Pointing while Looking Elsewhere: Designing for Varying Degrees of Visual Guidance during Manual Input

Barış Serim[1,3], **Giulio Jacucci**[1,2]
[1] Helsinki Institute for Information Technology HIIT, Dept. of Computer Science, University of Helsinki
[2] Helsinki Institute for Information Technology HIIT, Aalto University
[3] Department of Design, Aalto University
firstname.lastname@hiit.fi

ABSTRACT

We propose using eye tracking to support interface use with decreased reliance on visual guidance. While the design of most graphical user interfaces take visual guidance during manual input for granted, eye tracking allows distinguishing between the cases when the manual input is conducted with or without guidance. We conceptualize the latter cases as input with uncertainty that require separate handling. We describe the design space of input handling by utilizing input resources available to the system, possible actions the system can realize and various feedback techniques for informing the user. We demonstrate the particular action mechanisms and feedback techniques through three applications we developed for touch interaction on a large screen. We conducted a two stage study of positional accuracy during target acquisition with varying visual guidance, to determine the selection range around a touch point due to positional uncertainty. We also conducted a qualitative evaluation of example applications with participants to identify perceived utility and hand eye coordination challenges while using interfaces with decreased visual guidance.

Author Keywords

Gaze input; eye tracking; multimodal interaction; uncertain input; interaction techniques; interactive surface

ACM Classification Keywords

H.5.2. Information interfaces and presentation: Input devices and strategies

INTRODUCTION

In HCI, terms such as eyes-on or eyes-free input are used to describe the degree of visual guidance an input action is performed with, in other words, the extent sight is used to guide action. Input actions vary regarding their degree of visual guidance. While typing on a physical keyboard can be conducted with little visual guidance, selecting items from a graphical interface often demands users to look where they are pointing to. Visual guidance of input actions gains particular importance with the use of eye tracking as a real time input for interaction. Examples of gaze input often feature gaze as a pointer for selection [30, 33, 41], assuming user's visual focus at the region of interest [3, 21]. Human visual attention, however, is a limited resource and there are a number of reasons to support input without extensive reliance on visual guidance:

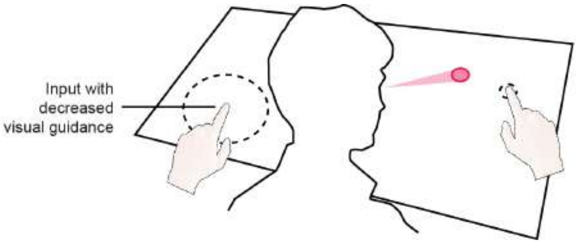


Figure 1. The user's manual input can be handled based on the degree of visual guidance it is conducted with. The input position is interpreted as exact when the action is realized eyes-on, while the system increases the potential selection range around touch and utilizes contextual resources and feedback techniques for input handling in the case of decreased visual guidance.

- It can be desirable or necessary to remain visually focused at a certain region of interest without having to redirect gaze to another region in the interface.
- Input accuracy can be uncritical for certain cases, when the user is casual or wishes to delegate a certain level of control to the system.
- The task can require concurrent pointing at multiple regions of interest within the interface.

In this paper, we propose *using eye tracking to support manual input in the absence of or with little visual guidance*. The design of most graphical interfaces takes user's full visual guidance during manual input for granted. System interpretation of input is accordingly definitive; pointing actions on the interface are processed as exact coordinates. As an alternative to the current adoption, eye tracking can be used to understand the degree of visual guidance that a manual action is accomplished with and adapt the system interpretation and handling of the user input. Our main design strategy is to use manual input as a direct input and utilize gaze to increase

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

ACM 978-1-4503-3362-7/16/0515.00
DOI: http://dx.doi.org/10.1145/2858036.2858480

its expressiveness. This approach is fundamentally different than many current examples of gaze interaction that complement gaze with indirect manual input [21, 30, 33].

We make a number of contributions to support manual input with varying visual guidance. We conceptualize user input performed with decreased visual guidance as input with uncertainty and adopt an uncertain input handling framework for adapting system behavior. We describe action mechanisms and novel feedback techniques for handling such input and demonstrate their use through three example applications. We conducted a two part study to guide future design. In the first part, we determine the selection range of manual input on a large touch screen by measuring the positional offset with different degrees of visual guidance. In the second part, we report the perceived utility and the hand-eye coordination challenges that emerge during the interaction through a qualitative evaluation of the applications.

BACKGROUND

We motivate our design approach by discussing earlier work on gaze input, hand-eye coordination and input with uncertainty.

Gaze as Input Modality

Our work falls under the design approach that utilizes gaze as an additional modality rather than replacing manual input. Research in this direction aims to compensate the lack of a confirmation mechanism in gaze input (known as "Midas Touch" [16]) by using physical keyboard [19], mouse [41], touch [30] or gesture.

Previous studies conducted in controlled, isolated settings show that gaze can be faster than other pointing devices for target selection [28, 34]. Thus, a strong motivation for most previous work that combine gaze and manual input has been motor performance gains in target acquisition [4, 18, 21, 30, 31, 33, 41]. A pioneering application is Zhai et al.'s MAGIC pointing [41], a manual and gaze hybrid pointing method, that eliminates part of the mouse movement by warping the mouse cursor to the eye fixation coordinates and then accomplishes the selection action through the mouse, thus cascading the two input modalities. Recently, interaction with large and distant displays, where direct input is impractical, has been an application area for utilizing gaze. In these applications gaze is complemented by touch input on a hand-held device [30, 31, 33] or free air gestures [18].

In general, previous work capitalizes on the rapid switching of spatial context afforded by gaze to decrease the amplitude of movement by hand. Thus, a common feature among them is the separation of the hand from the target, namely the indirect and relative use of manual input to complement the absolute coordinates provided by gaze. Two hybrid exceptions are GazeTouch [21] and Gaze-Shifting [22] that utilize manual input both as a direct and indirect input, based on the distance of gaze point to the input position.

Some of the examples cited above are similar to our approach in that they facilitate manual input without visual guidance. This is achieved through different means, such as using touch as a relative, indirect input [21] or in small handheld devices that enable eyes-free interaction [31]. Our approach departs from them by *always using manual input directly*, even in the case of input without visual guidance. We use manual input for selection and use gaze input to qualify manual input. While earlier work advocates the separation of the hand from the target, summarized as *"gaze suggests and touch confirms"* [30] or *"gaze selects, touch manipulates"* [21], we propose an alternative use in which *"gaze qualifies hand input."*

Hand-Eye Coordination

Our approach is partly motivated by the simultaneous use of gaze and manual input on multiple points of interest. Previous work in eye cursor coordination in web search shows that mouse use is not purely *incidental*, (i.e. performed for the purpose of clicking) [24]. Instead, the cursor can be used for other purposes such as keeping track of what is read and as a placeholder on interesting items, while eyes switch to other regions [15, 24]. Additionally, as Bieg et al. [3] argue, one assumption in techniques that aim to decrease the amplitude in target acquisition using gaze is that eye movement precedes pointing actions. Contrary to this, their study reports that pointing behavior is initiated without visual guidance for items whose approximate locations are known.

In the above described situations gaze and pointing accomplish parallel tasks in different regions within interface. However, such parallel use of eye and cursor movements might not be well supported by design approaches that cascade (i.e., sequence) manual and gaze input such as MAGIC pointing [41].

While approaches like MAGIC pointing focus on increasing the performance in a sequential set of actions using gaze, we target supporting concurrent access to multiple regions on the interface, without necessarily redirecting gaze. A usable distinction has been made by Fitzmaurice et al. [9, 10] between spatial and time multiplexing for user input. While time multiplexing refers to sequential and mutually exclusive techniques, spatial multiplexing refers to the concurrent access to dedicated input fields. Their observation of manual interaction with domino bricks is illustrative of spatial multiplexing: *"...Tactile feedback was often used to grab dominos while visually attending to other tasks. The non-dominant hand was often used to reposition and align the dominos into their final resting place while, in parallel, the dominant hand was used to retrieve new dominos..."* [10]. Fitzmaurice et al., accordingly, design for spatial multiplexing through graspable input devices by citing the benefits of tactile confirmation and possible use without visual guidance.

On the other hand, the advent of multi-touch devices enabled spatial multiplexing in graphical interfaces. Even though interaction with tactile interfaces has shown to be more robust and efficient [32, 37], multi-touch input surfaces allow similar benefits like bimanualism. Previous work on touch screens aims to support eyes-free interaction in various ways such as using touch as a gestural input or directing finger to predefined locations using magnetic attraction [36]. In contrast, we support eyes-free interaction through appropriate interpretation and handling of the user input. We interpret lack of visual guidance as situations of inputs with uncertainty.

Input with Uncertainty

Proliferation of inherently uncertain inputs, such as speech recognition, gestures and touch, motivated a number of techniques and frameworks for the flexible handling of user input and communicating system interpretation of input back to the user [20, 23, 25, 26, 38, 39].

A large body of research on input positional uncertainty deals with the "fat finger problem", namely the large touch contact area and visual occlusion caused by the finger [2, 26, 35, 38]. In the context of this paper, the source of uncertainty is the user's lack of exact information about how his/her manual input coordinates map to the visual content on the user interface *due to decreased visual guidance*. The most closely related work in this direction is by Hagiya and Kato [14], who use gaze point information to model touch distribution on a hand-size mobile display. Although particularly focused on text entry, their distinction between accurate and ambiguous touch is parallel to our approach. Different from their work, we consider the overall design space of input handling and demonstrate their use in diverse applications with multi-touch interaction on a large screen. Previous work on the accuracy of target acquisition using arm movements without visual guidance suggests that errors increase in relation to the amplitude of movement [5, 29] and cumulatively [6].

Users' lack of information about their exact input region allows system to interpret the input as positionally ambiguous and less decisive. Conversely, high visual guidance reinforces user's manual input. In HCI various design frameworks aim to adapt system behaviour depending on user's varying degrees of control. In vehicle design, "horse metaphor" [12] refers to a level of delegation of decision making to the system, based on how tight or loose the user's control is. Pohl and Murray-Smith propose design approaches for mobile systems that allow users to vary their level of engagement along a focused-casual continuum [23]. When user input is casual (i.e., lacking in precision and deliberation) the system partly takes control using available personal and contextual information. In the same spirit, we use the degree of visual guidance for a partial delegation of decision making to the system. However, while user attention is inhibited or reserved for another activity in mobile use [23], we are primarily interested in the cases in which user attention is divided between two actions related to the same task and two regions within the same interface. This enables using gaze position on the interface as a resource for interpreting the user's input and providing feedback to the user through various channels.

DESIGN SPACE

Previous work [20, 25] on handling input with uncertainty separates handling process into successive stages of modeling input, event dispatch, interpretation and action, in which the system component "mediator" is responsible for deciding on the action. We used a similar structure and provide an inexhaustive list of considerations and techniques that are particularly relevant for handling input with varying degrees of visual guidance.

User input involves both user manual input position and other contextual information. The system handles user input through various *action mechanisms* (i.e. select, defer or inaction). *Feedback techniques* aim to remedy users' lack of visual guidance by making manual input information and system interpretation of input available to the user.

User input

Manual input position. The primary resource for interpreting user's manual input is the position (such as x,y values) of the input. Potential selection range around a manual input position increases with decreasing visual guidance (Figure 2), due to positional uncertainty. We operationalized visual guidance as the distance between the gaze and manual input position and use it to compute the potential selection range around touch input. The interpretation of manual input is exact up to a certain distance threshold between gaze and input position. Beyond this threshold, the selection range increases in linear relation to the distance between gaze and manual input position. For touch input the selection range is greater than a single pixel, even when it is conducted with visual guidance, due to the inherent uncertainty of touch. During the design process we heuristically defined the threshold and linear relation values. A two stage target acquisition study described further in the paper shows how the threshold and linear relation can be empirically determined.

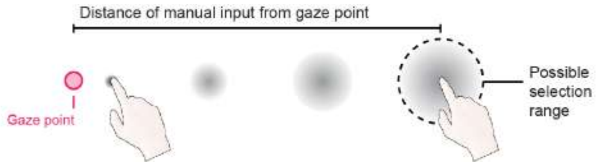


Figure 2. Input selection range increases the further gaze point is located from the manual input position (where the touch or cursor is located).

Additionally, the determination of visual guidance is dependent on a number of design decisions:

- **Continuous, discrete.** For a manual input event, visual guidance can be determined along a discrete (such as only covering the two opposite ends eyes-on and eyes-free) or continuous scale.
- **Conservative, liberal approaches.** The determination of visual guidance and thus the selection range around an input position can change upon eye movements (liberal approach) or only upon the movement of the hand (conservative approach). We borrow the terms from Zhai et al. [41], who used them to distinguish the cases in which a mouse cursor is continuously warped to the gaze point coordinates (liberal) or only upon a cursor movement event (conservative). For many cases, conservative approach can be more suitable, since elements pointed with visual guidance will persist in the user's short term memory even after the gaze shifts to another location.

  When scaling the range of uncertainty upon movement, our general principle is to a) decrease the uncertainty *instantly* when the user increases visual guidance and b) increase the uncertainty *gradually* when the user decreases the visual guidance. The difference is due to the gradual deviation in position with increasing amplitude of movement [5]. It should be noted that for cursor input (e.g., mouse), the uncertainty can always be determined upon movement as the cursor is always present at the interface. On the other hand, touch input involves finger enter and exit events.

Gaze context. Interpersonal interaction can involve "referential gaze" in combination with speech and manual pointing to ground and disambiguate meaning. The use of "dual pointing" to two different regions of interest in the environment, one with gaze the other with the hand, to semantically associate them, has been documented [13]. Similarly, eye tracking information can be used beyond determining the positional uncertainty of manual input. The information of where the user is gazing at the interface can be used to resolve uncertainty in manual input by prioritizing actions that are related to the gaze context.

There are multiple possibilities regarding how the gaze context can be determined in relation to manual input. First, the gaze context position can be determined at the beginning of manual interaction and remain fixed until the user ends the manual input (such as by releasing a mouse or in a touch up event). Second, the gaze context position can be constantly updated upon eye movements. However, during the design process we noticed that continuous synchronization of gaze context with eye movements can be intrusive and unstable. A more viable option is to update gaze context only upon touch or cursor movements.

Interaction history. Another resource for resolving positional uncertainty is the interaction history of the user. A possible reason for the lack of visual guidance and loose hand-eye coordination could be that the location of the item is remembered [11]. Thus, decreased visual guidance can be attributed to the user expectation of repeating a previous action. The system can also resolve uncertainty by keeping track of how user changes the application state. Application functionality makes certain action sequences more probable over others in an interface configuration. As an example, if the user previously opened a dropdown menu, it is more probable that a selection action on one of the menu items will follow.

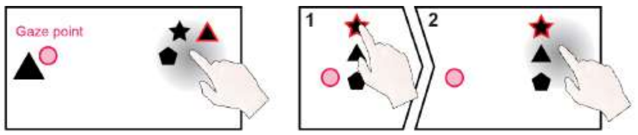


Figure 3. Two possible ways of resolving positional uncertainty are using gaze context (left) and interaction history (right).

Action Mechanisms

Select action. The system can respond to positional uncertainty in a number of ways for selecting action (Figure 4). One is positional selection between different actions, such as selecting between different discrete input fields like buttons. The selection can also occur between different actions that positionally overlap. For example, a touch action on a text field can be intended for scrolling or text selection [26]. These different actions require different degrees of visual guidance: scrolling has an area effect and does not require exact pointing, while selection requires accurate pointing. Decreased visual guidance in such cases can aid the system decision making between various actions types. Finally, if the input field allows range selection, positional uncertainty can be handled by expanding the selection range.

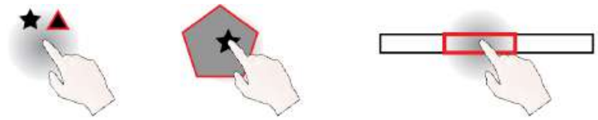


Figure 4. Action selection can involve positional selection (left), selecting between actions that positionally overlap (middle) or range selection (right).

Defer action. Another potential response is to defer action until enough information is gathered for disambiguation. A common example is the press-release sequence for inherently uncertain inputs such as touch [25] or gaze [19]. System interpretation of user command is communicated as a feedback after key or touch press event and the final action is deferred to a key or touch release event.

Inaction. Input without visual guidance can be interpreted as unintentional or unfocused, resulting in the system not taking action.

Feedback Techniques

In addition to possible non-visual notifications such as sound or tactile feedback, eye tracking allows various visual strategies to provide feedback about user actions and the system interpretation of them:

Support peripheral awareness. In the case of manual input without visual guidance the system can remedy the lack of information by supporting peripheral awareness. The system can increase the visual footprint of the cursor a) to support the peripheral awareness of where the cursor or finger is located within interface and b) to indicate the degree of positional uncertainty as determined by the system (Figure 5). The visual footprint of potential targets can also be increased, informing users about the system interpretation of their action.

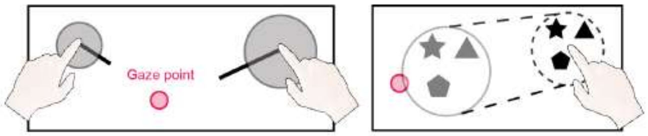


Figure 5. Providing peripheral awareness (left) and warping information content around manual input position to gaze point (right) are two possible visual feedback techniques to communicate system interpretation of user input back to user.

Warp information to gaze point. Warping information to the gaze point is the counterpart of warping cursor to the gaze point location (e.g. [41]). The information content around the user's manual input position or the system interpretation of user action is overlaid to where the user's gaze is directed (Figure 5).

Figure 6 provides a summary of the design space. It should be noted that providing feedback and uncertain input handling are two competing approaches, since feedback techniques decrease the uncertainty by providing information to the user. However, they are not mutually exclusive and can be integrated in various stages of interaction as can be seen in the applications described further in the paper.

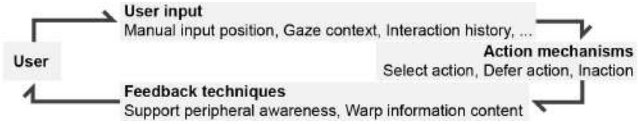


Figure 6. Summary of the design space.

DETERMINING SELECTION RANGE

An important design question related to positional uncertainty is determining selection range for input with varying levels of visual guidance.

We conducted a two stage study of target acquisition to determine the potential selection range around a touch point. 12 participants (4 female) aged between 20 to 34 (m=28, sd=3.95) took part in our evaluation. Each session started with 9 point eye tracking calibration and proceeded if the calibration was successful (below 2° deviation in accuracy). The height of the table on which the screen and eye tracker were mounted was adjusted for each participant and the participants remained standing during the evaluation.

Apparatus

The study has been conducted using a 10 finger multi-touch screen (27", 2,560x1,440 pixels) combined with an SMI RED eye tracker that is positioned below the touch screen running at 60Hz. The screen was tilted 30° to enable easier hand reach for touch input. The screen and eye tracker were positioned respectively 50cm and 70cm in front of the eyes (approximate values).

Study Design

Each participant performed two set of tasks. The first set of tasks aimed to determine positional inaccuracy (variable) for varying distances between gaze and target (invariable). The second task set aimed to determine the distance between gaze and the target (variable) for accurate pointing tasks (invariable).

We logged gaze, touch and target positions for each task. During the study there were brief moments when eye tracking signal was not available due to hand occlusion or head movement. Thus, a task was completed only when the gaze point was available to the system.

Position Inaccuracy for Gaze Distance

Earlier work suggests a decreased positional accuracy for motor target acquisition without visual guidance [5, 29]. At the first stage, we aimed to determine the positional uncertainty for target acquisition in varying degrees of visual guidance, which we operationalized as *the distance between the target and where the gaze is directed*.

To complete a target acquisition task, the participants had to keep their gaze (controlled by eye tracking) inside a circle while tapping on one of the 15 targets (on a 5 × 3 matrix) on the touch screen (Figure 7). The participants were instructed to tap as correctly and as fast as possible to determine the positional offset for acquisition. We used the 6 lower middle points within the matrix, where the eye tracking is most accurate, as gaze fixation points. While keeping their gaze within the defined area, the participants tapped on all the defined targets on the matrix at a randomized order. The target acquisition tasks were accepted only if the participants kept their gaze within the circle (indicated to the participants by changing the circle area to green).

To prevent giving any visual cues, the entire target matrix was visible during the tasks. However, information of which target to tap on was shown within the circle. The pairing of 6 eye fixation regions with 15 target positions resulted in 90 tasks, that show varying distances between the target and the gaze point. To prevent the misidentification of the target, each column was assigned a different shape (from left to right: circle, cross, triangle, square and pentagon).

The degree of visual guidance as we operationalized in the study is not easily comparable to index of difficulty in Fitts' law. First, the amplitude of motion is not dependent on the distance of the target from the gaze, since the participants initiated the movement from the previous target location in the matrix. Second, the visual boundary of the targets does not accurately represent target width as the system did not require the participants to touch on the exact position.

At the same time Fitts' law has implications for acquisition with restricted visual guidance. It has been shown that Fitts' law is valid for restricted visual guidance on the target or hand [40]. By increasing the selection range for manual input, we increase the target width and thus decrease the index of difficulty. How the increased target width compensates the lack of visual guidance for acquisition performance is a highly relevant question for future research. However, we limit the scope of this study to determining the selection range around touch point.

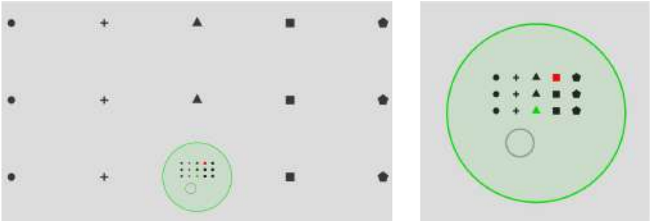


Figure 7. The experimental screen (left) and the close up of the circle (rad = 52.5mm) in which the participants need to keep their gaze inside (right). The target is shown in red inside the circle, while the position of the circle is shown in green.

Gaze Distance for Accurate Acquisition

While the first stage aimed to understand the positional inaccuracy with varying visual guidance, second stage aimed to understand the distance of gaze from the target for accurate pointing. Participants were instructed to touch circular targets, without any constraints on where they look. This stage forced participants to be accurate since a task was considered complete only when the touch point fell within the circular target (rad = 5.8mm). Each participant completed 3 repetitions of 6 target acquisition tasks (randomized order).

Results

The first stage yielded 1080 trials from 12 participants (× 90 tasks). The scatter plot in Figure 8 shows the relationship between the distance of the gaze point to the target position (invariable) to the positional offset (distance between the touch and target positions). The outliers in the scatter plot refer to the trials in which there was a large positional offset (M=120.8 mm, sd = 13.4), but the touch point was close to an adjacent target on the matrix (M=19 mm, sd = 8.8). We categorize these 31 outliers as cases in which participants misidentified the target, and exclude them from the analysis.

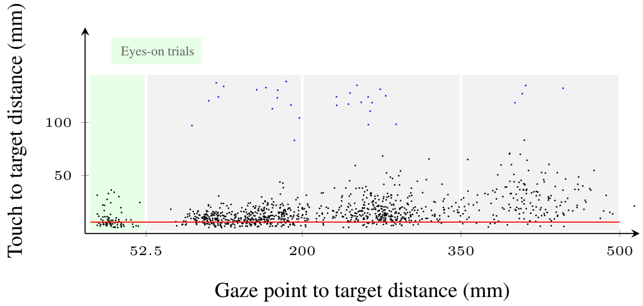


Figure 8. Scatter plot of peripheral target acquisition tasks across all participants. Outliers are shown in blue. Horizontal red line indicates the visual boundary of the circle target (rad=5.8mm). Green background indicates the eyes-on in which the target was within the boundary of the circle the participants had to keep their gaze inside(rad=52.5mm).

We divide the data in four continuous bins that correspond to varying levels of visual guidance. The intervals of the first bin (0-52.5) were determined by eyes-on tasks, in which the distance between the gaze point and the target are smaller than the radius of the circle in which the participants had to keep their gaze inside. We divided the rest of the data in three bins of even intervals. Figure 9 shows the distribution of touch points relative to the target across all users for four chosen levels of visual guidance. The deviation in the distance between gaze point and target is due to the large diameter of circle in which the participants need to keep their gaze within. For varying visual guidance levels, 95% confidence values for positional offset can be used to determine touch selection range. The results (Table 1) suggest an increasing positional inaccuracy with increasing distance between gaze point and target.

An unusual result from the first stage is the very large deviation for eyes-on tasks (0-52.5mm) when compared to a previous study [2] that reports an accuracy rate higher than 95% for 5mm radius target acquisition. We relate the unusual result to the experimental design of the first stage, in which participants did not have to touch within the target to complete the task. This is in contrast to the second stage, in which participants had to touch within the target visual border. In 216 total trials gathered from the second stage, 95% of gaze points were within a 61.0mm radius range around the target (M=26.3mm, sd=20.0).

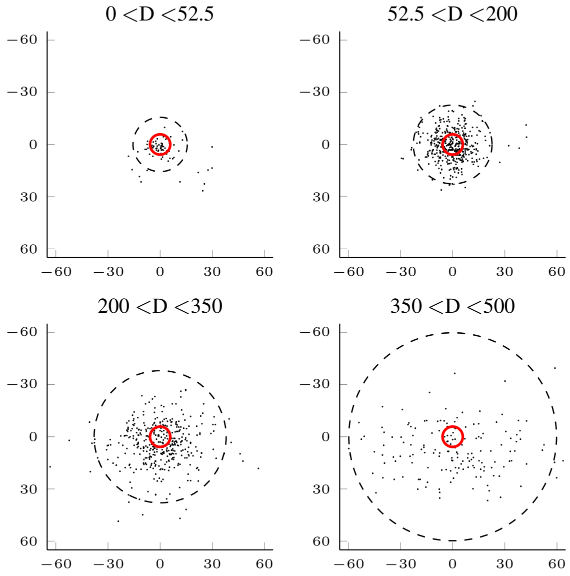


Figure 9. The distribution of touch points relative to the target across all users for chosen ranges of distance between gaze point and target. The dashed circles are the 95 % confidence circles. The red circles show the target visual boundary. All units in mm.

Distance range (Gaze to Target)	Mean distance (Gaze to Target)	95% confidence (Touch offset)
0-52.5mm	20.7mm (sd=9.3)	15.6mm
52.5-200mm	146.0mm (sd=29.6)	22.6mm
200-350mm	272.8mm (sd=33.1)	37.9mm
350-500mm	414.8mm (sd=32.1)	56.7mm

Table 1. 95% confidence values of positional offset for distance between touch and target for different visual guidance ranges. The range 0-52.5mm represent eyes-on tasks.

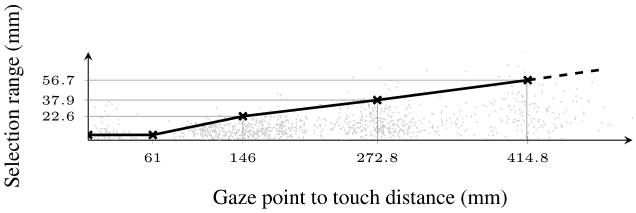


Figure 10. Selection range profile showing 95% confidence ranges from the first and second stages. Note that the minimum selection range on y axis does not need to start from a single pixel width for inputs that are inherently uncertain, such as touch.

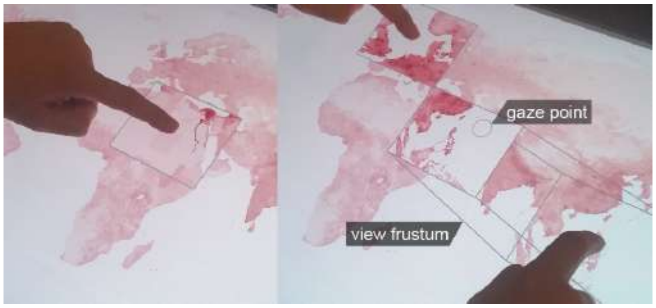
We use the 95% confidence value (61.0mm) for the distance between gaze and target in accurate pointing tasks as a threshold for increasing positional uncertainty due to decreased visual guidance. Together with the values from the peripheral target acquisition tasks at the first stage (Table 1), we plot a tentative profile of selection range with varying degrees of guidance (Figure 10). We should stress that the profile is intended as a provisional design guide. More importantly, not every potential action within the input selection range should be given the same weight for input handling. We consider a discrete confidence threshold to be most useful for cases that require making the selection range visually explicit to the user.

APPLICATIONS

We developed a number of applications to demonstrate the applicability of our design approach for a variety of use cases. The applications feature different combinations of input handling components. Below, we provide a conceptual breakdown of each application in terms of user input, action mechanisms and feedback techniques described above.

Application 1: Multifocus Image Exploration

A potential application case is multifocus interfaces (e.g., [7, 8]) that involve spatial juxtaposition of multiple points of interest. As opposed to time multiplexed methods such as zooming, multifocus interaction utilizes spatial multiplexing to display information [7]. A common aim in juxtaposition is to compare and correlate between multiple foci [8] and avoid redirecting gaze over long distances. The process of declaring multiple foci can be sequential or concurrent (e.g. using multi-touch). Gaze input can be a useful addition to multifocus interaction tasks, both as a focus point and for supporting input with decreased visual guidance.



User input: Manual input position (discrete, conservative), Interaction history (previously zoomed in regions are used to resolve positional ambiguity)

Action mechanisms: Select action (resolve positional uncertainty, increase selection range by zooming out)

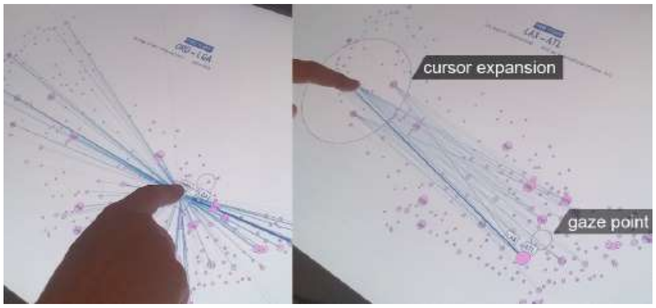
Feedback techniques: Peripheral awareness (show view frustum from finger to the gaze context), warp information content (warp lens near the primary touch point)

Figure 11. The degree of visual guidance is used to determine the position of lenses in image exploration. In the case of a touch event with visual guidance, the lens is shown at the touch position (left). In the case of a touch event with decreased visual guidance, the lens is warped near an existing lens (right).

Here, we demonstrate the use of input without visual guidance for exploring a map image that shows world population density (Figure 11). The application allows creating multiple lenses that are aligned edge to edge and controlled by individual touch points. The degree of visual guidance on different touch events is used to determine the primary touch point near the other lenses are aligned by its edge. The primary touch point is reevaluated with each touch down event. In the case of a touch event with visual guidance, the other lenses are warped to the new touch location. In the case of a touch event with decreased visual guidance, the lens is warped near the primary lens that the user's gaze is directed. In the latter case, positional uncertainty is handled in two ways. First, the lens covers an increased areal range. Second, positional uncertainty can be resolved by using interaction history by zooming into a previously viewed location.

Application 2: Exploring Relational Data

Another multifocus application case is interaction with relational data. We created a geospatial visualization of flight connections in the US (Figure 12). Interaction with the graph allows filtering flight connections based on the airports near the manual input. The degree of visual guidance is used to determine the positional uncertainty of touch points. For touch actions with high visual guidance, the application visualizes the connections from a single node that the user is pointing to. Positional uncertainty is increased and the cursor is expanded for pointing actions with decreased visual guidance. In this case, the gaze context of the user is used to resolve positional uncertainty; only the connections between manual selection range and airports near gaze point are visualized. In contrast to multifocus image application, the gaze context is not associated with an existing touch point and updates continuously with the movement of the touch.



User input: Manual input position (continuous, conservative), Gaze context (updates based on manual movement on touch surface, resolve positional uncertainty using relational data)

Action mechanisms: Select action (resolve positional uncertainty)

Feedback techniques: Peripheral awareness (expand the cursor to show positional uncertainty), Warp information content (show airport code on the gaze context)

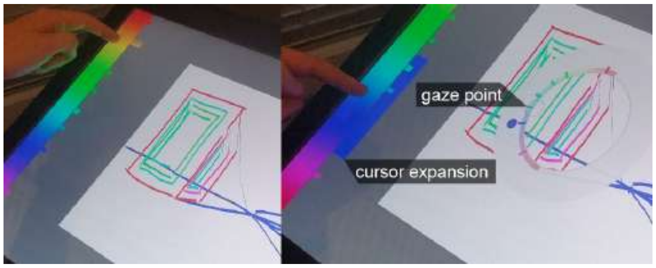
Figure 12. If the manual positions are interpreted as exact, the application visualizes all the connections from a single node (left). If the user gaze is directed elsewhere at the graph, the cursor is expanded to show increased positional uncertainty (right).

Application 3: Color Switching in Paint

Many interfaces involve sequences of tool switching and manipulation actions. Tool switching can be realized using toolbars, keyboard shortcuts or in-place selection techniques such as pie menus. Direct input devices such as touch screens allow bimanual action, where one hand performs tool selection while the other manipulates the target. In these cases, visual guidance can be used to qualify tool selection and manipulation actions. In this application we take color selection and painting on a canvas as an example of pair actions. Our application features two input fields: a continuous color selection bar and a virtual painting canvas.

Input field: Color palette

The color palette is configured as a vertical bar with varying hues along the y axis (Figure 13). The degree of visual guidance is determined continuously, corresponding to increased positional uncertainty for selection. In the case of selection with visual guidance, the system only considers the touch position for selecting a hue. In case of decreased visual guidance, the positional uncertainty is increased and the system uses additional user input resources for selection, namely the previous color selections (interaction history) and the colors in the gaze context of the user. The actual color selection is deferred to a touch release event. In the meantime, feedback of selection is displayed to the user through a radial color palette that appears on the gaze point. The gaze context is fixed at the start of a touch event and remains constant until the touch is released.



User input: Manual input position(continuous, conservative), Gaze context (the colors on the canvas region that the user looks at are used to resolve uncertainty for selection), Interaction history (previously used colors are used to resolve uncertainty)

Action mechanisms: Select action (resolve positional uncertainty), Defer action (defer the actual color selection to touch up event)

Feedback techniques: Peripheral awareness (increase the cursor size to show positional uncertainty), warp information content (create a color palette around gaze the context)

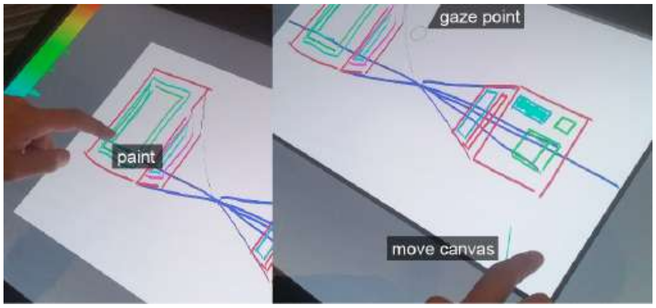
Figure 13. When conducted with a high degree of visual guidance the manual input is interpreted as exact (left). If the user gaze is directed elsewhere, such as on the canvas, the positional uncertainty is increased. In this case system provides feedback by visualizing a radial color palette at the gaze point (right).

Input field: Canvas

For input on canvas (Figure 14), the degree of visual guidance is determined discretely and is used to select action type. The degree of visual guidance at the moment of touch is used to determine if the action type is intended for painting (requiring fine degree of control) or moving the canvas (has area visual effect, thus requiring less visual guidance). The input type associated with a touch point remains stable until the touch is released.

EVALUATION OF APPLICATIONS

The various applications we developed combined different input handling techniques. We evaluated different applications to investigate 1) possible hand-eye coordination challenges that are general for input with decreased visual guidance and 2) the particular interaction challenges related to the action mechanisms and feedback techniques that vary among applications.



User input: Manual input position (discreet)

Action mechanisms: Select action (between types "paint" or "move")

Figure 14. Touch event on a canvas can be interpreted either as a paint (left) or move the canvas function (right) depending on the degree of visual guidance.

After finishing the target acquisition tasks, the participants proceeded into using the three applications. They were asked to perform open ended tasks with the example applications until they felt comfortable using the system (approximately 5 minutes). While our design approach does not require gaze point to be made explicit to the user, we still visualized the gaze as a translucent gray ring to inform participants in case the system loses track of their gaze. In this case, the translucent ring turned opaque, warning participants to correct their posture.

We video recorded participant interactions and collected their feedback after using each application. We also interviewed the participants at the end of the study to gather their overall feedback. In this section we report the participant feedback and observational data.

Participant Feedback & Video Analysis

Adjustment through Use

A common reaction among participants was the reported difficulty of "touching without looking" at the start of the session followed by gradual adjustment. Words "unnatural", "unintuitive", "strange" were often used to describe the initial experience, while the participants described their later experience as "natural" and "easier". Deliberately "avoiding drifting" of gaze to the touch location was observed during interaction and was also reported by participants as one reason for initial difficulty. *"At first it was of course quiet strange, pointing to a place where you can't see and your gaze tries to go there and you try to use your peripheral vision, but that also gets easier as you use it, you get used to the feeling of touching somewhere that you don't see."* (P4). The experience has also been compared to typing on a keyboard: *"... you start to write in the keyboard without looking, initially you look but you can try to do without looking..."* (P1)

Some participants described input handling mechanisms as "forgiving" and assistive of eyes-free input. *"I don't trust at all what I am seeing in my peripheral vision... knowing that it (eye tracking) being taken into account I trusted it even more and could predict and had some expectation of what will happen"* (P12). while others highlighted the need to know *"exactly where everything is"* before being able to point without looking. Although participant reaction differed regarding the degree of proficiency needed, individual confidence during eyes-free or peripheral pointing was a common dimension of use experience.

Being able to concentrate on the task such as drawing and *"using peripheral vision to do other specific tasks that are very obvious"* were highlighted as benefits. One participant also reported that gradual adjustment was useful for *"using two hands"*.

Gaze as Additional Pointer

The three applications are different regarding how explicitly they use the gaze context and how gaze updates in relation to touch events. The participant feedback helped identify potential benefits and drawbacks of different ways of using the gaze context. Explicit use of the gaze context has been welcomed as an additional "third hand" in flight visualization application and enabled concurrent access to three different locations (Figure 15). *"So the same feature can be done with two hands... but then I realized we only have two hands so maybe some possibility could be use your gaze as a third hand."* (P1)

Figure 15. Gaze being used as a third pointer in addition to two hands.

On the other hand, the use of gaze as an additional pointer, especially with dynamic update, caused hand eye coordination challenges. An additional challenge is the difference between the system and user interpretation of gaze context. Participants compared between the use of gaze in the multifocus image and flight visualization applications. *"I think it is easier to use it updating in a static way, because there is nothing that constantly change, I can compare more easily, there is nothing unexpected, but in dynamic I had more options, was good that it updates fast where I look, but then it was loosing..."* (P9).

Misinterpretation of Positional Uncertainty

We observed a number of instances in which a manual input action was wrongly interpreted as positionally uncertain due to the system's lack of awareness of the movement before the actual touch event. The instances usually involved the participant keeping his finger just above a specific point on the interface and performing the touch action while looking elsewhere (Figure 16). In these cases, although the participants knew exactly where they were pointing to, the application interpreted it as positionally uncertain and handled accordingly. The participant occasionally identified this as a "problem".

Figure 16. Interaction sequence leading to misinterpretation of positional uncertainty. The participant placed his right index finger on a region on the image (1). After a brief look, he lifted his right index finger from the touch screen but held it just above the surface (2), while pointing to another location with his left index finger (3). This was followed by a touch on the same point with the right index finger (4).

Screen Edge as Ambiguous Border and Tactile Guide

Although a touch screen is an input field with definite boundaries, decreased visual guidance can cause ambiguity for users regarding whether they are *addressing* the system during touch. In some instances, while aiming for the color palette near the edge of the screen, participants touched the insensitive bezel area of the screen (Figure 17). The lack of visual feedback (color palette warped to the gaze point) in this case communicated that the system is not addressed, which led the users to a repeated touch action. On the other hand, device borders provide potential tactile cues for eyes-free use. This was observed again for selecting colors, when participants *anchored* their left hand on the screen edge for sliding along the color palette with their thumb or index finger while keeping their gaze on the canvas (Figure 17).

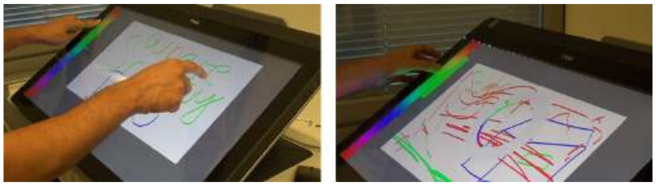


Figure 17. Instances of tapping on the bezel rather than the display area (left) and using the screen edge as a tactile guide (right).

DISCUSSION

We presented a design approach that targets supporting manual input with decreased visual guidance. Informed by previous literature [5, 14, 29], our main assumption was an increased positional uncertainty for input with decreased visual guidance. The first part of evaluation confirmed our assumption and provided a tentative profile for scaling selection range for different levels of visual guidance. The conceptualization of input with decreased visual guidance as input with uncertainty led to the design of various input handling and feedback techniques. The feedback we gathered from the participants during evaluation provides evidence for the viability of input with decreased visual guidance, although cognitive challenges related to hand-eye coordination and confidence during input were reported. Moreover, some such as addressing challenges and misinterpretation problems are mainly communicative challenges that might emerge in sensing or adaptive interfaces [1, 17, 27].

While we formulated manual input with decreased visual guidance as a design motivation, the specific benefits like increased functionality or satisfaction depends on a number of contextual factors such as physical setup and task. Manual input with decreased visual guidance can be forced by application context or preferred by the user. Similarly, various feedback techniques can be necessary when decreased visual guidance is forced by application context while they are not as essential when the users are able to direct their gaze to the input location.

At this point, we would like to discuss what we learned along the design process and evaluation through the analytical lens of spatial multiplexis [9]. Tuddenham et al. [32] limit the use of *bimanualism* to "two-handed one-object interaction", while using the terms *concurrent unimanualism* and *lateral sequential unimanualism* for "two-handed two-object interaction". The distinction is highly relevant regarding the degree of visual guidance during input. We argue that the benefit of supporting manual input with decreased guidance is most valid for multiple object interactions, since these cases involve direct and concurrent access to multiple interface regions. In painting application, this was observed when participants switched between colors with left hand and painted with their right hand (lateral sequential unimanualism) without redirecting their gaze. For multifocus image exploration and flight visualization the main pattern of interaction was concurrent unimanualism.

At the same time, gaze input requires revisiting the scope of space multiplexed user input. In their seminal paper, Fitzmaurice and Buxton investigate space-multiplexis through parallel use of hands [9]. However, the parallel use of manual input and eyes on different regions of interest [3, 15, 24] suggests that the scope of spatial multiplexis can be extended to the concurrent use of visual perception and input actions at different regions. In addition, spatial multiplexis can be extended to the interactions in which gaze is used not only for perception but also as a pointer. Many recent examples that combine gaze with indirect touch can be described as "one-hand+gaze one-object" or "two-hand+gaze one-object" interactions [21, 30, 33]. In contrast, during evaluation sessions, participants interpreted the explicit use of the gaze context as an additional "third hand", and performed "one-hand+gaze two-object", or "two-hand+gaze three-object" interactions.

Finally, gaze input requires revisiting the rationale of multiple object, space multiplexed input. Unlike multi-touch, gaze is a single channel but a very rapid input. Thus, using gaze coordinates for selection might favor sequential interaction over simultaneous selection of multiple targets, as in the case of sequential multiple target acquisition by gaze [21]. In this paper, we presented a design approach that targets multiple object interaction by supporting the use of touch with varying visual guidance as a direct input. However, further research is needed to evaluate the drawbacks and benefits of both approaches for different applications.

Limitations

Determining Visual Guidance

In the experimental design and applications we operationalized visual guidance as the distance of gaze point from the touch position at the moment of touch. This does not account for the complex hand-eye coordination over time that leads to a touch event: the motor movements can be accompanied by different levels of visual guidance between the initiation of the movement and touch. Planar input on the screen alone is not always sufficient for sensing this coordination, occasionally resulting in misinterpretation of positional uncertainty as reported above. A potential solution is over-the-screen sensing of the hand and finger movement to increase resources available to the system. The problem is not as significant for mouse or other cursor based movement, since the cursor position information is always available to the system.

Stability of Input Field

Our design approach assumes the stability of the input field. Keyboard, fixed toolbars and geospatial data are relatively stable input fields, which users either have prior knowledge of or get accustomed to. However, more information is needed to determine the viable scope of performing manual input with decreased visual guidance.

Non-planar Surfaces

It should also be noted that we limited our scope of investigation to a single planar interactive surface. However, our design approach can be applied to a range of settings that are spatially more diverse, such as distributed, on-body or virtual interfaces. These settings raise a number of questions such as determining positional uncertainty and solving the addressing problems described earlier.

CONCLUSION

Potential ubiquity of eye tracking in the near future calls for reevaluating the division of labor between different input modalities and the role of gaze within. In this paper, we contributed to the ongoing discussion on how gaze can be integrated with other modalities. Our proposed design approach utilizes gaze to qualify direct manual input by taking into account the level of visual guidance the input is performed with. In the case of input with visual guidance the system allows familiar interaction, while the input with decreased visual guidance is supported through various action mechanisms and feedback techniques for handling input with uncertainty. Adaptive handling of input, in return, supports concurrent access to multiple locations in an interface. We consider the design space we developed as a starting point for a systematic exploration of interfaces that adapt manual input handling in relation to visual guidance.

ACKNOWLEDGMENTS

This work has been partly supported by MindSee (FP7 ICT; Grant Agreement #611570). We would like to thank the anonymous reviewers for their insightful comments and to the participants of the study.

REFERENCES

1. Bellotti, V., Back, M., Edwards, W. K., Grinter, R. E., Henderson, A., and Lopes, C. Making sense of sensing systems: five questions for designers and researchers. In *Proc. CHI'02*, ACM (2002), 415–422.
2. Bi, X., and Zhai, S. Bayesian touch: a statistical criterion of target selection with finger touch. In *Proc. UIST'13*, ACM (2013), 51–60.
3. 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. ETRA'10*, ACM (2010), 89–92.
4. Blanch, R., and Ortega, M. Rake cursor: improving pointing performance with concurrent input channels. In *Proc. CHI'09*, ACM (2009), 1415–1418.

5. Bock, O., and Eckmiller, R. Goal-directed arm movements in absence of visual guidance: evidence for amplitude rather than position control. *Experimental Brain Research 62*, 3 (1986), 451–458.
6. Brown, L. E., Rosenbaum, D. A., and Sainburg, R. L. Limb position drift: implications for control of posture and movement. *Journal of Neurophysiology 90*, 5 (2003), 3105–3118.
7. Butscher, S., Hornbæk, K., and Reiterer, H. Spacefold and physiclenses: simultaneous multifocus navigation on touch surfaces. In *Proc. AVI'14*, ACM (2014), 209–216.
8. Elmqvist, N., Riche, Y., Henry-Riche, N., and Fekete, J.-D. Mélange: Space folding for visual exploration. *IEEE TVCG 16*, 3 (2010), 468–483.
9. Fitzmaurice, G. W., and Buxton, W. An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. In *Proc. UIST'97*, ACM (1997), 43–50.
10. Fitzmaurice, G. W., Ishii, H., and Buxton, W. A. Bricks: laying the foundations for graspable user interfaces. In *Proc. CHI'95*, ACM (1995), 442–449.
11. Flanagan, J. R., Terao, Y., and Johansson, R. S. Gaze behavior when reaching to remembered targets. *Journal of neurophysiology 100*, 3 (2008), 1533–1543.
12. Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., and Schutte, P. C. The h-metaphor as a guideline for vehicle automation and interaction.
13. Goodwin, C. *Pointing: Where language, culture and cognition meet*. Lawrence Erlbaum Associates, 2003, ch. Pointing as situated practice, 217–241.
14. Hagiya, T., and Kato, T. Probabilistic touchscreen text entry method incorporating gaze point information. *Journal of Information Processing 23*, 5 (2015), 708–715.
15. Huang, J., White, R., and Buscher, G. User see, user point: gaze and cursor alignment in web search. In *Proc. CHI'12*, ACM (2012), 1341–1350.
16. Jacob, R. J. What you look at is what you get: eye movement-based interaction techniques. In *Proc. CHI'90*, ACM (1990), 11–18.
17. Jacucci, G., Spagnolli, A., Freeman, J., and Gamberini, L. Symbiotic interaction: a critical definition and comparison to other human-computer paradigms. In *Symbiotic Interaction'14*. Springer, 2014, 3–20.
18. Kosunen, I., Jylha, A., Ahmed, I., An, C., Chech, L., Gamberini, L., Cavazza, M., and Jacucci, G. Comparing eye and gesture pointing to drag items on large screens. In *Proc. ITS'13*, ACM (2013), 425–428.
19. Kumar, M., Paepcke, A., and Winograd, T. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proc. CHI'07*, ACM (2007), 421–430.
20. Mankoff, J., Hudson, S. E., and Abowd, G. D. Providing integrated toolkit-level support for ambiguity in recognition-based interfaces. In *Proc. CHI'00*, ACM (2000), 368–375.
21. Pfeuffer, K., Alexander, J., Chong, M. K., and Gellersen, H. Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In *Proc. UIST'14*, ACM (2014), 509–518.
22. Pfeuffer, K., Alexander, J., Chong, M. K., Zhang, Y., and Gellersen, H. Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze. In *Proc. UIST'15*, ACM (2015), 373–383.
23. Pohl, H., and Murray-Smith, R. Focused and casual interactions: Allowing users to vary their level of engagement. In *Proc. CHI'13*, ACM (2013), 2223–2232.
24. Rodden, K., Fu, X., Aula, A., and Spiro, I. Eye-mouse coordination patterns on web search results pages. In *Proc. CHI EA'08*, ACM (2008), 2997–3002.
25. Schwarz, J., Hudson, S., Mankoff, J., and Wilson, A. D. A framework for robust and flexible handling of inputs with uncertainty. In *Proc. UIST'10*, ACM (2010), 47–56.
26. Schwarz, J., Mankoff, J., and Hudson, S. E. An architecture for generating interactive feedback in probabilistic user interfaces. In *Proc. CHI'15*, ACM (2015), 2545–2554.
27. Serim, B. Querying and display of information: symbiosis in exploratory search interaction scenarios. In *Symbiotic Interaction'14*. Springer, 2014, 115–120.
28. Sibert, L. E., and Jacob, R. J. Evaluation of eye gaze interaction. In *Proc. CHI'00*, ACM (2000), 281–288.
29. Spijkers, W., and Spellerberg, S. On-line visual control of aiming movements? *Acta psychologica 90*, 1 (1995), 333–348.
30. Stellmach, S., and Dachselt, R. Look & touch: gaze-supported target acquisition. In *Proc. CHI'12*, ACM (2012), 2981–2990.
31. Stellmach, S., and Dachselt, R. Still looking: investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proc. CHI'13*, ACM (2013), 285–294.
32. Tuddenham, P., Kirk, D., and Izadi, S. Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In *Proc. CHI'10*, ACM (2010), 2223–2232.
33. Turner, J., Alexander, J., Bulling, A., Schmidt, D., and Gellersen, H. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *INTERACT'2013*. Springer, 2013, 170–186.
34. Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. In *Proc. CHI'87*, ACM (1987), 183–188.
35. Weir, D., Pohl, H., Rogers, S., Vertanen, K., and Kristensson, P. O. Uncertain text entry on mobile devices. In *Proc. CHI'14*, ACM (2014), 2307–2316.
36. Weiss, M., Wacharamanotham, C., Voelker, S., and Borchers, J. Fingerflux: near-surface haptic feedback on tabletops. In *Proc. UIST'11*, ACM (2011), 615–620.

37. Weiss, M., Wagner, J., Jansen, Y., Jennings, R., Khoshabeh, R., Hollan, J. D., and Borchers, J. Slap widgets: bridging the gap between virtual and physical controls on tabletops. In *Proc. CHI'09*, ACM (2009), 481–490.
38. Wigdor, D., Williams, S., Cronin, M., Levy, R., White, K., Mazeev, M., and Benko, H. Ripples: utilizing per-contact visualizations to improve user interaction with touch displays. In *Proc. UIST'09*, ACM (2009), 3–12.
39. Williamson, J. *Continuous uncertain interaction*. PhD thesis, University of Glasgow, 2006.
40. Wu, J., Yang, J., and Honda, T. Fitts law holds for pointing movements under conditions of restricted visual feedback. *Human movement science 29*, 6 (2010), 882–892.
41. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (magic) pointing. In *Proc. CHI'99*, ACM (1999), 246–253.