Gaze interaction enhances problem solving: Effects of dwell-time based, gaze-augmented, and mouse interaction on problem-solving strategies and user experience

Roman Bednarik, Tersia Gowases and Markku Tukiainen

Department of Computer Science and Statistics
University of Joensuu
Finland

It is still unknown whether the very application of gaze for interaction has effects on cognitive strategies users employ and how these effects materialize. We conducted a between-subject experiment in which thirty-six participants interacted with a computerized problem-solving game using one of three interaction modalities: dwell-time, gaze-augmented interaction, and the conventional mouse. We observed how using each of the modalities affected performance, problem solving strategies, and user experience. Users with gaze-augmented interaction outperformed the other groups on several problem-solving measures, committed fewer errors, were more immersed, and had a better user experience. The results give insights to the cognitive processes during interaction using gaze and have implications on the design of eye-tracking interfaces.

Keywords: gaze interaction, dwell-time, gaze-augmented, problem-solving, evaluation.

Introduction

Eye-tracking technology is advancing rapidly, it is becoming cheaper, more robust and easier to apply (Duchowski, 2007). The increase in availability has fueled an increase in applications of interactive usage of eye-tracking and many nowadays believe that, in some way, gaze-based interaction may eventually become a standard human computer interface (Jacob et al., 2007).

Gaze input, in comparison with other interaction devices, provides a more immediate way of communicating user intentions, thoughts and behavior. Eye-movements require little conscious effort and are therefore a good candidate for being a convenient computer interaction medium. As a natural means of selection, gaze input is easy to learn, interaction using the eyes is not only faster (Sibert & Jacob, 2000) but also less fatiguing, as the user does not have to engage in physical movement to select targets.

There are, however, several problems associated with eye gaze input. In addition to somewhat lower accuracy (Majaranta & Räihä, 2002), the eyes are primarily used to gather information about our immediate environment. Unlike a manual pointing device, the eyes cannot be turned off neither can they be placed in a state of idleness - where no selection takes place. Hence, a "Midas Touch" problem arises (Jacob, 1991); every time a user dwells on a screen target, whether it is intentional or accidental, an item on the screen is activated.[1]

What is not well understood at the moment is how the very application of interactive eye-tracking affects its users, where one has to use the eyes not only to visually perceive the problem domain, but also to interact with the problem. In particular it is still largely unknown, what are the effects of gaze-based interfaces on a user's search or problem-solving strategies. This is an important question which needs to be answered before gaze-based interfaces are implemented on a larger scale.

This issue is important both from the perspective of cognitive psychology and user-interface design. Proper interaction methods can support users' skill acquisition and learning and thus user interfaces generally aim at minimizing user's efforts by supporting task planning, completion, stress management, or even visual attention skills (Cockburn, Kristensson, Alexander, & Zhai, 2007). However, when people have to use the same –primarily perceptual– interactive modality bidirectionally, the increased efforts in coordination of the limited resources are likely to interfere with problem-solving activities. For example, if the cost of issuing a command via gaze-based interaction is higher than when using a mouse or keyboard, traversing the problem space becomes more difficult, and presumably more internal. Thus, a critical question arises about how to design gaze-based interfaces that scaffold the user strategies.

In this paper, we approach the problem through problem-solving interface design, two gaze-based interaction techniques, and an empirical evaluation. We report on an empirical experiment in which participants were randomly assigned to use the interface either with a dwell-time based input, gaze-augmented input, or a traditional computer mouse. We hypothesized that the varying efforts related to using an

---

[1] Istance, Bates, Hyrskykari, and Vickers (2008) single out the Midas Touch problem as one of three main problems associated with using gaze as an input medium.

interaction method influence problem-solving strategies and performance. In particular, we predicted that solving the problem using gaze-based interaction would increase planning activities and decrease action and manipulation activities. A post-experimental questionnaire was administered to juxtapose user experience data and behavioral data.

Related Work

Over the past couple of decades, several studies have investigated the use of gaze input for direct interaction and manipulation tasks such as selecting items on the screen using alternative gaze-interaction techniques. These techniques can be summarized under two main categories: those that solely use eye gaze interaction (e.g. blinking, winking, gaze directional gestures, and dwell time) (Špakov, 2005; Ohno, Mukawa, & Kawato, 2003; Hansen, Johansen, & Hansen, 2003; Majaranta & Räihä, 2002) and those that use gaze together with another input device such as a mouse or keyboard (Zhai, 2003; Kumar, Paepcke, & Winograd, 2007).

Surakka, Illi, and Isokoski (2003) argue that to evaluate a gaze-based interaction technique means to "empirically test different features and attributes of the technique used." To show the possibilities and advantages of using eye-movements a large number of studies compared gaze-based interaction to using interfaces with the traditional computer mouse.

One of the oldest comparison between gaze and mouse is that conducted by Ware and Mikaelian (1987). They first investigated three types of selection methods, then the effects of target size. The results suggested that eye selection can be fast if the size of the target is large enough and that dwell-time selection can be as fast as selection using a button press.

Sibert and Jacob (Sibert & Jacob, 2000) conducted two experiments where they compared the selection speeds of dwell-time interaction to the mouse. The first experiment involved selecting a highlighted circle from a series of blank circles. While in the second experiment, each of the circles in the series contained a letter and users had to select the circle whose letter was called out over an audio speaker. Dwell-time interaction outperformed the mouse in both experiments in terms of speed.

Zhai et al. (Zhai, 2003) introduced a Manual and Gaze Input Cascaded (MAGIC) pointing technique which automatically positions the cursor to the point where the user is looking on the screen, while final target selection was achieved using a mouse. Their reported benefits for MAGIC pointing included greater accuracy, elimination of Midas Touch, reduced physical effort and fatigue, and selection speeds faster than with manual pointing. However, some participants felt that they had to use more effort to coordinate their eyes and hands.

Other studies have used gaze input in a non-command manner, where the interfaces monitor the user's eyes and responds appropriately "without the user explicitly giving a command" (Nielsen, 1993). For example, Ashmore, Duchowski, and Shoemaker (2005) used a fish-eye lens to magnify the region of the screen being looked at. This mag-nified region could then be used for a number of activities such as dwell-time based object selection. *Little Prince Storyteller* is a gaze-based storytelling game, where a narration is given in synthesized speech; the narration's generality and specificity depend on the user's attention (Starker & Bolt, 1990).

Any interactive behavior requires perceptual, motor, and cognitive operations (Gray, Sims, Fu, & Schoelles, 2006). Previous eye-tracking studies investigated rather simplistic situations and tasks, and approached gaze-based interaction mainly from a motor-perceptual perspective. Interaction with present-day computer interfaces, however, cannot be considered simply as just pointing and selecting targets. Users interaction is composed by of a range of complex strategies, including goal searching, planning, handling interruptions, and internal and external information coordination, in which selection and pointing have an important, but not sole, role.

To get a better understanding of the mechanisms involved in problem-solving with interactive displays, it is not sufficient to evaluate only low-level performance measures such as pointing time. In addition, we should: 1) study the interaction between an internal representation of the problem and the external environment (O'Hara & Payne, 1998); and 2) investigate the effects of the interaction with the display on the strategies. Both parts of the challenge have been previously addressed by numerous experiments, particularly in cognitive science and partly in HCI research (e.g. (O'Hara & Payne, 1999; Van Nimwegen, Van Oostendorp, & Tabachneck-Schijf, 2004, 2005; Gray et al., 2006; Cockburn et al., 2007)). The interaction between the user's physical actions and strategies, the task, and the design of the task environment for the problem, in the presence of gaze-controlled environments, has not been studied.

In this research we initially considered the claim of Van Nimwegen et al. (2005) that inducing planning, by not externalizing information, facilitates more internal processing and planning and thus provides advantages to learning and problem-solving. In a preliminary work we informally observed (Gowases, Bednarik, & Tukiainen, 2008) that mouse-based interaction with a problem supported information externalization. Users often perform a tinkering behavior, in which they use the external display as a work space – instead of working on an internal representation – trying out different configurations, paths, and approaches to solve a problem. We further formed a hypothesis that gaze-based interaction would cause users to internalize the problem-solving activities more, because the efforts related to the interaction, such as the cost of a move, would be perceived higher than when using a computer mouse. This would materialize in better problem-solving performance.

In this paper we aim to fill the lack of knowledge about three questions: 1) what are the strategical differences in problem-solving between gaze and mouse interaction, 2) whether and how interaction patterns change when gaze-interaction occurs, and 3) what are the outcomes of such a change. We describe problem-solving and user interaction using the traditional performance measures such as moves per minute, using fixation duration, and indirect user-

experience measures.

Method

To investigate the effects of gaze-based interaction methods on user strategies, we designed a study where users solved an 8-tile puzzle using either a dwell-time selection method, gaze-augmented selection, or conventional computer mouse.

User Interface and Interaction Techniques

We selected the 8-tile puzzle because it presented a generally familiar and well defined task, and thus the range of user strategies and problem solving analysis are reasonably constrained. The puzzle has been studied not only in the domain of psychology, but has important implications in the domain of HCI (O'Hara & Payne, 1998, 1999). Figure 1 shows a screen shot of the user interface of the puzzle.
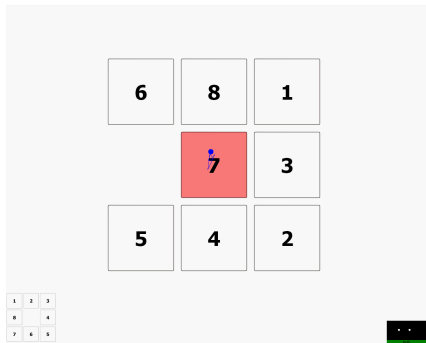

Figure 1.   8-tile slide puzzle, after the tiles have been shuffled. User is gazing to the center and could see the target solution (bottom left) and the status of the eye-tracking (bottom right).

The target application uses either a computer mouse as a manual input device, or *dwell time* activation, or *gaze-augmented input* activation for gaze input. Dwell time activation (Hansen et al., 2003) is a selection method that allows a command to be executed once a user's gaze is on an interaction element for a predefined amount of time. The executed command is equivalent to a single (left button) mouse click. Color animation using hue volume (see Figure 2) indicates the amount of time left before the tile is selected – the longer the user looks at the button the more red it will turn[2]. If the user's gaze leaves before the tile has been selected, the tile's color gradually animates back to white. If the users gaze returns to the tile before the tile has turned white, the tile's color animation – and the respective selection timer – continues from the current state.

From an implementation perspective, the application and eye-tracker interaction in the two gaze conditions is achieved using a call-back function via an application programming interface. The function is automatically invoked every time new gaze-data is available. The function simultaneously checks whether the current gaze point (the point the user is looking at) lies on one of the tiles, and identifies the tile.

Before the study, we derived a series of settings and implemented a range of design alternatives for the two gaze-based

techniques. Using information available from related work (Jacob, 1991; Stampe & Reingold, 1995; Sibert & Jacob, 2000; Jacob & Karn, 2003) we pilot-tested the settings for button activation timers, gaze-position filters, feedback, and target size. A button size of 200 x 200 pixel (approximately $5.1°$ of visual angle at viewing distance 60cm and screen resolution 1280 x 1024) was selected in order to reduce the Midas touch effect under the gaze conditions.

Figure 2 shows the final version of the dwell-time selection mechanism employed in this study. A dwell-time duration of one second was selected. Although the dwell-time for selection can be as short as 150ms, Stampe and Reingold (1995) recommend that for more difficult tasks the the dwell time should be longer. This allows users to overcome gaze disturbance problems such as blinking, re-fixation, or other type of disturbance. During the pilot test for dwell-time interaction, participants reported that a synthesized click sound gave them an additional confirmation that the command was issued and thus clink sounds were included as auditory feedback.

In the dwell-time input a continuous timer that measures the duration of user's gaze is attached to each interaction element. If the gaze remains on the element, the timer starts and the button's fill gradually turns red. If user's gaze leaves before the button is activated, the animation is paused and, after a short while, timer is reversed and returns the button back to white. This solution allows gaze, that is inherently noisy, to leave the button, for example during blinks.

Figure 2.   Dwell-time button color animation, using hue volume to indicate the remaining time before selection.

During gaze-augmented input an on-line algorithm constantly monitors user's gaze point with respect to the interaction elements. Once a tile has been identified as being gazed upon, it will immediately be highlighted green. The user can then issue a command on the selected the tile, for example by clicking the left mouse button. In both gaze-based methods, the mouse cursor is not displayed on the screen, therefore moving the mouse has no impact on tile selection.

Finally, in the mouse-controlled condition, users could point at a tile and click the left mouse button to initiate the move. A green color was selected to highlight buttons under the gaze-augmented and mouse conditions to indicate that the selection was ready to 'Go' at any time.

Experiment

The experimental settings, design, procedure, and analysis of the data were largely inspired by the studies of O'Hara

---

[2] An online animated figure showing the three interaction modes in action is available at: http://www.youtube.com/watch?v=oXBKXRqxVnU

and Payne (O'Hara & Payne, 1998, 1999). While their experiments focused mainly on the manipulation of high- and low-cost versions of the interface, we manipulated the interaction method.

Apparatus.

A Tobii ET 1750 eye tracker (http://www.tobii.se) was used for tracking the eye movements and gaze interaction. The Tobii 1750 (sampling rate of 50Hz) eye tracker is a non-intrusive dark-pupil binocular eye movement tracker that is integrated within a 17 inch monitor with a maximum resolution of 1280 x 1024. The eye tracker is able to accommodate the normal head motions a person makes when sitting at a distance of 60cm from the screen. The voice protocol was recorded using the control program of the eye-tracker.

Participants.

A sample of forty-one participants ranging in age from 21 to 53 (mean age 28.9 (7.22)) took part in the study. Of these, five participants' data were excluded from further analysis, due to technical problems leading to low-quality eye-tracking data. The remaining group of thirty-six participants was composed of 23 male and 13 female. The typical educational background of a participant was a university-level computer science undergraduate student. All participants had normal or corrected-to-normal vision. All participants were experienced computer users and understood the logic behind solving the puzzles.

Procedure and Design.

We made use of a mixed subject design similar to O'Hara and Payne (1999). The between-subject factor in our study was the interaction device with three levels (dwell-time, gaze-augmented, and mouse conditions). The within-subject factor was the session number with three levels. Participants were randomly divided into three groups, corresponding to the three input modalities. Each group consisted of 12 participants.

The experiment was conducted in a quiet laboratory. Participants were asked to complete a consent form and a background information questionnaire. Participants were then briefed on the rules of the 8-tile puzzle and interaction method to be used. Participants were then seated in front of the eye tracker and had to pass a calibration test.

Participant first had to complete a short think-aloud practice task followed by a warm-up puzzle. The starting configuration of the warm-up puzzle was aimed at getting participants familiar with both the interface and the assigned interaction method. Once participants were comfortable with the warm-up puzzle they started solving the target puzzles.

Participants completed three trials. Each of the three puzzles had a unique starting configuration (Figure 3); however the complexity of each puzzle was similar. The order of the three configurations was randomized and counterbalanced for a participant, so that each configuration appeared the same number of times at each of the tree possible trials.

Once a participant had completed all the puzzles under the specific condition, we allowed the participants from the gaze-augmented and dwell-time conditions to experience shortly

the other mode of interaction with the application. This aimed at providing participants the referential experience for answering the comparative questions after the study. Finally, the participants were given the post-test questionnaire and interviewed.

Figure 3. Three starting configurations.

Data Analysis.

The dependent variables in this study included: completion time, number of moves to solution, time per move (moves per minute), time to first move, inter-move latency, occurrence of backtracking behavior (due to interaction errors and problem-solving errors, see section Data Analysis), and mean fixation duration. We also collected post-experimental data to evaluate user experience, including subjective satisfaction, immersion, easiness, and naturalness of the interaction.

We included data from the three target sessions into the analysis. For each participant and trial, an interaction protocol was recorded that included timestamped logs containing the tile movement information. Completion time, number of moves, time to first move, and backtracking were computed or identified automatically from the interaction using custom tools. Results were analyzed using analysis of variance (ANOVA) and the post-hoc analyses of significant main effects employed comparisons using Bonferroni correction.

Errors in the strategies were post-experimentally evaluated from the interaction protocols, verbal protocols, and video-recordings. To evaluate what type of errors occurred, we looked at situations where participants exhibited a backtracking behavior. According to O'Hara and Payne (1999), *backtracking* occurs when users make a move or sequence of moves and then realize that they have made a poor choice. Trying to reverse the sequence of moves in order to return to the starting point before the move sequence is called backtracking. To distinguish problem-solving as a source of backtracking from interaction problems, we further defined two sources of backtracking behavior. An *interaction error* results in a backtracking when user in-adverdly selects a wrong target. For example, in the dwell-time condition, a participant would want to select a tile number 1, but due to the Midas touch selected tile number 2. A *problem solving error*, on the other hand, would result from a poor problem-solving strategy. In these cases, the participants would select the targets according to their intentions, but then would realize the strategy is not working. To further help the distinction between an interaction error and a problem solving error, we used the verbal protocols.

All results were statistically analyzed using SPSS v14 for Windows. For the analysis of eye-tracking data and computation of the eye-tracking measures, we used the built-in fix-

ation detection filter with a threshold of 100ms for minimal fixation duration. Fixational data were computed and visualized using available software packages (e.g. Clearview, see http://tobii.se).

Results

All results were analyzed considering the potential effect of the starting configuration on the resulting data. No effect of initial configuration was found, indicating that all three configurations were equal.

Before we report on the detailed quantitative results, we characterize the general interaction patterns and user strategies based on observations made during the study. The dwell-time gaze interaction and gaze-augmented interaction were new to all participants. Participants in the dwell-time condition quickly developed strategies aiming to avoid Midas touch, for example by gazing at tiles that could not be moved even after prolonged time. However, this type of compensation lead to distractions. Also, they often reported a feeling of being under pressure not permitting them to "think freely". At later trials they reported increased fatigue. Even though the mouse-cursor was disabled and not visible in the gaze-augmented condition, several participants attempted to move the computer mouse as it would actuate the position of the selection cursor. Few participants reported on difficulties caused by having to use and coordinate two modalities at the same time.

Performance and Problem-solving

We first consider the following behavioral measures as indicators of user performance and problem solving: completion time, number of moves, and moves per minute.

Overall, *completion times* for all three trials were 233.68, 217.96, and 425.03 seconds for mouse, gaze-augmented, and dwell-time conditions respectively. Figure 4(a) shows the break-down of the completion times for the three trials and three interaction modalities separately.

The average *number of moves to solution* for the three interaction modalities were 105.00, 74.83, and 127.61 for mouse, gaze-augmented, and dwell-time conditions respectively. *Number of moves per minute* is a composed normalized measure obtained by dividing the number of moves to solution by the completion time in minutes. The average number of moves were 35.52, 22.18, and 19.10 for mouse, gaze-augmented, and dwell-time conditions respectively. Table 1 and Table 2 summarize the number of moves to solution and the number of moves per minute, respectively, for the three trials in detail.

We performed a series of 3 (trial) x 3 (interaction modality) repeated-measures ANOVA on the data presented above, and all data were log-transformed to stabilize the variance. Results related to main effects are summarized in Table 3. Interaction effects were generally non-significant and weak. Post-hoc comparisons discovered significant differences between mouse and dwell-time (p = .006), gaze-augmented and dwell-time (p = .022) conditions in completion time. Further, gaze-augmented and dwell-time conditions also significantly
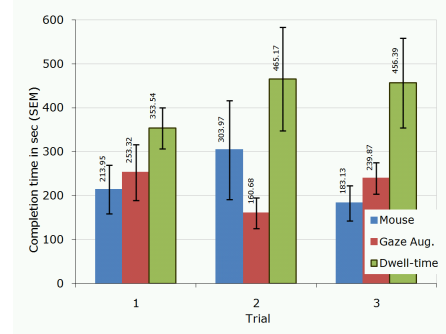
(a) Completion time for each of the three trials and interaction modality.
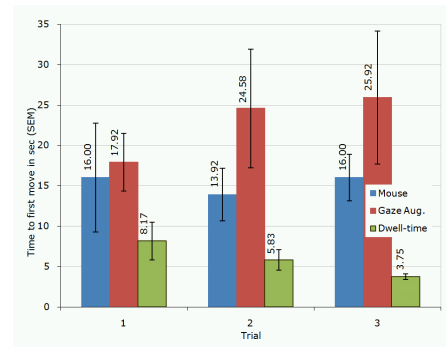
(b) Time to first move for each trial and interaction modality.

Figure 4. Performance measures.

	Trial 1	Trial 2	Trial 3
Mouse	104.5 (118.6)	116.8 (107.4)	93.7 (76.4)
Gaze-augm.	78.3 (80.7)	54.3 (46.4)	74.8 (33.0)
Dwell-time	98.1 (42.8)	151.0 (118.0)	127.6 (98.3)

Table 1
The mean number of moves to solution (with standard deviation) during the three trials.

differed in number of moves (p = .039). Finally, the number of moves per minute significantly differed in the mouse and dwell-time conditions. Other comparisons were not significant.

To deeper evaluate the user-strategies, we computed the time to first move, inter-move latencies and counted the number of backtracks in terms of interaction and problem-solving errors.

The overall mean *times to first move* were 15.31, 22.81,

	Trial 1	Trial 2	Trial 3
Mouse	33.1 (17.0)	35.9 (20.9)	37.5 (19.8)
Gaze-augm.	21.6 (10.2)	21.6 (7.4)	23.3 (9.3)
Dwell-time	17.0 (3.8)	21.4 (10.5)	18.9 (4.8)

Table 2
The mean number of moves per minute (with standard deviation) during the three trials.

Measure	Trial		IM	
	F(2,66)	p	F(2,33)	p
Completion time	0.206	.81	6.56	.004
Number of moves	0.838	.44	3.46	.043
Moves per minute	1.311	.28	4.98	.013

Table 3
The summary of the main effects of trial and interaction modality (IM).

and 5.92 seconds for mouse, gaze-augmented, and dwell-time conditions respectively. Figure 4(b) shows a detailed view on the measure for each of the trials and interaction modalities.

Higher *inter-move latency*, according to O'Hara and Payne (1998), indicates more plan-based strategies and increased cognitive efforts. The overall mean inter-move latencies were 2019.5, 3221.9, and 3544.6 ms for mouse, gaze-augmented, and dwell-time conditions, respectively. Table 4 summarizes the inter-move latencies for all three interaction modalities and trials.

Backtracking behavior is characteristic of search strategies with external displays rather than for strategies carried out on internal representations (O'Hara & Payne, 1998). Overall, the instances of backtracking behavior were 6.8, 6.5, and 27.1 for mouse, gaze-augmented, and dwell-time conditions, respectively. Applying the dichotomy of errors to the overall number of backtracking, there were 0.1, 2.7, and 20.6 interaction errors in mouse, gaze-augmented, and dwell-time conditions, respectively, and 6.6, 3.8, and 6.5 backtracks due to problem-solving errors in mouse, gaze-augmented, and dwell-time conditions, respectively. Table 5 summarizes the backtracking behavior in terms of number of such occurrences, together with a breakdown to the backtracking caused by a problem-solving error (PSE) and the backtracking caused by an interaction error (IE).

The main findings of the analysis of the problem-solving measures are summarized in Table 6. There were no interaction effects of trial and interaction modality on the time to first move and on inter-move latency. However, there was an ordinal interaction effect on the interaction errors F (4, 66) = 7.44, p < .001). A closer investigation of this effect (Table 5) reveals that it was mostly produced by the steep increase in the number of interaction errors in the dwell-time condition, starting during the second trial (30.17 on average) and slightly decreasing in the third trial (26.50 on average).[3]

A post-hoc analysis of the significant main effects discovered significant differences between gaze-augmented and dwell-time (p = .006) conditions in time to first move. The other two comparisons showed no significant differences. To delineate the causes of the main effect of trial on inter-move latency, we performed three additional pair-wise comparisons. There were significant two-tailed differences between the first and second (p = .003) and fist and third trial (p = .017), and no significant difference between the second and third trial. Comparisons of inter-move latencies with regard to interaction modality showed significant differences

between mouse and gaze-augmented conditions (p = .016) and mouse and dwell-time conditions (p = .002); the difference between gaze-augmented and dwell-time conditions was not significant.

Mean fixation duration is a measure often linked with the depth of required processing (Just & Carpenter, 1976; Goldberg & Kotval, 1999). Figure 5 summarizes the results related to that measure. An ANOVA discovered a significant effect of trial F (2,66) = 3.59, p = .033, η^2 = .098, and a non significant effect of interaction modality F (2,33) = 2.48, ns, η^2 = .129. There was no interaction effect between trial and interaction modality. The main effect of trial was analyzed by a series of pair-wise t-tests. The only statistically significant difference was between the first and third trial, t (35) = 2.73, p =.010), where the absolute difference was 10.8 ms between first and third trial.

Although not showing a significant effect of interaction modality, the effect size indicated that there may be differences in the mean fixation duration between interaction modalities in general. An independent sample t-test of all mean fixation durations showed that there indeed were significant differences between mouse and dwell-time condition of 70 ms (p = .0004) and between mouse and gaze-augmented condition of 39.4 ms (p = .015).

User Experience

Participants rated their experiences with the interaction method that they just used. Ratings were made on a five-point Likert scale. A rating of 1 indicated a strong agreement (e.g., I like the interaction very much/very natural means of interaction), a rating of 3 indicated that the user had a neutral opinion of the interaction, whereas a rating of 5 indicated that the user opposed the interaction (e.g. the game was very difficult to control using this interaction). The results were analyzed using χ^2. Table 7 summarizes the results related to user experience and immersion.

The results obtained from the last user experience question, that is, "How immersive did you find the game using this interaction?" clearly show that although the gaze-based interaction methods were not found the easiest and most natural, they were recognized as immersing the users into the

---

[3] The number of interaction errors committed in these two trails and dwell-time condition accounted for 81% of all interaction errors, therefore both main effects and interaction effects need to be interpreted with care.

Measure	Trial		IM	
	F(2,66)	p	F(2,33)	p
Time to 1st move	.73	.93	5.56	.008
Inter-move latency	6.92	.002	7.98	.001
PSE	.05	.95	0.90	.42
IE	5.98	.004	21.49	< .001

Table 6
The summary of the main effects of trial and interaction modality (IM) on problem-solving strategies. Note, PSE denotes Problem-solving Errors, IE denotes Interaction Errors.

	Trial 1	Trial 2	Trial 3
Mouse	2176.2 (1434.2)	1929.7 (986.3)	1952.5 (1206.4)
Gaze-augm.	3567.9 (1042.2)	2908.0 (1144.5)	3189.7 (1910.6)
Dwell-time	4248.0 (1133.8)	3250.2 (733.5)	3135.6 (450.9)

Table 4

Inter-move latency in milliseconds (with standard deviation) as a function of trial and interaction modality.

		Trial 1	Trial 2	Trial 3
Mouse	Backtracking	6.33 (9.21)	7.33 (9.96)	6.67 (9.31)
	PSE	6.08 (8.83)	7.17 (9.70)	6.67 (9.31)
	IE	0.25 (0.45)	0.17 (0.39)	0.00 (0.00)
Gaze-augm.	Backtracking	9.67 (8.80)	4.75 (4.67)	5.00 (4.67)
	PSE	6.00 (6.32)	2.50 (2.97)	2.83 (3.46)
	IE	3.67 (4.46)	2.25 (2.30)	2.17 (2.48)
Dwell-time	Backtracking	9.17 (9.59)	37.67 (36.18)	34.42 (22.65)
	PSE	4.17 (4.80)	7.50 (8.46)	7.92 (6.47)
	IE	5.00 (6.16)	30.17 (30.09)	26.50 (17.02)

Table 5

Mean number of backtracking (with standard deviations) for all three trials and interaction modalities with a breakdown to backtracks due to Problem-solving Errors (PSE), and Interaction Errors (IE).
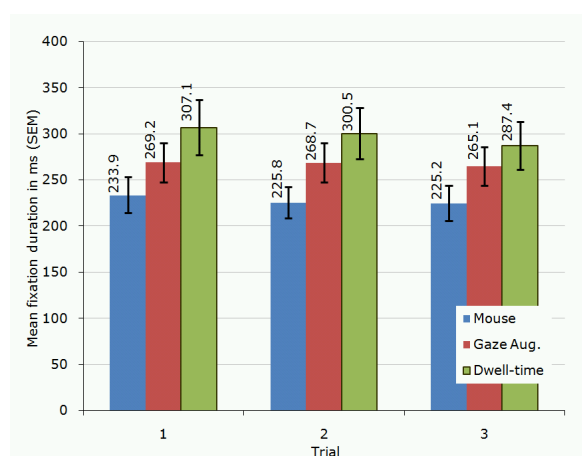

Figure 5. Mean fixation duration for each trial and interaction modality.

game experience. The most immersive mode of interaction was the gaze-augmented input, followed by the dwell-time interaction. One participant mentioned that she enjoyed the fact that she was using more of her senses during interaction, *"I wasn't even aware that I was thinking"*, while another stated *"You are able to see all your eye movements and in a way it highlights the numbers more and makes you remember what you have already looked at. In this way you are more aware of the numbers, their position and in time can manage to solve the problem faster"*. On the other hand, participants felt that there was *"nothing special"* about interaction with the mouse. In summary, dwell-time based interaction received generally the worse subjective feedback from users. On the dimension of immersion, gaze-augmented problem-solving was found to be subjectively perceived as the best.

Discussion

The results presented above only partially confirm the hypotheses we formed about the effects of interaction modality on problem solving. Our results show that the way people interact with a problem-solving interface indeed depends on the interaction device they use, however, interaction modalities have different effects on user strategies, performance, and user experience.

When gaze-input interferes with user strategies, a performance decrease can be observed in terms of completion time, number of moves, and moves per minute. It is not surprising that completion time was higher in the dwell-time condition, given the fact that each move required at least 1 second to initiate a selection. The number of moves were also higher in the dwell-time condition. The measure of moves per minute is an equivalent to the words-per-minute measure used in eye-typing studies. In our experiment, the participants were however not instructed to perform as fast as possible and thus the completion time and moves per minute have to be considered with care.

In the often cited pioneering study of Ware and Mikaelian Ware and Mikaelian (1987), hardware button selection, a technique similar to the gaze-augmented condition in the present study, was faster than the dwell-time selection. While our data cannot be directly compared to the selection times of Ware and Mikaelian, a possible comparison arises with those of eye-typing research. Stampe and Reingold (1995) reported 7 wpm, Ward and MacKay (2002) reported 25 wpm or even 34 wpm for an expert eye-typist. In our study, the fastest participants performed as many as 45 moves per minute in the gaze-augmented condition, and 53 moves per minute in the dwell-time condition.

The number of moves to solution presents a more fair comparison in the domain of problem-solving. The participants in the gaze-augmented interaction completed the task

Question	Dwell Time	Gaze Augm.	Mouse	$\chi^2(2)$	p
1 Did you like this interaction?	1.83	1.67	1.83	0.33	0.846
2 How easy was it to control the game?	3.17	1.92	1.42	13.5	.001*
3 How natural was it to use this interaction?	2.58	2.17	1.67	4.81	0.09
4 How immersive did you find the game?	1.83	1.75	2.50	7.44	.024*
Average rating	2.35	1.88	1.86	8.11	.017*

Table 7

The summary of the user experience questionnaire. Lower number means better result.

using only half the number of moves than participants in the mouse condition, and only one-third of the number of moves than the participants in the dwell-time condition. This would indicate that most of the problem-solving activities occurred internally and participants had to use more planning activities on the internal representation of the problem before invoking a change in the external display.

More immediate quantitative indicators of the nature of problem-solving strategies, such as the inter-move latency, backtracking due to problem-solving error, or time to first move inform about the supremacy of gaze-augmented inter-action in problem-solving. Time to first move was longest in the gaze-augmented condition, indicating that participants investigated the problem and spent longer time on planning activities. Compared to the users with mouse, gaze-augmented participants also exhibited longer inter-move latencies, further indicating that gaze-augmented interaction supports plan-based strategies. Although the number of problem-solving errors was generally lowest in the gaze-augmented condition, statistically the backtracking behavior, in terms of problem-solving errors, did not differ between the conditions.

The indicators of plan-based problem-solving raise considerations about the interplay of dwell-time based interaction and problem-solving. For example, the inter-move latency was longest for dwell-time interaction, and, even when the one-second threshold would be subtracted, the inter-move latency in that condition was longer than in the mouse condition. Users committed approximately same number of problem-solving errors as in the mouse condition. Finally, the number of moves to solution in the dwell-time condition – were the interaction errors removed – was comparable to that in the mouse condition. These results indicate that even though the dwell-time based interaction partially restricted users to freely analyze the problem-space, this inhibition did not have serious effects on problem-solving strategies.

The increased mental efforts spent on planning activities were also reflected in the longer mean fixation durations. The longest mean fixation durations were found in the two gaze-based conditions, indicating that participants in these conditions had to invest more mental processing to solve the tasks than the participants using mouse. This measure triangulates with other indicators of increased internal efforts in the gaze-augmented condition, but in the dwell-time, interaction can be affected by the efforts to initiate a command by spending one second within a certain region.

The decomposition of the overall backtracking behavior unearths the reasons for the dwell-time condition's poor per-

formance, compared to the other two conditions. In particular, dwell-time based interaction was clearly prone to interaction errors. This finding is similar with the higher error rate reported in studies comparing mouse and dwell-time selection, e.g. (Hansen et al., 2003). We observed that when users committed an interaction error, often they committed more errors on the way to correct the original error. The constant fixing of interaction errors triggered a cyclical cumulation of errors. When observing the video protocols with a superimposed gaze-path, it becomes obvious that the source of the interaction errors often resulted either from 1) users ending the current fixation just before the move-command was invoked (i.e., before the end of the dwell-time counter) or 2) fixating on a target other than the one intended. We hypothesize that the increase in the number of interaction errors in the dwell-time condition from the second trial could be caused by fatigue, loss of motivation, or a combination of both. This possibility will be studied in future.

There are several viewpoints from which to discuss the results presented above.

If the user interface *externalizes* data by displaying relevant information pertaining to a task, then the user need not recall certain knowledge. The user can perform actions directly on the external display, therefore relieving working memory and the related cognitive efforts. When faced with a problem, computer users do not generally plan the complete sequence of actions, which they will carry out, before attempting to solve their problem (O'Hara & Payne, 1998); instead they often execute an action as soon as their first actions are decided upon and plan further action as they go along the task. This type of problem solving strategy is a consequence of the externalization. We showed that also by manual input can cause more external problem-solving. It causes users to rely on information obtained from the user interface and interact directly with its externalized form.

If however the interfaces supports data *internalization* then "certain information is less directly available" (Van Nimwegen et al., 2005) causing the user to rely on and work with information stored internally while doing problem solving task, such as initiating some type of plan. Van Nimwegen et al. (Van Nimwegen et al., 2004, 2005) show that internalizing information makes users to plan more, get better knowledge and not reach implausible situations. In our study, the user interface remained the same and provided the same information to all participants. However, it can be suggested that gaze-based interaction affects users to carry out internalization of the information.

O'Hara and Payne (1998) argued that high-cost inter-

action operators improve performance and that training on an interface with these high-cost operators results in better subsequent performance in the same domain. O'Hara and Payne's implementation cost of an operator include factors such as the amount of time, and physical and mental efforts. They also claim that the characteristics of a certain task encourage more or less planning. Cockburn et al. (2007) showed that interfaces that induce extra costs in terms of mental efforts support spatial learning.

Contrary to our hypothesis, the dwell-time based gaze interaction, although requiring extended costs of operator, did not have a greater positive effect on the performance and problem-solving. On the other hand, we presented an evidence that gaze-augmented interaction is characterized by increased planning activities and internalization of a problem. Our explanation to this finding is that participants were able to effectively use the freed resources that would otherwise be allocated for physical movement of the cursor. Although dwell-time interaction can possibly free up the resources for strategical planning in the similar way, they were used for correcting the problems in the interaction, rather than for the process of problem-solving.

Interacting with a problem space using a computer mouse facilitated exploration and problem-solving on the external representation. With the computer mouse, users engaged in display manipulation instead of performing operations internally. The findings confirm that interacting with a problem using computer mouse does not trigger users into more deeper, plan-based reflective cognition, but instead it supports what Norman calls "experiential cognition" (Norman, 1993; Van Nimwegen et al., 2005).

The gaze-augmented interaction was generally preferred over mouse interaction. In some cases users could not discern between mouse and gaze-augmented interaction in terms of both preference and ease of use. Smith and Graham (2006) also found that gaze input could increase the level of immersion and enjoyability during play. The results of the present study are in line with their findings.

Conclusions and Future Work

Design and interaction mechanism of user interfaces can support beneficial activities during problem solving with instructional systems. We presented a study investigating the effects of interaction methods on user strategies, problem-solving, and user experience. In particular, we compared dwell-time based, gaze-augmented, and mouse-based interaction in problem-solving tasks and analyzed the interaction with the problem in terms of several measures.

Gaze-augmented interaction allowed users to spend more time on problem-analysis and on planning, resulting in improved problem-solving performance. The dwell-time method introduced a serious interaction obstacle that caused users to engage in constant correction of the problem-space, however, their problem-solving performance was not seriously affected and was comparable to the performance of participants using computer mouse.

We found that users felt more immersed in problem-solving when using either of the two gaze-based input methods than when using a conventional mouse. This result is a promising one, taking into consideration the role of engagement and immersion in human computer interaction in general, and in educational interactive systems in particular.

In summary, gaze-based interaction with a problem introduces an interesting dilemma to both the user and the designer of problem-solving interfaces. The efforts connected with initiating a command are lower than when using mouse and can be spent on strategic planning. However, although the selection of the next target is faster and provides a clear benefit, the large number of interaction errors in the dwell-time condition presents an obstacle to benefit from the increased easiness. The gaze-augmented condition seems to solve the problem of Midas Touch, and in addition it improves the problem-solving performance and strategies. When pointing with their eyes and issuing a command using a button, users plan more before conducting a move, an activity resulting in better problem-solving strategies.

This study raises important questions about the mechanisms by which the gaze-based methods support planning and goal oriented behavior in problem solving with external displays. In other words, to understand what the qualitative differences are in problem-solving strategies is a task for future research. To answer that question, we plan to conduct a protocol analysis on the content of the verbal statements, and replicate the method of O'Hara and Payne (1998).

Another way to extend our findings is to conduct a longitudal study with increased number of trials, to provide an account of extended effects of gaze-based interaction on learning. Extending the current experimental settings with more trials, however, has to be done carefully, since even the most immersed participants became tired when repetitively presented with a problem-solving task.

Long dwell times might hinder interaction performance, whereas short dwell times result in inaccuracy in selection (MacKenzie & Zhang, 2008). Therefore, another extension of the present results lies in implementing and testing online user-adjustable dwell-time, as previously applied in eye-typing systems (Majaranta, MacKenzie, Aula, & Räihä, 2006). Apart from dwell-time, other gaze-based interaction methods such as gaze gestures, can be evaluated in a similar manner.

References

Ashmore, M., Duchowski, A. T., & Shoemaker, G. (2005). Efficient eye pointing with a fisheye lens. In *GI '05: Proceedings of Graphics Interface 2005* (pp. 203–210). Canadian Human-Computer Communications Society.

Cockburn, A., Kristensson, P. O., Alexander, J., & Zhai, S. (2007). Hard lessons: effort-inducing interfaces benefit spatial learning. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1571–1580). New York, NY, USA: ACM.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Goldberg, J., & Kotval, X. P. (1999). Computer Interface Evaluation Using Eye Movements: Methods and Constructs. *International Journal of Industrial Ergonomics*, *24*, 631–645.

Gowases, T., Bednarik, R., & Tukiainen, M. (2008). Gaze vs. mouse in games: The effects on user experience. In *ICCE2008: Proceedings of The 16th International Conference on Computers in Education, ICCE 2008* (pp. 773–777).

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006, 7). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review,*, *113*(3), 461-482.

Hansen, J. P., Johansen, A. S., & Hansen, D. W. (2003). Command without a click: Dwell time typing by mouse and gaze selections. In *Human-Computer Interaction – INTERACT'03* (pp. 121–128). IOS Press.

Istance, H., Bates, R., Hyrskykari, A., & Vickers, S. (2008). Snap clutch, a moded approach to solving the midas touch problem. In *ETRA '08: Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 221–228). New York, NY, USA: ACM.

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 of Information Systems*, *9*(2), 152–169.

Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M., Shaer, O., Solovey, E. T., et al. (2007). CHI2006: what is the next generation of human-computer interaction? *interactions*, *14*(3), 53–58.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 573–603.

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

Kumar, M., Paepcke, A., & Winograd, T. (2007). Eyepoint: practical pointing and selection using gaze and keyboard. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 421–430). New York, NY, USA: ACM.

MacKenzie, I. S., & Zhang, X. (2008). Eye typing using word and letter prediction and a fixation algorithm. In *ETRA '08: Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 55–58). ACM.

Majaranta, P., MacKenzie, I. S., Aula, A., & Räihä, K.-J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Univers. Access Inf. Soc.*, *5*(2), 199–208.

Majaranta, P., & Räihä, K.-J. (2002). Twenty years of eye typing: systems and design issues. In *ETRA '02: Proceedings of the 2002 symposium on Eye tracking research & applications* (pp. 15–22). New York, NY, USA: ACM.

Nielsen, J. (1993). Noncommand user interfaces. *Communications of ACM*, *36*(4), 83–99.

Norman, D. A. (1993). *Things that make us smart: defending human attributes in the age of the machine*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator cost of planfulness of problem solving and learning. *Cognitive Psychology*, *35*, 34–70.

O'Hara, K. P., & Payne, S. J. (1999). Planning and the user interface: the effects of lockout time and error recovery cost. *Int. J. Hum.-Comput. Stud.*, *50*(1), 41–59.

Ohno, T., Mukawa, N., & Kawato, S. (2003). Just blink your eyes: a head-free gaze tracking system. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems* (pp. 950–957). New York, NY, USA: ACM Press.

Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 281–288). New York, NY, USA: ACM.

Smith, J. D., & Graham, T. C. N. (2006). Use of eye movements for video game control. In *ACE '06: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology* (p. 20). New York, NY, USA: ACM.

Stampe, D. M., & Reingold, E. M. (1995). Selection by looking: A novel computer interface and its application to psychological research. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 467–478). Elsevier Science.

Starker, I., & Bolt, R. A. (1990). A gaze-responsive self-disclosing display. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 3–10). New York, NY, USA: ACM Press.

Surakka, V., Illi, M., & Isokoski, P. (2003). Voluntary eye movements in human–computer interaction. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 473–491). Elsevier Science.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2004). Externalization vs. internalization: The influence on problem solving performance. In *ICALT '04: Proceedings of the IEEE International Conference on Advanced Learning Technologies* (pp. 311–315). Washington, DC, USA: IEEE Computer Society.

Van Nimwegen, C., Van Oostendorp, H., & Tabachneck-Schijf, H. J. M. (2005). The role of interface style in planning during problem solving. In *In Proceedings of The 27th Annual Cognitive Science Conference* (pp. 2771–2776). Lawrence Erlbaum.

Špakov, O. (2005). Eyechess: The tutoring game with visual attentive interface. In *G. evreinov (ed.) proceedings of "alternative access: Feelings and games 2005", report b-2005-2* (pp. 81–86). Department of Computer Sciences, University of Tampere.

Ward, D. J., & MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction. *Nature*, *418*(6900), 838.

Ware, C., & Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input. *SIGCHI Bulletin*, *17*(SI), 183–188.

Zhai, S. (2003). What's in the eyes for attentive input. *Communications of ACM*, *46*(3), 34–39.