

Look & Touch: Gaze-supported Target Acquisition

Sophie Stellmach and Raimund Dachsel
User Interface & Software Engineering Group
University of Magdeburg
Magdeburg, Germany
{stellmach, dachsel}@acm.org

ABSTRACT

While eye tracking has a high potential for fast selection tasks, it is often regarded as error-prone and unnatural, especially for gaze-only interaction. To improve on that, we propose *gaze-supported interaction* as a more natural and effective way combining a user's gaze with touch input from a handheld device. In particular, we contribute a set of novel and practical gaze-supported selection techniques for distant displays. Designed according to the principle *gaze suggests, touch confirms* they include an enhanced gaze-directed cursor, local zoom lenses and more elaborated techniques utilizing manual fine positioning of the cursor via touch. In a comprehensive user study with 24 participants, we investigated the potential of these techniques for different target sizes and distances. All novel techniques outperformed a simple gaze-directed cursor and showed individual advantages. In particular those techniques using touch for fine cursor adjustments (*MAGIC touch*) and for cycling through a list of possible close-to-gaze targets (*MAGIC tab*) demonstrated a high overall performance and usability.

Author Keywords

Gaze input; mobile touch interaction; selection; target acquisition; gaze-supported interaction

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces - Evaluation/methodology – Input devices and strategies

General Terms

Design, Human Factors

INTRODUCTION

The diversity of display setups is increasing and with that is the need for more efficient means to interact with them. While traditional mouse input works excellent for pointing tasks in desktop environments, it does not apply well for situations in which the user is standing in front of a powerwall or sitting on a couch to interact with a large-sized television set.

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 to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Copyright 2012 ACM 978-1-4503-1015-4/12/05...\$10.00.

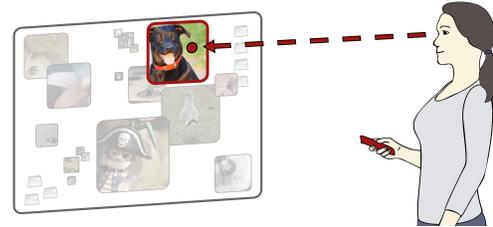


Figure 1. Basic idea: Gaze-supported interaction in combination with a handheld touchscreen and a distant display.

Regardless of the specific interaction setup used, the selection of targets is one of the fundamental tasks that need to be supported in any application. Gaze is a promising input modality to bridge the gap between a user and a distant display as illustrated in Figure 1. In this respect, it can even be a more efficient means for pointing tasks than traditional input devices [4, 21, 24].

Even though target acquisition seems to be a simple process which basically involves positioning a cursor and confirming a selection, it imposes several challenges when using eye gaze. Among them are *inherent inaccuracies* caused by the physiological nature of our eyes and by measurement errors of the tracking systems which lead to *jitter* and *offsets* [17, 23]. Thus, for more precise selections it is essential to address these two problems. Jittering can, for example, be compensated by stabilizing the gaze cursor (e.g., [28]). Offsets are difficult to handle as the degree of the offset is usually not known. Common solutions for gaze interaction include:

- *Large-sized* or *magnified* graphical user interfaces (GUIs) [2, 13, 14, 20],
- A combination of *gaze and manual (mouse) input* to perform exact positioning manually [7, 27],
- Invisibly *expand targets in motor space* [16],
- *Intelligent* algorithms to estimate the object of interest [17, 18, 23].

In this paper, we focus on the first two approaches as they provide high flexibility for diverse settings. They neither require any changes of conventional GUIs, nor substantial a priori knowledge about the distribution of items.

Beside accuracy issues, the *Midas touch* problem [11], the unintentional execution of actions, is often described as one of the major challenges for gaze-based interaction. The combination of gaze with other input modalities can solve these problems [5, 18, 22, 24]. Nevertheless, up to now dwell time activations are mostly used for gaze-based selections (e.g., [9,

10, 15]). This especially supports people who are not able to use their hands to interact with a digital system, for example, because of disabilities or because their hands are busy.

To alleviate the stated problems, this work aims at supporting precise target acquisition in a natural way while still maintaining sufficient task performance. In this context, eye movements may serve well as a supporting interaction channel in combination with other input modalities, such as speech, hand and body gestures, and mobile devices. We call this type of interaction *gaze-supported interaction*. For this work, we propose to conveniently use eye gaze in combination with touch input from a handheld device following the principle *gaze suggests* and *touch confirms*. Note that we do not aim at replacing or beating the mouse, but instead motivate gaze-supported interaction for diverse display setups. This may include public screens or large projection walls, for which we see a particularly high potential for this type of distant interaction. Another specific goal of this work is to support accurate selections even of *small* and *densely positioned* targets.

In this paper, we contribute a set of practical and novel gaze-supported selection techniques using a combination of gaze and touch input from a handheld touchscreen. The techniques utilize principles such as target expansions and separating coarse and fine positioning of a selection cursor by means of gaze vs. manual touch input. We carefully investigated the design space and developed five solutions which were tested and compared in a comprehensive user study. In particular those techniques using touch for fine cursor adjustments (*MAGIC touch*) and for cycling through a list of possible close-to-gaze targets (*MAGIC tab*) resulted in very promising results with respect to their overall performance and perceived usability.

The remaining paper is structured as follows: First, we discuss how gaze has been used for target acquisition tasks in previous work. Based on that we elaborated five gaze-supported selection techniques that are described in the *Design* section. These techniques have been tested by 24 participants which we report in the *User Study* section. The paper concludes with a discussion of the results from the user study and an outlook to future work.

RELATED WORK

In general, gaze input has a high potential for fast pointing tasks and may even outperform traditional selection devices such as a mouse [4, 21, 24]. In the following, we will investigate gaze-based selections in combination or in context with (1) target expansions, (2) a manual input device, (3) distant displays, and (4) a supportive modality.

Target Expansions for Gaze-based Selections

One way to ease gaze-based target selections is to magnify the display either locally at the point-of-regard [2, 8, 13, 14, 16, 20, 22] or globally [1, 3, 9]. Empirical evidence shows that eye pointing speed and accuracy can be improved by target expansions [2, 16, 20]. In this respect, Ashmore et al. [2] describe different types of dwell activated fisheye lenses that either follow the user's gaze at all times (*eye-slaved*) or remain fixed at a fixated position. They point out that users

would find it irritating if the lens is active at all times. Fono and Vertegaal [8] use eye input with either dwell time or a key for zoom activation. The latter was preferred by users over automatic activations. Kumar et al. [13] present *EyePoint*: a combination of gaze and keyboard input for selecting GUI elements. For this, they introduce the concept of *look-press-look-release*. On pressing a keyboard button, the viewed region is enlarged. Different keys on the keyboard are assigned to various actions, such as single click, mouse over, and double click. However, the magnified view is based on a screen capture. Thus, no dynamics (e.g., an animation) are possible during this mode. Hansen et al. [9] present *StarGazer*: a 3D interface displaying groups of keyboard characters for gaze-only target selections (in this case gaze typing) using continuous pan and zoom. The point-of-interest moves towards the center of the screen while zooming and thus provides a better feedback for more precise selections. Skovsgaard et al. [20] use a local zooming lens to gradually increase the effective size of targets. They distinguish between discrete and continuous zooming tools for step-wise zooming. While their zooming tools improve hit rates, it takes longer to perform a selection compared to the non-zooming interface.

Gaze & Manual Input

Zhai et al. [27] present the *MAGIC* (i.e., *Manual And Gaze Input Cascade*) pointing technique, a combination of mouse and gaze input for fast item selections. The idea is to warp the cursor to the vicinity of the user's point-of-regard prior to moving the mouse. Then the cursor can be manually positioned using the mouse for more precise selections. Drewes and Schmidt [7] point out that the problem of this technique is *overshooting*: the cursor is only set to the gaze position after a mouse timeout and after the mouse is then moved again. Thus, the mouse is already in motion when the pointer is positioned which is difficult to coordinate. Zhai et al. [27] acknowledge this problem and propose to dampen the cursor movement based on the initial motion vector and distance to the previous cursor position. Instead, Drewes and Schmidt [7] use a touch-sensitive mouse button. Thus, when touching the mouse key (and before performing any mouse movement), the mouse pointer is set to the gaze position.

Gaze & Distant Displays

Several studies indicate that gaze *can* be faster than mouse input [9, 11, 13, 24]. In this respect, gaze-based input is acknowledged a particularly high potential for a more convenient interaction with high-density information on large (e.g., public) displays [1, 7, 9, 19, 26]. However, a main issue remains for accuracy versus speed. Kumar et al. [13] report, for example, that while task completion times were similar for gaze and mouse conditions, error rates are usually much higher for gaze input. San Agustin et al. [19] present a gaze-based navigation of a *digital bulletin board*. If messages lie on top of each other, the user can look at them and they will get separated from each other. Yoo et al. [26] combine gaze data (head orientation) and hand gestures for the interaction with large-sized displays. A 3D push-and-pull gesture is used to control the zoom for different applications, such as a geographical information system.

Gaze-supported Selection

Ware and Mikaelian [24] compare three gaze-supported selection techniques: a button press, gaze dwell, and an on-screen button to confirm a selection. Although dwell time and button activations resulted in similar completion times, dwell-based activations were more error-prone. Salvucci and Anderson [18] also use a button press for confirming a gaze-based selection for which they report errors due to a *leave-before-click* issue. This means that the gaze was already fixating a new target when pressing the button. After all, Ware and Mikaelian [24] conclude that eye tracking can be used as a *fast selection* device, if the target size is sufficiently large. Monden et al. [17, 25] present three gaze-supported selection techniques in combination with a mouse. First, when clicking the mouse, the closest item to the current gaze position is selected. This is especially useful for selecting small-sized targets, however, it has no advantage for closely positioned objects. Second, the cursor position can be manually adjusted with the mouse. Third, the first two approaches were combined and showed to be the fastest technique, even beating mouse input.

So far, few researchers have investigated a combination of gaze with a mobile touch-enabled device for interacting with distant displays (e.g., [7, 22]). As indicated by Stellmach et al. [22], it is very important to design the interaction with the mobile device in a way that the need to switch the user's visual attention between the distant and local (mobile) display is minimized. In this respect, Cho et al. [6] compare tilt, button, and a click wheel (as on the Apple iPod) input for exploring image collections. While participants found tilt most interesting to use, buttons offered the most control.

DESIGN OF GAZE-SUPPORTED SELECTION

For the design of *gaze-supported* target selection techniques, the first design decision is the choice of an additional input modality. In this work, we decided to use a small touch-enabled device, because smartphones are very commonplace and easy to use. It can be held in the user's hand and combined with his/her direction of gaze. For this, we assume that the eyes are tracked to deliver gaze positioning information with respect to a distant display. In addition, the handheld display allows for confirming a selection and for additional functionality addressing the problems of small targets and targets being too close to each other to be easily selected with gaze input only. Moreover, for advancing gaze-supported selection in combination with a mobile touch display, we elaborated the following design goals:

- Possibility to interact with standard GUIs
- Possibility to select *small* and *closely positioned* targets
- Prevent performing involuntary actions (*Midas Touch*)
- Subtle gaze interaction - should not overwhelm the user
- Support of *eyes-free* interaction with the mobile device
- One-handed interaction: hold mobile device in one hand and interact with the thumb only (based on [22])

Considering these design goals, we envisioned and investigated different approaches for combining gaze and touch-enabled input. We thereby distinguish between three basic

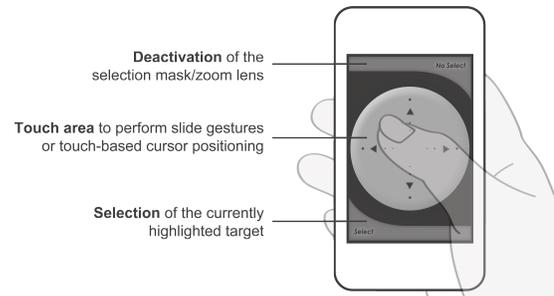


Figure 2. Interface prototype for the mobile touch device.

gaze-supported selection types that are further discussed in this paper (specific variants in *italic*):

- Gaze-directed cursor (basic selection)
- Gaze-supported manual selection: *MAGIC touch* and *tab*
- Gaze-supported expanded target selection: *Eye-slaved* and *semi-fixed zoom lens*

All types use a different combination of touch and gaze, but follow the underlying principle *gaze suggests* and *touch confirms*. The basic interaction vocabulary – besides gaze for cursor positioning – is briefly outlined here: On mobile touch-enabled devices held in one hand, the simplest way of interaction is to tap a particular button which is typically used for confirming a selection, changing modes or activating other functions. Another one is to use a sliding gesture for controlling a numeric value such as a zoom factor. Next, the thumb can be positioned and dragged around continuously within a particular area, which can often be used for panning and positioning tasks. More complex gestures are not easy to be performed with a single finger in eyes-free operation and are therefore not further considered here (cf. design goals five and six). Finally, built-in sensors such as accelerometers can be employed to recognize tilting or rotating the device. Since this is continuous input, too, it can be used for zooming, panning or adjusting other values.

Using these basic interaction techniques and combining them with gaze allowed us to contribute novel gaze-supported selection techniques. To help preventing involuntary actions, the mobile touch device is used to issue a selection event (*touch confirms*), thus avoiding the Midas Touch effect (cf. design goal 3). For this, we developed an interface prototype that is shown in Figure 2. For our current prototype, we use virtual buttons on the mobile device to confirm selections as they offer more control compared to tilt input [6]. Further details (including the terms *selection mask* and *zoom lens* used in Figure 2) are discussed in context with the individual selection techniques in the following.

For each technique, we provide a brief description first. Then, we go into detail about the specific mapping of interaction methods to the envisioned functionality as we have used it for our implementation. Finally, we briefly discuss particular advantages and disadvantages of each technique. Please note that we explicitly aim for diverse techniques that can be later combined in a *complex interaction set* benefitting from their particular advantages.

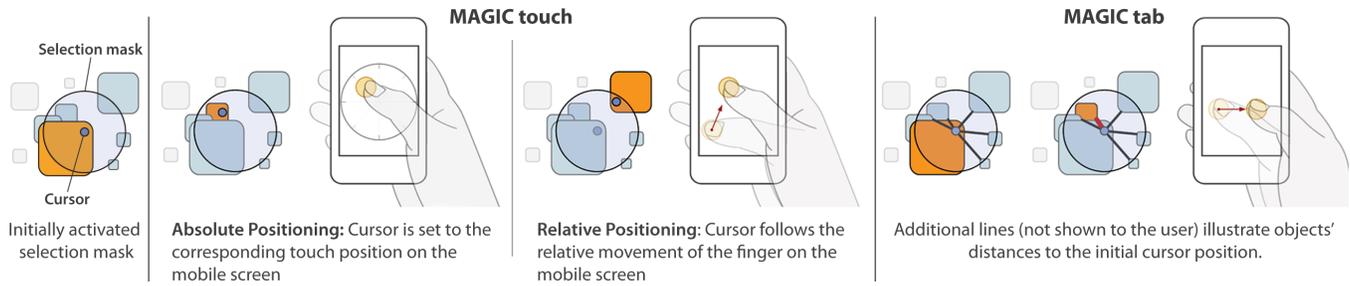


Figure 3. Manual gaze-supported selection: Initial situation (left), MAGIC touch – Absolute and relative adjustment of the cursor within the selection mask (middle) and MAGIC tab – Slide gesture to iterate through an item collection to select a target (right).

Gaze-directed Cursor

A *gaze-directed cursor* is the most basic technique, depicting an icon at the user’s point-of-regard (*gaze suggests*). Internally it is represented by a single gaze position (in contrast to area cursors [12]). This is a common approach for substituting mouse with gaze input (e.g., [13, 25]). Different zones on the mobile display can be used for supporting toggle and multiple item selection (similar to Shift or Ctrl keys).

Interaction design. The user simply touches the mobile screen (anywhere) to highlight currently viewed objects. When releasing the touchscreen, the currently highlighted item is selected. If the user does not want to select an item, he/she can simply look at a void spot or look away from the distant display and lift the finger from the touchscreen.

Discussion. An advantage of this pointing technique is that it is easy to adapt to common mouse-based interfaces. However, the pointing is imprecise as it does not take inherent eye tracking inaccuracies into account. As mentioned before, jittery gaze movements can be compensated by stabilizing the gaze cursor (e.g., [28]). However, the offset problem remains.

Gaze-supported Manual Selection

As presented by Zhai et al. [27], the idea of MAGIC pointing is to roughly position the cursor based on a user’s gaze position and then let the user make manual fine adjustments with the mouse. We adapt this concept to touch input and further extend it. For that, we first define a selection mask in whose proximity more precise selections can be performed using the mobile touch device (see Figure 3, left). As with the previously described technique gaze suggests the target, but here touch allows for fine adjustments before confirming the selection. Once the selection mask is activated (e.g., after a touch on the mobile device), the cursor does not follow the gaze anymore. For performing the manual fine selection, we contribute two variations: *MAGIC touch* and *MAGIC tab*, which are described in the following.

MAGIC touch. The cursor can be manually moved according to the touch position on the mobile screen. For this purpose, a representation of the selection mask is shown on the mobile screen (see Figure 3).

Interaction design. We propose a differentiation between *absolute* and *relative* positioning. This means, if the mobile screen is only briefly touched in the circular *touch area* (see Figure 2), the cursor will jump to the respective absolute po-

sition within the selection mask on the distant display. On the other hand, if the finger is dragged across the touchscreen, the cursor will move according to the relative movement from the initial touch position (see Figure 3, middle). This aims at supporting the user in keeping the view on the distant screen instead of switching his/her attention to the mobile device for manual cursor positioning.

Confirming a selection. The user can activate the selection mask by touching the mobile screen. As illustrated in Figure 2, the selection mask can be deactivated without performing a selection by touching the *no selection* area at the top of the mobile screen. Analogous, a selection can be confirmed by touching the *selection* area at the bottom of the mobile screen. Furthermore, large targets can be *directly* selected without the need to activate the selection mask first. This is achieved by looking at the target and touching the *selection* area immediately.

MAGIC tab. The cursor remains at the center of the selection mask and does not move. Items intersecting the selection mask can be discretely cycled through by using, for example, a continuous sliding gesture on the touchscreen. Another interaction option includes tilting the device to the left or right to browse through the collection of intersected items. Thus, *MAGIC tab* is similar to using the *tab button* on a keyboard.

Interaction design. The closest item to the user’s point-of-regard when activating the selection mask is automatically highlighted (see Figure 3, right). For going through the other objects intersecting the selection mask, we suggest using a *horizontal slide gesture* (to the left or right). If the user performs such a gesture and does not release the touch, the list is further passed through. Confirmation of an item’s selection is again done via touch on a virtual button as described for MAGIC touch (subsection *Confirming a selection*). In addition, we propose a *vertical slide gesture* to alter the size of the selection mask from very small (1 px) to a maximum value (100 px). This helps in confining the number of selectable items. Concerning the order of highlighted items, the sliding direction could indicate a clockwise or counterclockwise selection. Alternatively, items could be cycled through according to their distances to the mask’s center.

Discussion. With the MAGIC techniques gaze is no longer required after activating the selection mask and control is entirely handed over to manual fine selection. An advantage of the MAGIC techniques is that, despite inaccurate

gaze tracking, small and closely positioned targets can be selected. *MAGIC touch* allows for fine tuning the cursor position with the finger. *MAGIC tab* is decoupled from the size and distance of targets, because candidate objects are discretely highlighted one after each other. This may have a disadvantage for the selection of an item from a larger group of close targets as more objects need to be cycled to reach the desired target. The possible manual change of the selection mask’s size alleviates this problem. In addition, the order in which the items are stored in the list may not be clear to the user at all times. We propose to sort them according to the distance to the center of the selection mask. While this has the advantage that items close to the current cursor position are highlighted first, it may have the disadvantage that items with a similar distance are positioned at opposite sides and that their highlighting order may confuse the user.

Gaze-supported Expanded Target Selection

A local zoom lens can be activated at the current gaze position to facilitate target selections. We refer to this approach as *gaze-supported expanded target selection*. The lens activation can be done via a manual command, e.g., by pressing a button, issuing a touch event, or performing a gesture. Within the magnified area the user can select items more accurately with his/her gaze. Inspired by the gaze-based fisheye lenses from Ashmore et al. [2], we contribute two variations, which are illustrated in Figure 4 and described in the following.

Eye-slaved zoom lens. The lens follows the user’s gaze. Thus, the cursor remains always at the center of the lens.

Interaction design. After activating the zoom lens by tapping on the touch device, the user can move the lens based on his/her gaze position. A target can be selected in the previously described way by touching the selection area at the bottom of the mobile screen (*touch confirms*). To decrease jittery movements because of the target expansions, the gaze cursor is further stabilized (i.e., by increased filtering). The magnification level can be altered using a *vertical slide gesture* on the mobile screen (i.e., moving the finger up results in a higher magnification).

Semi-fixed zoom lens. A zoom lens is activated at the user’s point-of-regard, and the cursor can be freely moved within the lens using eye gaze. The lens does not move itself until the user looks beyond its boundary. In this case, the lens is dragged towards the current gaze position.

Interaction design. Similar to the *eye-slaved zoom lens*, a *vertical slide gesture* can be used to change the magnification level. Furthermore, we suggest a rate-based control for moving the lens while looking outside its border: the further the distance between the gaze position and center of the lens, the faster the lens will move.

Discussion. The proposed zoom lenses have the advantage of improving the visibility of small targets. However, a local magnification may not necessarily improve pointing accuracy, if the cursor speed remains at the same level. This means that the cursor movement may become more jittery when further zoomed in. Thus, target expansions may facili-

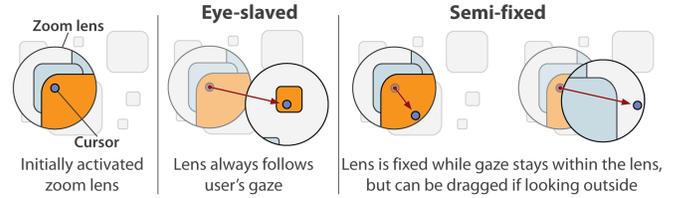


Figure 4. Initial situation (left) and variations of the gaze-supported expanded target selection - *eye-slaved* and *semi-fixed* zoom lens.

tate gaze-based target selections, but do not entirely overcome eye tracking inaccuracies (e.g., *offset* problems).

USER STUDY

We conducted a user study testing the five described gaze-supported selection techniques for different selection conditions. In particular, we were interested in the performance and suitability of each technique with regard to different target sizes and distances, which were both used as independent variables. Tasks ranged from selecting targets from an abstract grid of items to a more realistic desktop-like setting. This aims at better assessing the suitability of the developed techniques for different situations and how they can be further enhanced. Besides investigating the performance of each technique, we put a particular emphasis on qualitative user feedback for gaining better insights, identifying potential problems, and possible solutions for them.

Design. We used a within-subjects design with five interaction conditions:

C	Gaze-directed cursor
M_{tch}	MAGIC touch
M_{tab}	MAGIC tab
Z_{es}	Eye-slaved zoom lens
Z_{sf}	Semi-fixed zoom lens

We decided to test C, the two M and the two Z conditions together in a counterbalanced order to prevent the influence of order effects. This was done since M_{tch} and M_{tab} are very similar and belong to the same selection type (i.e., gaze-supported manual selection) and Z_{es} and Z_{sf} respectively (i.e., gaze-supported expanded target selection).

Participants. Twenty-four paid volunteers (14 male, 10 female) participated in the study, aged 22 to 31 (Mean (M) = 26.3) with normal or corrected-to-normal vision. In an initial questionnaire we asked participants about their background and to rate several statements on a 5-point Likert scale from 1 - *Do not agree at all* to 5 - *Completely agree*. Based on this, participants stated that they mainly use mouse and keyboard for computer interaction ($M=4.79$, Standard Deviation (SD)=0.50). While participants are interested in novel input devices ($M=4.33$, $SD=0.80$), many participants do not frequently use multitouch devices (such as smartphones) ($M=3.50$, $SD=1.58$). Finally, while all participants use computers on a daily basis, only eight had already used an eye tracker for interaction before.

Apparatus. For gathering gaze data we use a Tobii T60 table-mounted eye tracker: a binocular eye tracker is inte-

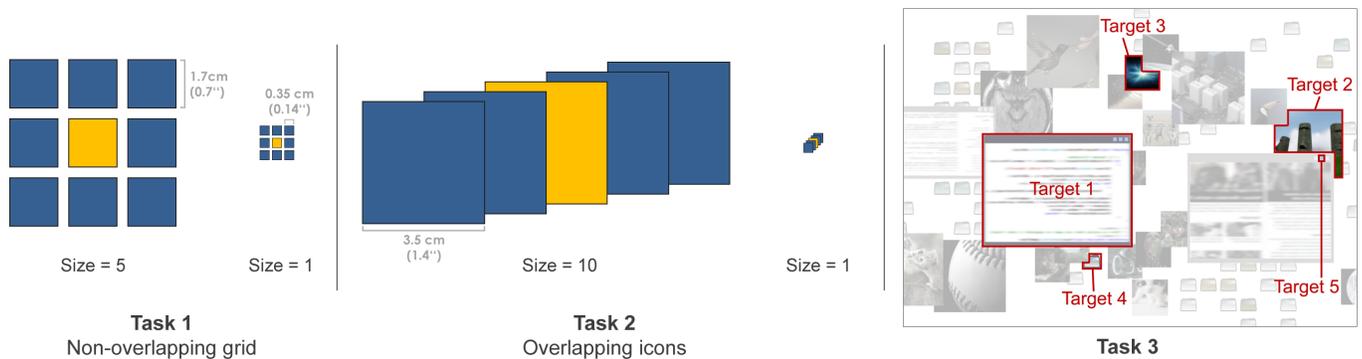


Figure 5. Schematic exemplary illustrations for the three task blocks T1, T2, and T3. Sizes and distances varied among the runs for T1 and T2. For T3, five targets had to be selected in the same order from a scene resembling a cluttered desktop interface.

grated in a 17-inch TFT flat panel monitor with a resolution of 1280x1024, a 0.5° accuracy, and sampling rate of 60 Hz. The gaze position is stabilized using the *speed reduction* technique [28]. Based on initial tests before the user study, we use a ratio of 8% of the current with 92% of the previous gaze position. The described gaze-supported selection techniques have all been implemented as suggested in the respective *Interaction design* paragraphs. We use a similar system setup for the gaze-supported multimodal interaction as proposed by [22]. We use Microsoft’s XNA Game Studio 3.0 (based on C#) for the creation of a test environment allowing for the selection of targets on the Tobii display (cf. Figure 5). An iPod Touch is used for the interaction on a mobile touchscreen. The GUI on the iPod is designed according to the screen prototype illustrated in Figure 2.

Procedure. The user study started with a brief introduction and an initial questionnaire about participants’ background (see subsection *Participants*). Participants were seated approximately 60 cm from the eye tracker display and were instructed to sit fairly still without restricting their movement. For each selection technique the same procedure was followed. First, a 9-point eye tracker calibration sequence was performed. Then, one selection technique at a time was described and the user could directly play around with it. The participant could test the technique until s/he felt sufficiently acquainted with it (usually less than 5 minutes). Three task blocks had to be completed in the same order with each selection technique. An overview is presented in Figure 5. The overall task was to select a single target from a set of given objects, whereby the alignment and size of objects and their distances differed among the task blocks:

- T1 Non-overlapping 2D items aligned in a grid
(3 sizes x 4 distances = 12 runs)
- T2 Overlapping 2D items aligned in a row
(3 sizes x 4 distances = 12 runs)
- T3 Desktop mockup: Overlapping items varying in size
(5 differently sized targets = 5 runs)

For task block T1 and T2, the object sizes and distances varied. The sizes differed from 10 (large) to 5 and 1 (small). The distances ranged from 1 (large distance) to 0.5, 0.25 and 0 (objects touch each other). The size and distance values are based on an internal virtual unit. Based on the Tobii T60’s

screen resolution, a size of 10 equals 3.5 cm (1.4”). Thus, assuming a distance of 60 cm to the eye tracker screen, the visual angle of targets ranged between 3.3° (size 10) to 0.3° (size 1). Item sizes and distances were differed across runs but not within the same run (see Figure 5). The same order was used for alternating target sizes and distances: First, all large targets (size=10) were tested with differing distances to distractors; then, this was repeated for the other target sizes as well. At the beginning of each run, participants needed to look at the center of the screen and touch the mobile device to confirm readiness. This was meant to improve the comparability between selection times. Targets had always the same distance to the screen center, however, they were positioned at alternating corners depending on the respective run.

For task block T3, a prototype was used that should resemble a desktop environment containing windows and icons. Participants had to select five targets as illustrated in Figure 5. The targets always had to be selected in the same order, starting with the largest (target 1) and finishing with the smallest and most difficult one (target 5).

Measures. Our quantitative measures included logged target acquisition times and error rates (an error is issuing a selection without the target being highlighted). Furthermore, as previously pointed out we aimed for substantial user feedback for a better assessment of the individual selection techniques. An *intermediate questionnaire* was handed out after T1-T3 have been completed with a single selection technique. The *intermediate questionnaires* consisted of three types of questions, for which all quantitative questions were based on 5-point Likert scales from 1 - *Do not agree at all* to 5 - *Completely agree*:

- Q1 Six general usability questions (see Figure 8) that were the same for all techniques
- Q2 Several questions concerning the particular selection technique - the number of questions differed from 5 to 7 depending on the respective technique
- Q3 Two questions asking for qualitative feedback on what the users particularly liked and disliked about the tested selection technique

In the *final questionnaire*, participants had to rate how they liked each selection technique, answer some concluding

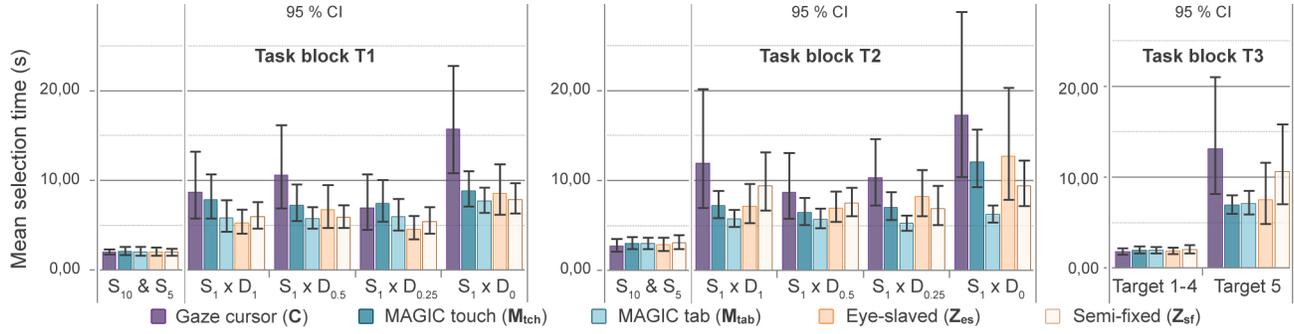


Figure 6. Overview of mean target acquisition times for the three task blocks. Results are summarized for S_{10} and S_5 (no significant differences).

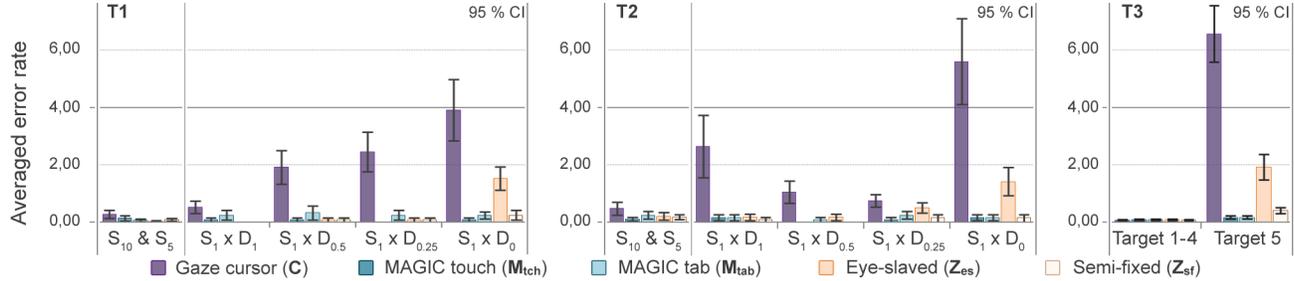


Figure 7. Overview of the averaged error rates for the three task blocks. Results are summarized for S_{10} and S_5 (no significant differences).

questions on the potential of gaze-supported selections, and give concluding qualitative feedback. On average, each session took about 120 minutes with instructions, carrying out the described procedure, and completing the questionnaires.

RESULTS

For the evaluation, we were particularly interested in qualitative feedback about how the developed gaze-supported selection techniques were experienced by the participants. The quantitative measures such as *target acquisition times* and *error rates* give a good indication about the practicality of these techniques. A repeated-measures ANOVA (Greenhouse-Geisser corrected) was used to investigate differences in task completion times and usability rankings. Post-hoc paired samples t-tests with a Bonferroni correction were used to further investigate which particular values differed significantly (pairwise comparisons).

Target Acquisition Times and Error Rates

Mean selection times and error rates are shown in Figure 6 and 7 with error bars indicating 95% confidence intervals (CI). They will be further discussed in the following based on the independent variables *distance* (D_1 , $D_{0.5}$, $D_{0.25}$, D_0), *target size* (S_{10} , S_5 , S_1), and *selection technique*. The influence of D and S on the mean selection times for T1 and T2 are summarized in Table 1.

For T1 and T2, different **target sizes** (S) had a *highly significant* impact on the selection times for all techniques (cf. Table 1). Selection times were significantly worse for S_1 than for S_5 and S_{10} for each selection technique. Large targets (S_5 and S_{10}) on the other hand could be quickly selected with all five techniques with no significant performance differences. Therefore, we accumulated values in Figure 6 and 7 for S_5

	T1	T2
C	$F_D(3,216)=4.50$, $p=0.004$ $F_S(2,72)=41.21$, $p<0.001$	$F_D(3,180)=3.40$, $p=0.036$ $F_S(2,60)=23.92$, $p<0.001$
M_{tch}	$F_D(3,216)=0.27$, $p<0.81$ $F_S(2,72)=31.33$, $p<0.001$	$F_D(3,216)=7.11$, $p=0.001$ $F_S(2,72)=52.77$, $p<0.001$
M_{tab}	$F_D(3,216)=0.81$, $p=0.40$ $F_S(2,72)=20.12$, $p<0.001$	$F_D(3,216)=5.14$, $p=0.003$ $F_S(2,72)=36.19$, $p<0.001$
Z_{es}	$F_D(3,216)=6.13$, $p=0.001$ $F_S(2,72)=29.02$, $p<0.001$	$F_D(3,213)=5.04$, $p=0.013$ $F_S(2,71)=21.59$, $p<0.001$
Z_{sf}	$F_D(3,216)=0.84$, $p=0.47$ $F_S(2,72)=40.43$, $p<0.001$	$F_D(3,210)=3.80$, $p=0.04$ $F_S(2,70)=13.25$, $p<0.001$

Table 1. Overview of the influence of target distances (D) and sizes (S) on the mean selection times. Significant results are printed in bold.

and S_{10} . Selection of small targets using **C** was very difficult and partly not realizable. In fact, on average only 81.5% of the participants managed to select a target of S_1 . As a result, selection times for **C** were significantly worse than for all other techniques for S_1 , except in comparison to **Z_{es}** for T2. In addition, **M_{tab}** performed significantly faster than the other techniques for S_1 in T2. A dependence between increased **error rates** and decreased target sizes exists for **C** and in an alleviated way for **Z_{es}** (see Figure 7). The error rates for **M_{tch}** , **M_{tab}** , and **Z_{sf}** do not significantly differ with respect to S .

While the mean selection times and error rates were not significantly influenced by varied **distances** (D) for S_{10} and S_5 , D had a significant effect on the task completion time for the smallest target size S_1 both for T1 ($F(4,384)=20.95$, $p<0.001$) and T2 ($F(4,288)=12.33$, $p<0.001$). This was es-

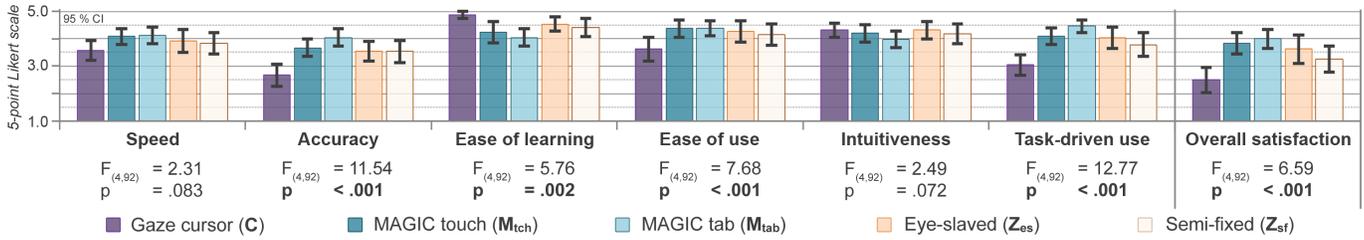


Figure 8. User-ranked gaze-supported selection techniques based on six usability factors and the overall satisfaction.

pecially the case for C and Z_{es} (cf. Table 1). As shown in Figure 6, for S₁ mean selection times did not significantly vary between different distances for M_{tch}, M_{tab}, and Z_{sf}. **Error rates** significantly increased for C with a decreasing distance for very small target sizes and in an alleviated way for Z_{es}. The error rates for M_{tch}, M_{tab}, and Z_{sf} remained relatively stable for varied distances.

For task T3, selection techniques significantly differed for the five targets (see Figure 6 and 7, right) with respect to target acquisition time ($F(4,440)=3.32$, $p=0.04$) and error rate ($F(4,440)=9.27$, $p<0.001$). Selecting target 5 was considerably harder than the other targets: a significant increase in target acquisition time occurred for all selection techniques. However, the error rates for M_{tch}, M_{tab}, and Z_{sf} remained fairly constant, while there was a significant increase for selecting the smallest target (5) with C, $t(108)=3.56$, $p<0.05$ (cf. Figure 6). In addition, selecting target 5 with C was significantly slower than with M_{tch} ($t(107)=3.13$, $p=0.02$) and M_{tab} ($t(107)=3.15$, $p=0.02$).

Quantitative User Feedback

For each selection technique, users rated how satisfied they were with different usability aspects (Q1 in the intermediate questionnaires), such as how *intuitive* the technique felt and whether tasks could be achieved as anticipated (*task-driven use*). The results are summarized in Figure 8. C received significantly lower ratings for task-driven use, accuracy, and ease of use which reflects the low task performance of C (high selection times and high error rates). On the upside, users assessed C significantly easier to learn. No significant differences occurred among M_{tch}, M_{tab}, Z_{es}, and Z_{sf}. Finally, Figure 8 (on the right) presents how participants rated the individual selection techniques in the final questionnaire after having tested each of them. C was assessed significantly worse than the other gaze-supported selection techniques. Although the other techniques were rated very similar, MAGIC tab received the best overall rating.

Qualitative User Feedback & Discussion

In the following, we discuss the gaze-supported selection techniques based on the user feedback to Q2 and Q3 (intermediate questionnaires) and the previously presented results. In general, the combination of gaze- and touch-based input for the interaction with a distant display was found very useful ($M=4.63$, $SD=0.56$). As one participant summarized well: *A prolonged working with the eyes will be wearisome, if the gaze input demands high precision and concentration from the user. The MAGIC and zoom lens techniques ease this and*

make gaze input more comfortable. In this regard, precise and fast selection of small targets was particularly wearisome for C, Z_{es}, and Z_{sf}, as participants sometimes had to look at a location near or beside a target to actually select it. This is reflected by high selection times and error rates especially for C, which partly prevented users from selecting a small target at all. Participants preferred the MAGIC techniques ($M=4.58$, $SD=0.57$), for which a rough gaze cursor position is sufficient and manual fine adjustments can be made. Seven participants mentioned that this gave them more control.

Mobile touchscreen. Although participants liked the virtual buttons on the touchscreen to confirm a selection ($M=3.76$, $SD=0.90$), this was sometimes error-prone. Especially for M_{tch} some users accidentally hit the selection button when trying to manually position the cursor. Instead, a selection could, for example, be performed with a double tap, and the selection mask could be deactivated by looking away from it for a specific amount of time.

Gaze-directed cursor. Although the gaze-directed cursor C was not suitable for the selection of small targets (high selection times and error rates), 19 participants emphasized its high potential for very easy, fast, and intuitive selections of large objects (also see Figures 6 and 7). Participants appreciated that items are highlighted as long as touching the display ($M=4.38$, $SD=0.63$).

MAGIC touch and tab. In general, the idea to manually position the cursor precisely via touch for M_{tch} and M_{tab} was found very good ($M=4.06$, $SD=0.85$), because it is more relaxing for the eyes (mentioned by 8 participants). In this respect, the MAGIC techniques were assessed very positively with respect to several usability aspects that are listed in Figure 8. The high grade for task-driven use can be explained by consistently low error rates for M_{tch} and M_{tab}. Furthermore, the selection mask was found very helpful to confine the cursor movement or the amount of selectable targets respectively ($M=4.08$, $SD=0.93$).

MAGIC touch. The current implementation of the relative ($M=3.63$, $SD=1.18$) and absolute cursor positioning ($M=3.29$, $SD=1.37$) was not particularly liked by participants. The relative cursor movement was sometimes too fast, which led to overshooting problems. This was especially a problem for very small targets, which 14 participants described as tedious. The absolute positioning was regarded as most suitable for quickly selecting large targets.

MAGIC tab. Participants liked the discrete selection of targets very much ($M=4.42$, $SD=0.70$), as it was fast and precise. Particular advantages of M_{tab} include its robustness against inaccurate gaze data, its suitability to select small targets, and that it is more relaxing for the eyes. However, the task performance is affected by the amount of targets intersecting the selection mask. Hence, it is important that users can easily narrow down the amount of selectable items, as this was criticized by participants. The size of the selection mask changed too slowly and was therefore considered as less useful (5 participants). In general, the slide gestures were described as slow and imprecise (9 participants), because they were sometimes not immediately recognized due to the current implementation. In addition, five participants complained that the tab order was unintuitive. It was not clear where the target was in the item collection and whether going forward or backward would be faster to reach it. Instead of using slide gestures, backward and forward buttons may offer more control. It was also suggested that the selection could be based on the gesture's direction to perform more systematic selections, for example, by flicking the cursor towards a desired target and snapping the cursor to it. However, this may cause problems if objects are positioned behind or close to each other. Another idea is to automatically adapt the initial size of the selection mask depending on the current context, for example, based on the number, size, and distance of targets.

Gaze-supported zoom lenses. The combination of an enlarged view and a slowed movement of the cursor/lens are considered helpful for selecting small targets more precisely. However, they do not overcome eye tracking inaccuracies completely, which becomes apparent for small and closely positioned targets (see Figure 6). In addition, participants liked the vertical slide gesture to adjust the zoom level ($M=3.80$, $SD=1.01$). However, especially for the zoom lens techniques it showed that users desired more possibilities for *customization*, for example, of the lens size, the maximum zoom level, and the lens speed. This is reflected in highly diverging ratings in the questionnaires. For example, the decreased speed, although preferred by several users, received an overall moderate rating ($M=3.46$, $SD=1.25$). In this respect, five participants explained that they found the *lens speed* too slow and three *not slow enough*.

Eye-slaved. While Z_{es} was on the one hand described as fast and intuitive (by three users), it was also characterized as distracting, laborious, and imprecise, because it does not take eye tracking inaccuracies into account. Participants liked that the cursor was always at the lens's center ($M=3.96$, $SD=1.06$) and that the lens immediately moved according to the gaze ($M=4.04$, $SD=1.21$).

Semi-fixed. Z_{sf} was found less irritating than Z_{es} (four participants), however, it was considered imprecise (8 participants). It was positively mentioned that Z_{sf} provides more stability ($M=3.58$, $SD=0.95$) while still being flexible to be panned if the user looks outside the lens ($M=3.79$, $SD=1.19$). However, seven users particularly disliked the way the Z_{sf} lens moved, because users had to look away from the lens to move it and look back to see if the desired target has been

reached. Thus, it was suggested that the lens should already move when looking at the lens's border from within so that there is no need to leave the lens with the gaze.

Combination of techniques. All in all, each selection technique has its unique advantages as described above. Even though C achieved the lowest performance and overall user rating, it was assessed very positively for the selection of large targets. In this respect, a seamless integration of the proposed selection techniques into one complex interaction set seems highly promising to cater for diverse selection contexts. While large targets can be selected with a simple gaze-directed cursor (C), smaller items could be quickly selected with a *gaze-supported zoom lens* and/or manual selection (*MAGIC*) technique. Providing more customization features for the zoom lenses, a user could dynamically adjust how fast the lens follows the gaze and with that a smooth transition between Z_{es} and Z_{sf} could be achieved. The lenses could then be advanced by allowing the user to move it via touch or select items within the lens using M_{tch} or M_{tab} .

In summary, the combination of gaze- and touch-based input for the interaction with a distant display was perceived as very promising. However, since this study was restricted to a desktop setting, the techniques need to be further investigated and adapted to distant large displays.

CONCLUSION

In this paper, we contributed several novel and practical ways for combining gaze and touch input from a mobile touchscreen for selection tasks on a distant display. According to the principle *gaze suggests* and *touch confirms*, we described five gaze-supported selection techniques ranging from a simple touch-enhanced *gaze-directed cursor* to novel *gaze-supported local zoom lenses* and our *MAGIC touch* and *MAGIC tab* techniques. While *MAGIC touch* allows for manually positioning the cursor via touch input, *MAGIC tab* allows for discretely going through a list of objects that are spatially close to the user's point-of-regard via a slide gesture. These techniques were evaluated in a user study providing important insights for further improvements. With the presented techniques we could overcome prevalent problems associated with gaze-based interaction, such as the *Midas touch* problem. Especially with our *MAGIC* techniques we could compensate eye tracking inaccuracies well. In this respect, the *MAGIC* techniques excel with their robustness against inaccurate gaze data, a high comfort as they are less straining for the eyes, and a high overall performance and perceived usability. In particular *MAGIC tab* shows a high potential for further investigations as it is suitable for quick and reliable selections of small, closely positioned and overlapping targets.

For future work, we will enhance the proposed gaze-supported techniques based on the collected user feedback and investigate possibilities to seamlessly combine them to take advantage of their particular strengths for different conditions. Furthermore, the revised techniques will be tested with large distant displays and in comparison with traditional input devices. Finally, the proposed selection techniques may also benefit target acquisition within virtual 3D environments, in which several objects may be positioned behind each other.

Therefore, as part of our future work we will investigate gaze-supported interaction in 3D environments.

ACKNOWLEDGMENTS

This research is supported by the German National Merit Foundation and the German Ministry of Education and Science (BMBF) project ViERforES-II (01IM10002B).

REFERENCES

1. Adams, N., Witkowski, M., and Spence, R. The inspection of very large images by eye-gaze control. In *Proc. of AVI'08* (2008), 111–118.
2. Ashmore, M., Duchowski, A. T., and Shoemaker, G. Efficient eye pointing with a fisheye lens. In *Proc. of Graphics Interface 2005, GI '05* (2005), 203–210.
3. Bates, R., and Istance, H. Zooming interfaces!: enhancing the performance of eye controlled pointing devices. In *Proc. of Assets'02*, ACM (2002), 119–126.
4. Bieg, H.-J., Chuang, L. L., Fleming, R. W., Reiterer, H., and Bülthoff, H. H. Eye and pointer coordination in search and selection tasks. In *Proc. of ETRA'10*, ACM (2010), 89–92.
5. Castellina, E., and Corno, F. Multimodal gaze interaction in 3D virtual environments. In *COGAIN '08* (2008), 33–37.
6. Cho, S.-J., Choi, C., Sung, Y., Lee, K., Kim, Y.-B., and Murray-Smith, R. Dynamics of tilt-based browsing on mobile devices. In *Proc. of CHI'07*, ACM (2007), 1947–1952.
7. Drewes, H., and Schmidt, A. The MAGIC touch: Combining MAGIC-pointing with a touch-sensitive mouse. In *Proc. of INTERACT'09*, Springer-Verlag (2009), 415–428.
8. Fono, D., and Vertegaal, R. EyeWindows: evaluation of eye-controlled zooming windows for focus selection. In *Proc. of CHI'05* (2005), 151–160.
9. Hansen, D. W., Skovsgaard, H. H. T., Hansen, J. P., and Møllenbach, E. Noise tolerant selection by gaze-controlled pan and zoom in 3D. In *Proc. of ETRA'08*, ACM (2008), 205–212.
10. Hansen, J. P., Johansen, A. S., Hansen, D. W., Ito, K., and Mashino, S. Command without a click: Dwell time typing by mouse and gaze selections. In *Proc. of INTERACT'03*, IOS Press (2003), 121–128.
11. Jacob, R. J. K. What you look at is what you get: eye movement-based interaction techniques. In *Proc. of CHI'90*, ACM (1990), 11–18.
12. Kabbash, P., and Buxton, W. A. S. The prince technique: Fitts' law and selection using area cursors. In *Proc. of CHI'95*, ACM Press/Addison-Wesley Publishing Co. (1995), 273–279.
13. Kumar, M., Paepcke, A., and Winograd, T. EyePoint: practical pointing and selection using gaze and keyboard. In *Proc. of CHI'07*, ACM (2007), 421–430.
14. Lankford, C. Effective eye-gaze input into windows. In *Proc. of ETRA'00*, ACM (2000), 23–27.
15. Miniotos, D. Application of Fitts' law to eye gaze interaction. In *Proc. of CHI EA'00*, ACM (2000), 339–340.
16. Miniotos, D., Špakov, O., and MacKenzie, I. S. Eye gaze interaction with expanding targets. In *Proc. of CHI EA'04*, ACM (2004), 1255–1258.
17. Monden, A., Matsumoto, K., and Yamato, M. Evaluation of gaze-added target selection methods suitable for general GUIs. *Int. J. Comput. Appl. Technol.* 24 (June 2005), 17–24.
18. Salvucci, D. D., and Anderson, J. R. Intelligent gaze-added interfaces. In *Proc. of CHI'00*, ACM (2000), 273–280.
19. San Agustin, J., Hansen, J. P., and Tall, M. Gaze-based interaction with public displays using off-the-shelf components. In *Proc. of Ubicomp'10*, ACM (2010), 377–378.
20. Skovsgaard, H., Mateo, J. C., Flach, J. M., and Hansen, J. P. Small-target selection with gaze alone. In *Proc. of ETRA'10*, ACM (2010), 145–148.
21. Smith, B. A., Ho, J., Ark, W., and Zhai, S. Hand eye coordination patterns in target selection. In *Proc. of ETRA'00*, ACM (2000), 117–122.
22. Stellmach, S., Stober, S., Nürnberger, A., and Dachselt, R. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. of NGCA'11*, ACM (2011), 1–8.
23. Špakov, O. Comparison of gaze-to-objects mapping algorithms. In *Proc. of NGCA'11*, ACM (2011), 1–8.
24. Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. In *Proc. of CHI'87*, ACM (1987), 183–188.
25. Yamato, M., Inoue, K., Monden, A., Torii, K., and Matsumoto, K.-i. Button selection for general GUIs using eye and hand together. In *Proc. of AVI'00*, ACM (2000), 270–273.
26. Yoo, B., Han, J.-J., Choi, C., Yi, K., Suh, S., Park, D., and Kim, C. 3D user interface combining gaze and hand gestures for large-scale display. In *Proc. of CHI EA'10*, ACM (2010), 3709–3714.
27. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. In *Proc. of CHI'99*, ACM (1999), 246–253.
28. Zhang, X., Ren, X., and Zha, H. Improving eye cursor's stability for eye pointing tasks. In *Proc. of CHI'08*, ACM (2008), 525–534.