

Gazemarks - Gaze-Based Visual Placeholders to Ease Attention Switching

Dagmar Kern

Pervasive Computing and User
Interface Engineering Group
University of Duisburg-Essen
Essen, Germany-
dagmar.kern@uni-due.de

Paul Marshall

Pervasive Interaction Lab
Open University
Milton Keynes, UK
p.marshall@open.ac.uk

Albrecht Schmidt

Pervasive Computing and User
Interface Engineering Group
University of Duisburg-Essen
Essen, Germany-
albrecht.schmidt@uni-due.de

ABSTRACT

Many tasks require attention switching. For example, searching for information on one sheet of paper and then entering this information onto another one. With paper we see that people use fingers or objects as placeholders. Using these simple aids, the process of switching attention between displays can be simplified and speeded up. With large or multiple visual displays we have many tasks where both attention areas are on the screen and where using a finger as a placeholder is not suitable. One way users deal with this is to use the mouse and highlight their current focus. However, this also has its limitations – in particular in environments where there is no pointing device. Our approach is to utilize the user’s gaze position to provide a visual placeholder. The last area where a user fixated on the screen (before moving their attention away) is highlighted; we call this visual reminder a *Gazemark*. *Gazemarks* ease orientation and the resumption of the interrupted task when coming back to this display. In this paper we report on a study where the effectiveness of using *Gazemarks* was investigated, in particular we show how they can ease attention switching. Our results show faster completion times for a resumed simple visual search task when using this technique. The paper analyzes relevant parameters for the implementation of *Gazemarks* and discusses some further application areas for this approach.

Author Keywords

Eye-gaze interaction, attention switching, *Gazemarks*

ACM Classification Keywords

H.5.2 User Interfaces, Subjects: Input devices and strategies

General Terms

Human Factors, Experimentation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

Copyright 2010 ACM 978-1-60558-929-9/10/04...\$10.00.

INTRODUCTION

As our work and leisure settings become suffused with an increasingly rich array of technologies, both on and off the desktop, issues of how to deal effectively with multi-tasking, switching focus and interruptions become ever more important (e.g., [20; 27; 34]).

In order to keep track of multiple activities over different timescales, people can either maintain internal representations of the current status of tasks or they can use external artefacts and representations to offload cognitive effort [39], pulling information into working memory only when required for the guidance of action (cf. [5]).

Theoretical work in both cognitive science and HCI has considered how environments might be structured to scaffold this kind of agent-environment coupling over extended timescales. Kirsh [28] introduced the concept of *entry points*: structures or cues in the environment that represent an invitation to do something or enter somewhere. Designers create entry points such as doorways or headlines in a newspaper to draw people into an interaction (cf. [32]). People also actively structure the environment themselves to create their own entry points. For example, they might leave an email application open as a reminder to finish writing an important email the next day or leave a post-it note on a pile of papers. Kirsh [28] proposes a number of dimensions along which entry points vary, including visibility, intrusiveness and importance.

Dix and co-workers [11] discuss the related concepts of *triggers* and *placeholders*. Triggers are environmental cues that tell you when to carry out an activity (e.g., an alarm on a mobile phone), whereas placeholders say what should happen (e.g., a flight strip or to-do list) or index where an activity was left off.

We also actively structure the environment at shorter timescales, to change the nature of cognitive operations necessary to carry out tasks. For example, Kirsh [28] describes how *complementary strategies* such as pointing at or rearranging objects while counting them can facilitate memory, attention and perception. On this timescale, a placeholder could be something as simple as keeping a finger on a line of text in a book while talking to a colleague. Here, the finger acts as a spatial index that

allows the reader to quickly find the point in the text reached before the activity was interrupted.

Technologies and representations provide different resources for the user to utilize strategies to manipulate the world to facilitate cognition and perception. Most screen-based representations, for example, are more constrained than physical artefacts in terms of the resources they offer. Some include annotation, layers, text highlighting and cursors, but others, such as navigation systems, offer little flexibility in the ways that users are able to create placeholders and the like. Because screens are typically positioned at some distance from the user and are usually in a vertical orientation, we predict that people will be less likely to use physical props and their own bodies.

In this paper we investigate mechanisms by which placeholders might be used to ease attention switching between screens. We first motivate our approach by presenting a short pilot study where we investigated people's use of placeholders while sifting through information presented both on screen and on paper. Our main contribution is the concept of *Gazemarks*, a new eye-tracking technique to provide visual placeholders automatically to the user. The basic idea is to make use of the user's eye-gaze behavior to determine where a placeholder could be beneficial. A system and implementation using eye-tracking equipment is described that provides automatic placeholders. Several parameters for the design of automatic visual placeholders are experimentally assessed and discussed. In a study, the feasibility and utility of the approach is investigated: results suggest that it may be beneficial for tasks that require attention switching.

RELATED WORK

Gaze tracking

Eye tracking has been in use for more than half a century. Early work focused mainly on the application of eye tracking in the field of psychological research. More recently, it has attracted the attention of HCI researchers who have used eye-tracking data to analyze interface usability and also to interact directly with computers. The first approaches using gaze tracking for interaction with computers date back to the early 80s and 90s [7, 25]. For more information about techniques and the historical background of this approach see [14, 24].

In 1990 Jacob [25] introduced gaze-based interaction techniques, like key-based and dwell-based activation, gaze-based hot spots and gaze-based context-awareness. These techniques can be used for object selection, moving an object, eye-controlled text scrolling, menu commands and listener windows. Much research in this area followed. For example, Yamato et al. [41] and Zhai [42] investigated a combination of gaze-based and mouse pointing. Lanford [30] proposed a dwell-based technique for pointing and selection including zoom functionality. Kumar et al. [29] also used a magnification view for zooming in his “look-

press-look-release action” approach. Salvucci and Anderson [38] looked into gaze-based interaction using a button for activation as an alternative to dwell-based activation. Laqua [31] and Fono [16] focused on gaze spaces for selecting a content area or a window. Ashmore et al. [3] introduced pointing techniques involving a fisheye lens. Drewes and Schmidt [12] explored a different selection approach, using eye gestures: for example, looking around a dialog box clockwise means “OK” and anti-clockwise means “canceled”. Hyrkykari et al. [22] developed a system that offers help while reading a text in a foreign language by detecting irregularities in eye gazes while reading the words. Trends towards using eye tracking in computer games [33, 40], with mobile devices like mobile phones [13] or in cars, e.g., for fatigue detection [15] are also evident in the literature.

Multi-monitor setups / large displays

Multi-monitor setups and large displays on personal desktops enjoy great popularity. Robertson et al. [37] reported in 2001 that “as many as 20% of the Windows™ OS information worker users today run multiple monitors from one PC or laptop.” It is likely that this percentage is significantly higher today. Qualitative as well as quantitative studies confirm the productivity benefit that multi-monitor setups and large displays can confer [9, 18, 21, 26].

Multi-monitor setups and large displays also bring up usability issues; a simple mapping from a single monitor setup to a multi-monitor setup is not tenable. Robertson et al. [37] addressed these issues and proposed alternative interaction techniques. The approaches of head and eye tracking have been investigated to support mouse movement over multiple screens, [6, 3]. Grudin [18] observed that the second monitor is often used for secondary activities. These might be related to the main task performed on the primary monitor or they might just provide peripheral awareness of information, in particular that used for communicative purposes, such as email or instant messaging.

Attention switching and Interruption

Application switching, and therefore attention switching, is a part of our daily working life when using computers. Since they support multitasking, we are able to work on different projects at the same time and our attention is switched between different tasks or even between different devices. González and Mark [17] found out that “people spend an average of three minutes working on any single event before switching to another event.” They also observed that “people spend on the average somewhat more than two minutes on any use of electronic tool, application or paper document before they switch to use another tool.” [17]. Interruptions that often lead to attention switching are both self-generated, for example in reacting to the peripheral notification that a new email has arrived, or are a response to external influences, for example a colleague who asks for help. People tend to have difficulties in

resuming the task they were working on before the interruption occurred [23]. The visibility of the suspended application plays an important role. Iqbal and Horvitz [23] observed that “windows largely occluded by application windows that users access during the disruption chain took longer to recover.” The participants in their study were found to use visual indicators within the application window to remember which task they were working at.

The work described in this paper aimed to investigate ways of reducing attention-switching costs between displays. In the following sections we present the development and evaluation of the concept of *Gazemarks*, which could help to achieve this goal. We began by conducting an exploratory pilot study to investigate uses of placeholders when switching between paper and on-screen representations.

PILOT STUDY ON THE USE OF PLACEHOLDERS

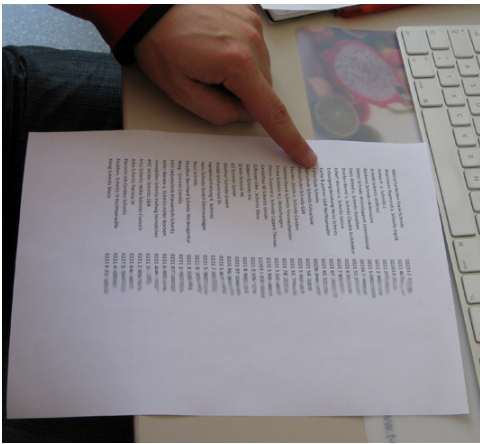


Figure 1. Finger is used as a placeholder on a paper list.

Figure 1 illustrates the inspiration for *Gazemarks*. The idea is based on the complementary strategies used on paper representations [28]. People use their fingers to mark their position on a list, enabling them to find it again quickly when cross-referencing with another representation. To test our initial prediction of differences in these complementary actions when using physical and digital representations, we conducted an informal study. We observed participants who were asked to compare a paper list and a digital list presented on a screen to find out if they used any strategies to mark a position either on the paper or on the digital list, and if so, what kind of placeholder strategies they would use.

We prepared a website with a telephone list consisting of 40 names and phone numbers. The same list was also available as a paper list, but there were three differences: for example, the phone numbers for two names were switched. 30 participants (10 female, 20 male; aged 23 to 61; mean age = 30.6) took part in the study. All participants sat in front of a PC or laptop at either their own desk or at an experimenter’s desk. At the beginning of the study they

were asked to go to the website with the telephone list. The experimenter gave them the paper list and asked them to compare the two lists to find any differences. There was no time limit for the task, but after finding the first error the experimenter aborted the task because enough information had been obtained to determine what search technique was being used.

Our hypothesis was that people would use fingers, pens or other objects to help keep track of where they were with the paper list.

The results of the study showed that 22 of the 30 participants used objects or fingers to mark the current line on the paper. They used both a finger and a pen (7); only one finger (9); two fingers (1); only a pen (4); or a ruler (1). Eight of these 22 participants also used a placeholder to keep track of the position on the display: a finger (1); a cursor (5) or by highlighting the name and/or number (2). Three participants of the 30 participants only marked the last position on the monitor by using the cursor (1), highlighting the line (1) or using the paper as a placeholder under the current line (1). Five subjects didn’t use any strategy to create placeholders on the lines either on the paper or on the screen. To summarize these findings, more than 5/6 of the observed participants used some kind of placeholder strategy on the paper to mark the current line, whereas only 1/3 used a placeholder strategy on the screen.

GAZEMARKS

The results of the pilot study on the use of placeholders added weight to our assumption that most people would try to find something to mark the position on a list when they had to switch attention between two different tasks: in this case between the physical world (paper list) and the digital world (list shown on a display). For the physical world people used physical placeholders like fingers, pens or rulers. In the digital world they tended to use the mouse cursor as a marker or highlight the last line by marking it. However, in both cases the user had to find a placeholder and actively manipulate it.

In the following section we introduce a method for visual placeholders, called *Gazemarks*, that doesn’t need active manipulation. Use of an eye tracker allows a system to be implemented to remember the last gaze position on a screen after visual attention has been switched away. Upon switching attention back to the screen, the system highlights the last gaze position (see figure 4).

Selection of parameters

There are three essential aspects that play an important role in the determination of the last gaze position:

- What is the definition of a gaze fixation?
- How long is the last conscious gaze fixation?
- How can showing a visual placeholder be avoided after a blink?

Gaze fixation

The human eyes are permanently in at least slight motion, and therefore a gaze cannot be determined as a fixation at a single pixel on a screen. A gaze on a screen is defined as a set of glances at a region with a specific radius around the first glance. That means the number of glances at this region is counted and after reaching a set threshold these glances are said to form a gaze. Figure 2 shows an example where a gaze can be identified (left) and another example where no gaze can be determined (right).

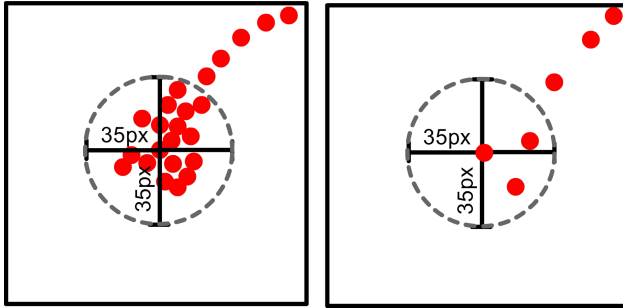


Figure 2. A gaze is defined as a specific number of glances in a predefined radius (e.g., 35px) around the first glance. The left picture demonstrates a gaze, while no gaze can be determined in the right picture.

Last conscious gaze fixation

We have to distinguish between two different kinds of gazes: on the one hand, gazes which are consciously at a specific location and on the other, unconscious gazes which are too short for the user to really recognize the content at the position. The latter kind of gaze occurs, for example, when the attention switches from one task to another and the user looks away from the display, which shows the task she is currently working on.

To avoid marking unconscious gaze positions we performed a type of user study, called a fixation study, with 13 participants (12 male, 1 female; aged 21 to 32, mean = 25.4) to find out how long the last conscious gaze fixation should be for our setup.

The setup consisted of an 8” display, a 42” display and an eye tracker (see figure 3). We used a Tobii™ eye tracker X120. The participants were asked to perform a search task on the 8” screen and their attention was randomly grabbed by animal pictures that were presented on the 42” screen. Participants were instructed to look at the 42” screen as soon as they recognized that there was a picture showing, to tell the experimenter which animal they saw and then to switch their attention back to the search task on the 8” screen. For the search task they were shown pictures with either 20 words or 20 digits and were asked to look for a specific word or digit. 10 different search task images were presented, and for each search image 2 animal pictures were shown on the large display. Altogether, each participant was requested to switch their attention 20 times, so that we recorded 260 attention switches. These were analyzed in order to calibrate the length of the last conscious gaze

position. We recorded the gazes on the 8” screen with Tobii Studio™ and determined from video analysis that the last conscious gaze position has to be longer than 0.13 seconds.



Figure 3. Setup of the fixation study. Participant performs a search task on an 8” display and his attention as grabbed by animal pictures on a 42”.

Blinking

Blinking is defined as the rapid closing and opening of the eyelid. On average, a blink takes approximately 0.3 to 0.4 seconds [36]. Humans are typically unaware of their own blinking and therefore it is necessary that blinks are ignored in the *Gazemarks* concept. Otherwise, after each blink the last gaze position would have to be marked. To achieve this we implemented a delay of 0.6 seconds before the last gaze position was marked on the screen.

Visualization options

Many different representations could be used to mark the last gaze position on the screen. The optimal representation is probably dependent upon the task that the user is performing. While searching in a list, for example, it would make sense to mark the whole of the last line. However, for searching on any graphical user interface that is not line based (e.g., a desktop environment or a navigation map), it makes more sense to mark a region or a point. Therefore, we decided to focus on the more generic variation.

We proposed three different visualization options:

1. Flag: Marking a point with a flag or an arrow (see figure 4a)
2. Spotlight: Marking a region by drawing a circle around the last gaze position. Outside the circle the representation is grayed-out (see figure 4b)
3. Focus area with gradient filter: Marking a region by drawing a circle around the last gaze position but, in contrast to option 2, using a gradient filter to illuminate seamless transition between the focus area and the non-focus area. (see figure 4c)

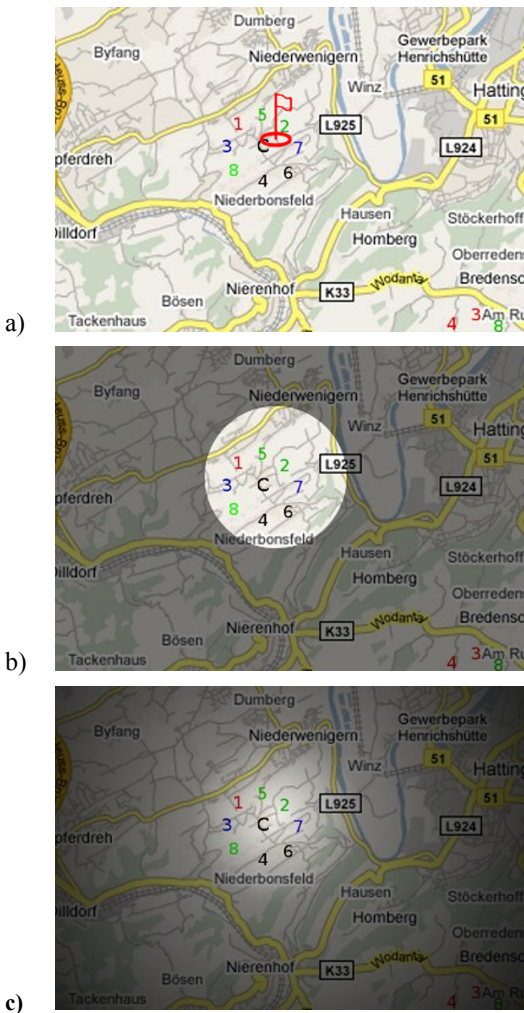


Figure 4. Three possible graphical representation of the *Gazemark* concept a) flag, b) spotlight, c) focus area with gradient filter.

We demonstrated the three different visual options on an 8” screen to 6 participants (1 female, 5 male, aged 21 to 32 mean = 24.3) and let them vote which one they preferred and indicate why they liked or didn’t like the representations. After calibrating the eye tracker to their eyes, they were presented with a map (as shown in figure 4). The task was simply to mark positions on the screen by looking at the display, looking away and looking at the display again. After looking back at the display the previous gaze position was marked either by the flag (see figure 4a), the spotlight (see figure 4b) or the focus area with gradient filter (see figure 4c). Each type of visual *Gazemark* was presented six times.

Afterwards they filled out a questionnaire. In the first question they were asked to express a preference for the three *Gazemark* options from 1 (preferred option) to 3 (least preferred option). The results are shown in table 1.

	voted as 1 st	voted as 2 nd	voted as 3 rd	average
flag	1	1	4	2.5
spotlight	2	4	0	1.7
focus area	3	1	2	1.8

Table 1. Results of users’ preferences to the three visual representation (1=preferred option to 3=least preferred option). The number of votes as well as the average value is presented.

They were also asked if they liked the representation or not and to provide an explanation. The flag was liked by 2 participants, but they didn’t specify why this was the case. 4 people didn’t like the flag representation. The main reason given was that detecting the flag was difficult because it was too small and not easy to distinguish from the map background. 5 participants liked the spotlight because it provides a larger focus area and therefore it was easier to find and they found it more accurate. Another advantage mentioned for the spotlight, was that it doesn’t hide information, in contrast to the flag representation. One participant didn’t like the spotlight because for him it was too vague. The focus area with gradient filter option was liked by 4 subjects, because the focus was clear, the representation guides the gaze automatically to the last gaze point and the gradient filter avoids sharp edges in the representation. 2 participants didn’t like this representation because they found the grayed out areas more distracting and they were concerned that if the position indicated for the last gaze was not correct it would be much harder to find information in these dark areas.

The results of this study indicated that just marking a single point using a flag was less acceptable to users because it is more difficult to find, especially when the background is colorful. Marking a region around the last gaze position seems to be more promising. The last gaze position is easier to find and the users perceive this technique to be more accurate; it is more robust against minor deviations caused by the eye tracker. Taking the disadvantages of the focus area with gradient filter into account we decided to select the spotlight representation in our prototype implementation. Nevertheless, the focus area with gradient filter might also be an option, especially after adjusting the filter so that the farthest points away from the focus point are not darker than in the spotlight example.

Taking on board the concern that finding information in dark areas might be harder if the last gaze was not correct, we decided that the *Gazemark* should be only visible for 3 seconds at most and should disappear directly after finding the last gaze position again.

IMPLEMENTATION

We implemented a prototype of the *Gazemarks* concept to research if it might help users to reorient faster on a screen after switching attention. We chose a setting for our lab

study where the user switches attention between a small display and a large screen.

Hardware

The following hardware components were used to demonstrate the concept. As an eye tracker, we used a Tobii X120¹ (with a data rate of 120Hz) and a consumer PC connected to a 42" display and an 8" display. To validate the utility of our approach on small as well as large screens, the last gaze position was marked on the 8" display. The eye tracker was placed underneath the display to capture the user's gaze. The eye tracker delivered its data via a LAN connection to the PC, where the data analysis was performed.

Software

For the communication between the eye tracker and our application we utilized EIToolkit². EIToolkit [19] is a component-based architecture which allows proxy-like objects, called stubs, to exchange messages over a general communication area, e.g., via UDP. A main benefit of this toolkit is that different hardware and software platforms as well as programming languages can be used. The stub concept allows hardware components to be replaced without changing anything in the application logic. In our case this means our application can run with any eye tracking hardware.

Eye tracker component

Using the Tobii SDK³, we developed an EIToolkit stub in C# that provides means for calibration and that receives data from the eye tracker, transforms them to normalized coordinates [0,1] and transfers data to the EIToolkit general communication area using UDP. The reason for normalizing the coordinates is that we are able to show gaze points on any other display with any resolution, this enables the experimenter to observe gaze points during a study on an observation screen.

Gazemark application

Our *Gazemark* application is implemented in Java. The application provides two mechanisms for showing the last gaze position: either on an image or using a transparent window on any screen background. The program registers with EIToolkit that it listens to eye tracking messages. It receives either valid data [0, 1], when the user looks at the display or invalid data [-1, -1], when the user looks away from the display or blinks. Received data are collected and stored in a vector so that the last gaze position can be calculated after the program has received invalid data for

0.6 seconds. Given that eye trackers send eye-tracking data with a specific data rate, times can be translated into a number of received values. That means that after receiving (0.6*eye tracker frequency) invalid values the last gaze position will be calculated, in our case where we have used a data rate of 120Hz, that means after receiving 72 invalid values. The last gaze position is determined by examining the last valid data in reverse order. In the fixation study we found that the duration of the last gaze position is 0.13sec. With a data rate of 120Hz that means 16 values have to be in the fixation radius around a valid value. The fixation radius is set to 10% of the width of the screen resolution; this makes the program independent of the screen resolution that is used. In our case the resolution was 600x800 so that we had a fixation radius of 80px. After indicating a fixation, this point was highlighted by the spotlight representation. To avoid distraction by marking a position that the user doesn't want to return to or by marking a false position, the spotlight is only shown for 3 seconds or as long as the user doesn't look at the highlighted area. As soon as a gaze fixation is recognized in the highlighted area it fades out in 100ms.

EXPERIMENT

We ran a user study using the implemented prototype to compare two conditions: a control condition of performing a search task on a screen without any visual placeholder and performing a search task with *Gazemarks*. Using the feedback from the demonstration of the visual options we decided to use the spotlight. The hypothesis that we tested was that users would be able to perform a simple visual search task faster when the last gaze position is highlighted.

Setup

We placed the participants in front of a 42" screen, 8" screen and an eye tracker, which tracked glances towards the 8" screen (see figure 5)⁴. We asked participants to perform an attention-switching task with a visual map searching task implemented on the small screen and a textual reasoning task presented on the large screen. We showed questions on the large display to direct participants' attention away from the small display where the searching task was performed.

On the 8" screen a map (chosen because of its colorful background) was shown with six letters randomly placed on the screen. Around each letter, eight numbers were equally spaced in a circle (see figure 4). Two different maps were prepared to assign them in counterbalanced order for the two conditions.

¹ http://www.tobii.com/scientific_research/products_services/eye_tracking_hardware/tobii_x120_eye_tracker.aspx

² <http://www.eitoolkit.de>

³ http://www.tobii.com/scientific_research/products_services/eye_tracking_software/tobii_software_development_kit.aspx

⁴ Note in figure 5 that the participant has his hands on a steering wheel. This two-display set-up is also used as a driving simulator in our lab. However, the steering wheel played no role in this study other than as a place for the participants to rest their hands

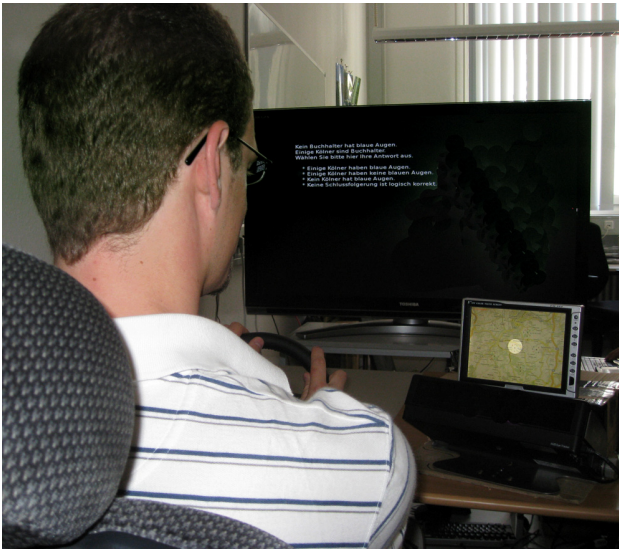


Figure 5. Experimental setup. Participant performs search tasks on the 8" display. The 42" display is used for providing questions to distract participant's attention away from the search task.

Design

A within-subjects design was employed, with each subject performing the task in both conditions (in counterbalanced order). First of all the eye tracker had to be calibrated to the users' eyes. Afterwards participants were introduced to the visual search task. The goal of the search task was to find letters on a map shown on the small display (as shown in figure 4). The participant was initially told to find one of the letters, tell the experimenter upon finding the letter and then look away from the small display at the 42" display where two questions were shown, one after the other. After answering both questions, an arrow appeared indicating one of eight directions. The subject had to look back at the small display, find the same letter again and tell the experimenter which number is shown in the indicated direction. This procedure was then repeated for each of the six letters before switching to the other condition. The experimenter asked for the six letters in counterbalanced order. The questions displayed on the 42" screen were selected from an IQ questionnaire, and were designed to be sufficiently challenging to fully engage the participants' attention. Examples of questions asked included "What number completes the following series? 5 15 12 4 12 9?" and "Which one of these five choices complete the best analogy? Finger is to Hand as Leaf is to: a) Twig, b) Tree c) Branch, d) Blossom, e) Bark". At the end, participants were given a questionnaire and asked to rate aspects of the *Gazemark* concept. Further open-text explanations for their statements were collected (e.g., advantages and disadvantages of the concept), as well as demographic data. The duration of the experiment was dependent on how quickly participants answered the question. It took between 10 and 20 minutes.

As a dependent variable we assessed search time by measuring the time between looking back at the navigation display and finding the requested letter again. The measurement was generated automatically by starting a timer in our software after recognizing the first gaze at the 8" screen and stopping this timer as soon as a gaze was inside the area, which would match the highlighted area around the last gaze position in the spotlight condition. If the participant was not able to find the letter in 3000ms this value was counted for the search time indicating that the participant was not able to find the letter in time.

Participants

16 participants took part in the study: 3 female and 13 male, aged 23 to 52 (mean = 28.44).

RESULTS

In the following we present and discuss the quantitative as well as the qualitative findings of our user study.

Analysis of search times

As the participants' search time had a ceiling value of 3000 ms, search times in the two conditions were compared with a non-parametric Wilcoxon signed-rank test. The median was used as a measure of central tendency; effect size is reported as Pearson's r .

As predicted, participants were found to be considerably faster in searching for letters on the map with *Gazemarks* (Mdn = 625.75 ms) than without (Mdn = 1999.50), $T = 1$, $p < 0.001$, $r = -0.87$.

Analysis of questionnaire results

The questionnaires asked participants to report whether they liked the spotlight representation. 15 of the 16 subjects liked the representation, explaining that this was because "the target was faster to find" (6 people); "focus leads the attention to the essential" (3 people), "it helps to orientate" (3 people) and "it reduces mental workload" (1 person). Only 1 participant didn't like the spotlight. He said that he didn't notice it or just ignored it.

In the next question participants were asked to score how helpful the presented system was, indicating their preference by crossing a line on a continuous Likert scale. The distance along the line was then measured and translated into a scale ranging from 0 (not at all helpful) to 5 (very helpful). The results showed that they typically found the spotlight helpful or very helpful (mean value: 4.09, standard deviation: 1.0). Only one participant rated helpfulness as less than 3.6 (0.65). The accuracy was addressed in the following question. Participants had to score on a scale from 0 (very inaccurate) to 5 (very accurate) how accurate the highlighted region corresponded to their last gaze position. The mean value was 3.24 with a standard deviation of 1.06. The accuracy depends highly on the eye tracker calibration and the movement the user carries out during the experiment, which can lead to inaccurate behavior of the system. This result also confirms our assumption that highlighting a region is more useful

than marking one single point. But more consideration must be given to how we can make the system more accurate.

Participants were also asked if they found *Gazemarks*, to make sense in general, independent of the presented representation. They indicated their preferences on a scale from 0 (completely senseless) to 5 (very sensible). Nearly all voted *Gazemarks* as very sensible (mean value 4.26, standard deviation 0.53).

Answers given to the open questions about advantages and disadvantages also indicate the benefits that such a system might have. The main advantages suggested were saving time, and the speed with which a search task could be performed, which was mentioned by 11 participants. 3 people liked the aid to memory of not having to remember the last position or letter. The main disadvantage was seen to be the loss of contextual information, because other important information in the greyed out area is harder to recognize (6 answers). One participant mentioned that it might get annoying after a while. Three also mentioned that the reliability has to be high otherwise it might be distracting.

In the last section participants were given the opportunity to suggest an alternative representation of the last gaze position. 3 suggested they would prefer marking only a single point with a flag, an arrow, or a pulsating point. 2 indicated they would like to zoom in to the focus area, and 1 suggested a fisheye view. 2 participants suggested a flashing light or an animation around the last gaze position.

Marking a single point makes the potential inaccuracy of the system more obvious, as we already observed in the pretest we performed. Zooming in to the focus area is counter-productive as it also means losing context information. Blinking and flashing might be applicable in some settings, but could become annoying. The fisheye seems to be an interesting approach to highlight the last gaze position and will be considered in future work.

DISCUSSION

In this paper we have described *Gazemarks* as an approach to automatically create visual placeholders based on users' eye-gaze. In tasks where users are required to switch their visual attention between displays, re-orienting to the last position of interest when coming back to a display can be a problem. When using paper-based representations, people often use fingers or physical props as placeholders, facilitating re-orientation when returning to focus on the representation. *Gazemarks* are designed as visual aids on digital displays to provide these placeholders automatically and hence support the user's attention switching between displays.

It remains to be seen in which contexts the *Gazemarks* concept might be of greatest utility. However, we propose three application domains where *Gazemarks* may have a role to play:

Desktop environment

As suggested by our pilot study, a potential benefit could be seen in work-related tasks where people have to type in paper forms or to cross-reference paper with digital information: for example, in an insurance company or in a university administration department, where results of examinations are filled into databases. *Gazemarks* could help users to find the input fields faster. Furthermore, an advantage might be seen in multi-monitor setups. As Grudin [18] highlighted, the secondary screen is often used as a space for supporting a primary task presented on the main screen, e.g., checking lists of variables while debugging program code. *Gazemarks* could facilitate finding the variable of interest quicker on the secondary screen. Furthermore, a role on the primary screen would also make sense, for example, in marking the last position in the program code, to extend our programming example.

We also see a potential benefit in a normal working environment where there are many interruptions by clients or colleagues, which can require turning attention away from the screen and from the current task. After looking back to the screen, *Gazemarks* could help the user to reorient on the screen, even in cases where the interruption takes an extended time, to remember what the unfinished task was. *Gazemarks* could also be used in a single monitor setup to support task switching between multiple windows. The last gaze position on each window before minimizing it could be highlighted after maximizing it again.

Small devices / mobile phones

Dickie et al. [10] have already illustrated how to use an eye tracker with a mobile phone to handle interruptions. A video sequence or a speed-reading application is stopped as soon as the user stops looking at the screen. With our approach, the last gaze position would also be highlighted. This probably doesn't make sense in the case of a video, but could support normal reading on mobile phones or while browsing the internet on small devices.

User Interfaces in cars

Another application domain that seems particularly promising is user interfaces in cars. The main task in a car is always driving. Secondary tasks like interacting with navigation or infotainment systems force the driver to split her attention, which leads to distraction from driving. One of the key challenges in designing user interfaces for cars is to keep driver distraction to a minimum, which also implies that each interaction sequence has to be interruptible. The performance of the secondary task is continually interrupted by driving and the driver often needs a number of attention shifts to complete it [35]. Interruptibility is therefore a design goal for applications in cars and stressed in guidelines for automotive user interfaces [1, 8].

For example, in driving a car while interacting with an infotainment system, the driver has always to split her attention between the primary task of driving and this secondary task. The time before the driver is able to return her attention to the secondary task again is filled with other

tasks that are demanding of attention and therefore, it can sometimes be difficult to reorient on the display. For example, the driver may forget her position in a list of music when searching for a title. A *Gazemark* would help her to remember where she was. With our approach, the time taken to reorient to the screen could be minimized and consequently the time looking away from the road could be reduced.

Navigation systems would be an interesting special case for *Gazemarks*: because of the fact that the car is in motion, the visual representation of a navigation map moves on the display as well. Consequently, highlighting the last gaze position without taking this into account would lead to confusion for the driver. Therefore the gaze point has to be set on the moving map such that the visual placeholder is moving, too.

Exploring the applications of *Gazemarks* will be a focus of future work.

CONCLUSION

We have introduced *Gazemarks*, an approach to automatically create visual placeholders based on users' eye-gaze, which add to the existing repertoire of complementary actions people use to keep track of where they are in an interrupted task or sub-task.

The paper describes the basic idea and an implementation of this approach. We have discussed implementation issues such as recognition of the last gaze position, filtering out blinking, and different visualization approaches for implementing a *Gazemark* system and give recommendations. Using the prototype implementation we conducted a study with 16 subjects where we compared the performance on a task that involved attention switching with and without *Gazemarks*. The results show a clear benefit in using *Gazemarks* for our simple visual search task.

We propose that *Gazemarks* might be used in different contexts ranging from multi-display setups, to mixed reality environments and in automotive user interfaces. By providing visual placeholders this approach lowers the perceptual cost of attention switching for the user. In particular, in the automotive domain this might increase safety, as users may be more willing to interrupt interaction tasks that require visual attention and interleave it with more gazes to the road and surroundings. In future studies we plan to explore *Gazemarks* in the context of driving scenarios (in simulations as well as in real driving situations). In such studies we expect to be able to assess the impact on driving performance as well as interaction performance.

ACKNOWLEDGMENTS

We would like to thank Lukasz Wolejko-Wolejszo, Simon Hufen and Fabian Schulte for implementing the *Gazemark* prototype and supporting us with the execution of the user study.

We are very grateful to those who participated in our user studies and provided us with feedback on our *Gazemarks* implementation.

REFERENCES

1. Alliance of Automobile Manufacturers: Statement of Principles, Criteria and Verification Procedures on Driver Interaction with Advanced In-Vehicle Information and Communication Systems, 2003.
2. Anderson, R.E. Social impacts of computing: Codes of professional ethics. *Social Science Computing Review* 10, 2 (1992), 453-469.
3. Ashdown, M., Oka, K., Sato, Y. Combining head tracking and mouse input for a gui on multiple monitors. Ext. Abstracts CHI 2005, ACM Press (2005), pp 1188-1191.
4. Ashmore, M., Duchowski, A. T., and Shoemaker, G. Efficient Eye Pointing with a FishEye Lens. *Proc Graphics Interface 2005*, pp. 203-10.
5. Ballard, D., Hayhoe, M., Pook, P., Rao, R. Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 723-767, (1997).
6. Benko, H., Feiner, F. Multi-Monitor Mouse. Ext. Abstracts CHI 2005. ACM Press (2005), pp 1208-1211.
7. Bolt, R. A. Gaze-orchestrated dynamic windows. *SIGGRAPH '81*, ACM Press (1981), 109-119.
8. Commission of the European Communities: Commission Recommendation of 22 December 2006 on safe and efficient in-vehicle information and communication systems: Update of the European Statement of Principles on human machine interface, 2006.
9. Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., Starkweather, G. Toward characterizing the productivity benefits of very large displays. *Proc. Interact 2003*, IOS Press (2003), 9-16.
10. Dickie, C., Hart, J., Vertegaal, R., Eiser, A. LookPoint: an evaluation of eye input for hands-free switching of input devices between multiple computers. *Proc. OZCHI 2006*, pp 119-126.
11. Dix A, Ramduny-Ellis D, Wilkinson J. Trigger analysis—understanding broken tasks. In: *Diaper D, Stanton N (eds) The handbook of task analysis for human-computer interaction*. Lawrence Erlbaum Associates, London (2002).
12. Drewes, H., Schmidt A., Interacting with the Computer Using Gaze Gestures. *Proc. INTERACT 2007*, pp 475–488.
13. Drewes, H., De Luca, A., Schmidt, A. Eye-Gaze Interaction for Mobile Phones. *Proc. Mobility 2007*.
14. Duchowski, A. T. Eye Tracking Methodology: Theory and Practice. Springer-Verlag New York (2003), ISBN:1852336668

15. Eriksson, M. Papanikotopoulos, N.P. (1997). Eye-tracking for detection of driver fatigue. *Proc. Seventeenth Annual Conference of the Cognitive Science Society 1995*, pp. 212-217.
16. Fono, D., Vertegaal, R. EyeWindows: Evaluation of Eye-Controlled Zooming Windows for Focus Selection. *Proc. CHI 2005*, ACM Press (2005), pp 151 - 160.
17. Gonzalez, V.M. and G. Mark, Constant, constant, multi-tasking craziness: managing multiple working spheres. *Proc. CHI 2004*, ACM Press (2004), pp 113-120.
18. Grudin, J. Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. In *Proc. CHI 2002*, ACM Press (2002), 458-465.
19. Holleis, P., Schmidt, A. MAKEIT: Integrate User Interaction Times in the Design Process of Mobile Applications. *Proc. of the Sixth International Conference on Pervasive Computing, Pervasive'08*. Springer LNCS Sydney, Australia (2008), S. 56-74.
20. Horrey, W. J., Wickens, C. D. Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors*, 46(4), 611-624, (2004).
21. Hutchings, D., Czerwinski, M., Smith, G., Meyers, B., Robertson, G. Display space usage and window management operation comparisons between single monitor and multiple monitor users. *Proc. AVI 2004*, ACM Press (2004), pp 32-39.
22. Hyrskykari, A., Majaranta, P., Aaltonen, A., Rähkä, K. Design issues of iDict: A gaze-assisted translation aid. *Proc of the Eye Tracking Research and Applications Symposium 2000*, ACM Press (2000), pp. 9–14.
23. Iqbal, S.T., Horvitz, E. Disruption and recovery of computing tasks: Field study, analysis, and directions. *Proc. CHI 2007*, ACM Press (2007), pp 677-686.
24. Jacob, R. J. K., Karn, K. S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (Section commentary). In J. Hyona, R. Radach, & H. Deubel (Eds.), *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Oxford: Elsevier Science (2003).
25. Jacob, R. J. What you look at is what you get: eye movement-based interaction techniques. *Proc. CHI '90*, ACM Press (1990), 11-18.
26. Kang, Y., Stasko, J. Lightweight task/application performance using single versus multiple monitors: a comparative study. In *Proc. GI 2008*, pp 17-24.
27. Kirsh, D. A Few Thoughts on Cognitive Overload, *Intellectica*, 2000 pp 19-51.
28. Kirsh, D. The Context of Work, *Human computer Interaction*, 2001 Vol 16(2-4), pp. 305-322
29. Kumar, M., Paepcke, A., Winograd, T. EyePoint: Practical Pointing and Selection Using Gaze and Keyboard. *Proc CHI 2007*, ACM Press (2007), pp 421-430 .
30. Lankford, C. Effective Eye-Gaze Input into Windows. *Proc. ETRA 2000: Eye Tracking Research & Applications Symposium*. ACM Press (2000). pp. 23-27.
31. Laqua, S., Bandara, S. U., Sasse, M. A.. GazeSpace: Eye Gaze Controlled Content Spaces. *Proc BCS HCI Group Conference*. (2007)
32. Lidwell, W., Holden, K., Butler, J.: *Universal Principles of Design*. Rockport (2005)
33. Lin, C-S., Huan, C-C., Chan, C-N., Yeh, M-S., Chiu C-C. The design of a computer game using an eye tracking device for eye's activity rehabilitation, *Optics and Lasers in Engineering* 42(1), 2004, pp. 91–108.
34. Mark, G., Gonzalez, V., Harris, J. (2005). No Task Left Behind? Examining the Nature of Fragmented Ext. *Abstracts CHI 2005*, ACM Press (2005) , pp. 321-330.
35. Monk, C. A., Boehm-Davis, D. A., Trafton, J. G.. The attentional costs of interrupting task performance at various stages. *Proc. Human factors and Ergonomics Society 46th annual meeting 2002*.
36. Moses RA. in *Adler's Physiology of the eye clinical application*, Robert A. Moses., Ed. (Mosby, 1981), chap. 1, pp. 1-15.
37. Robertson, G.G., Czerwinski, M., Baudisch, P., Meyers, B., Robbins, D., Smith, G., Tan, D. Large Display user experience. In *IEEE CG&A special issue on large displays*, 25(4), pp. 44-51, (2005).
38. Salvucci, D. D. Intelligent Gaze-Added Interfaces. *Proc CHI 2000*, ACM Press (2000). pp. 273-80.
39. Scaife, M. , Rogers, Y. External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, 45, 185-213, (1996).
40. Smith, J.D., Graham, T.C. Use of eye movements for video game control. *Proc. ACE 2006*.
41. Yamato, M., Monden, A., Matsumoto, K-c., Inoue, K., Torii, K. Quick Button Selection with Eye Gazing for General GUI Environment. *Proc. of International Conference on Software: Theory and Practice (ICS2000)*, pp.712-719.
42. Zhai, S., Morimoto, C., Ihde, S. Manual and Gaze Input Cascaded (MAGIC) Pointing. *Proc. CHI 1999*, ACM Press (1999), pp. 246-53, 1999.