500ms dwell time gaze-based condition (276). Chi-Tests revealed no significant difference between the mouse and combined condition but highly significant differences to and between all other conditions (all $p<.001$). Regarding the impact of the level of workload (see Figure 5), we found for the mouse and combined condition, that there is a significant increase in number of errors from the LW to the HW and HTW condition (all $p<.001$), but there is no difference between HW and HTW. In the gaze-based 500ms dwell condition we found a successive increase in number of errors for each level of workload (all $p<.001$). In the 250ms dwell condition we found no significant difference between the LW and HW condition, but a significant difference to the HTW condition ($p<.05$).

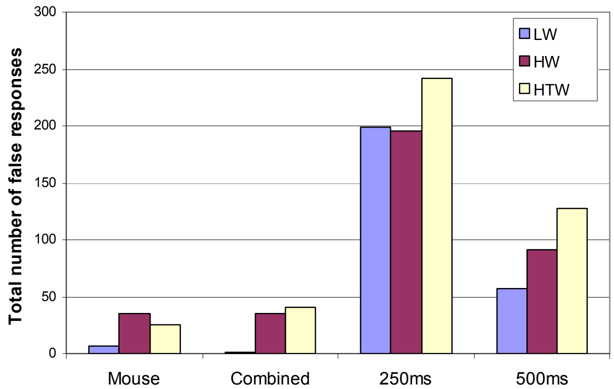

Figure 5. Total number of false response errors for kind of interaction and level of workload

CONCLUSION

In this experimental study we investigated the influence of imposed mental workload and visual distraction on gaze-including and mouse interaction. We found that mouse interaction was experienced as least stressful overall. This might be due to the "mouse-friendly" interface design we used in this investigation, but also to a lack of long-term experience with gaze-including kinds of interaction. It has to be pointed out, that even the combined condition was experienced as stressful as dwell time based gaze-interaction, although subjects had more control in this condition.

We found that the combined interaction led to comparably high reaction times for correct responses and was surprisingly sensitive to workload. Although error rates were comparable to mouse interaction, the necessity of precise eye-hand coordination in the combined condition led to a sort of reassurement behavior. Although subjects already looked at the correct button very quickly, they did not press immediately the key, but reassured themselves that it is really the correct button before pressing. In the HTW condition this reassurement behavior was obviously disturbed by the added visual distraction. From that point of view, it has to be taken into consideration that the coordination of different modalities leads to strong requirements for feedback design. In our study the borders of the button changed to indicate that it is chosen. More distinct feedback might have supported the combined interaction.

Finally we could show that visual distraction in combination with increased workload can lead to severe problems in gaze-based interaction. In both dwell time conditions error rates were highest in the condition where visual distraction was added. Although, as opposed to the combined condition, reaction times for correct responses were not affected in the same way.

From our results we want to develop design guidelines that are adjusted to gaze-including interaction and take into account future demands of HCI (e.g. complex information presentation, aggregated displays). We believe that this study is a base for further research and ongoing studies.

ACKNOWLEDGMENTS

We thank Mandy Dotzauer and Jessika Reissland for preparing and conducting the experiment, Mario Lasch for his technical support, and Marius Schwalbe for his programming work.

REFERENCES

1. Cooper, G. Research into Cognitive Load Theory and Instructional Design at UNSW. http://dwb4.unl.edu/Diss/Cooper/UNSW.htm.
2. Dzaack, J., Trösterer, S, Nicolai, T. and Rötting, M. iCOMMIC: Multimodal Interaction in Computing Systems. In *Proceedings of the 17th World Congress of the International Ergonomics Association* [CD], 2009, ID 3EP0100.
3. Eilers, K., Nachreiner, F. & Hänecke, K. Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft*, 40 (1986), 215- 224.
4. Jacob, R. J. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Human Factors in Computing Systems: CHI '90 Conference Proceedings*. ACM Press (1990), 11-18.
5. Myers, C.W., Gray, W.D. and Schoelles, M. The effects of stimulus configuration and cognitive workload on saccadic selectivity. *Journal of Vision 4,* 8 (2004), article 740.
6. Rötting, M. *Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen.* Shaker, Aachen, 2001.
7. Wolfe, J.M. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1 (1994), 202-238.