
Gaze and Mouse Coordination in Everyday Work

Daniel J. Liebling
Susan T. Dumais
Microsoft Research
One Microsoft Way
Redmond, WA 98052
USA
danl@microsoft.com
sdumais@microsoft.com

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. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UbiComp '14 Adjunct, September 13 - 17, 2014, Seattle, WA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3047-3/14/09...\$15.00.
<http://dx.doi.org/10.1145/2638728.2641692>

Abstract

Gaze tracking technology is increasingly common in desktop, laptop and mobile scenarios. Most previous research on eye gaze patterns during human-computer interaction has been confined to controlled laboratory studies. In this paper we present an *in situ* study of gaze and mouse coordination as participants went about their normal activities. We analyze the coordination between gaze and mouse, showing that gaze often leads the mouse, but not as much as previously reported, and in ways that depend on the type of target. Characterizing the relationship between the eyes and mouse in realistic multi-task settings highlights some new challenges we face in designing robust gaze-enhanced interaction techniques.

Author Keywords

Mouse; gaze tracking; target acquisition; multimodal input

ACM Classification Keywords

H.5.2. User Interfaces

Introduction

Most prior research in eye tracking to support better human-computer interaction has been conducted in laboratory settings. Observational studies have been used to characterize how individuals attend to stimuli or

interact with interfaces. At a high level, heat maps show overall attention to different regions or interface elements in aggregate [14], [21]. At a finer-grained level of analysis, gaze tracking has been used to improve our understanding of interactions such as how people point to items [7], how they search and select menu items [1], [3], and how they inspect search results [13] [15]. In addition to observational studies, new gaze-based interaction techniques have been developed to improve human-computer interaction. Examples of gaze-enhanced interactions include using gaze to select items [11], accelerate the cursor to the point of gaze [22], differentially render content based on gaze [13], and automatically scroll text [12].

Given the previous cost and size of gaze tracking devices, most research was conducted in laboratory settings with small numbers of participants, so very little is known about eye movements while using computers across tasks in the real world. Recently, small devices that can be mounted on any display are now available from several manufacturers for less than US\$100, along with open software development kits. The availability and portability of these devices creates new opportunities to collect gaze data in real-world settings and to use resulting insights to design new presentation and interaction techniques. Vrzakova and Bednarik argue that longer term perspective of attention in natural scenes is an important direction for the community [19], and our research provides a step in that direction. We examine the coordination of mouse and eye movements in the context of normal work patterns, characterize the varied patterns of coordination, and discuss implications for making some interaction techniques more robust.

Related Work

Eye tracking has been used to study pointing tasks in both the physical world and in computer settings. Helsen et al. [7] studied the spatio-temporal coupling of the eyes and hand movements in physical pointing tasks. Using a reciprocal pointing task with two fixed targets, they found consistent patterns of eye-hand coordination. On average, eye movements are initiated 70 ms earlier than hand movements, the eyes make two saccades to acquire the target, and stabilize on the target at 50% of the total hand response time.

In computer applications, pointing is often carried out with the aid of a cursor, which does not have a direct absolute mapping to target location or proprioceptive feedback that characterizes pointing in the physical world. A common observation in computer interaction is that the “eyes lead the mouse” in pointing, with the eyes first acquiring the target and the cursor then following for selection. Zhai et al.’s Manual and Gaze Input Cascaded (MAGIC) technique [22] builds on this assumption to warp the cursor to the vicinity of the point of gaze, thus reducing the distance that the cursor needs to move to acquire the target. In their experiments, participants were asked to point and click at targets appearing at random positions on the screen. Two techniques were used to warp the cursor to the point of gaze – a *liberal* technique that warped the cursor whenever the eyes moved more than 120 pixels from the starting position, and a *conservative* technique that did not warp the cursor to a target until the cursor was activated. The liberal method was easy to use since it did not require coordinated action and sometimes led to faster selection than the cursor alone. Automatic activation, however, can result in a “Midas touch” problem [20] in which everything a user looks at

is selected, and this is likely to be more problematic in non-laboratory environments that are not carefully controlled and contain more than a single target.

Smith et al. [16] studied eye-cursor coordination in target selection. They used a reciprocal pointing task in which participants alternately selected two fixed targets, and a random pointing task in which participants selected targets presented at random locations. Although gaze was often near the target before the cursor, coordination patterns varied across tasks, pointing devices, and individuals. Three types of patterns were: the eyes lead the cursor, the eyes follow the cursor, and eye gaze switches back and forth from target to cursor. The frequency of these patterns is unclear, except that switching was not common. Bieg et al. [2] examined eye-mouse coordination in visual search and selection. They considered three tasks: a single target to the right of a fixation point, a single target (specified by its color and shape) in a grid of targets; and a single target (specified by its color and shape) in a random field of targets. When target location was unknown (third task), the eyes lead the mouse by 300 ms on average. When the approximate location of the target was known (first and second tasks), the cursor often led gaze in acquiring the target, and fixations on the target occurred later in the pointing process. Knowledge about the target location is likely to be important in non-laboratory settings.

Mouse and gaze alignment has also been studied in somewhat richer tasks, to evaluate the extent to which mouse position could be used instead of gaze. Chen et al. [4] examined mouse and gaze movements during web browsing. During certain subtasks, mouse and

gaze movements were often correlated. They found that the average distance between mouse and gaze was 90 pixels during transitions from one area of interest (AOI) to another, and that 40% of the distances were closer than 35 pixels. Similarly, Huang et al. [10] evaluated mouse and gaze behavior during Web search. They found that the average distance between the eye and mouse was 178 px, with the differences in the x -direction being larger (50 px) than in the y -direction (7 px).

Understanding how the eyes and mouse interact in the real world provides a basis for developing models, grounded in data, which take into account the richness and variety of interactions in practical settings. Since gaze provides a measure of a user's attention, knowing when the mouse and gaze are aligned can help strengthen models that use the mouse as proxies for attention [5], [10]. Furthermore, clicks can provide ground truth for improving calibration [9] by leveraging highly correlated mouse movements to compensate for changes in tracker accuracy over time.

In this paper, we present the results from a study of eye and mouse coordination outside of the laboratory. Using a system which simultaneously records mouse and gaze movements as well as metadata about controls that were clicked, we provide a rich picture of different coordination patterns. We characterize several different relationships between gaze and cursor activity around the time of mouse clicks, depending on target type and application. These insights have implications for the design of gaze-enhanced selection techniques.

System Design

For our study, we used a Tobii REX Laptop Edition 30 Hz eye tracker. In the manufacturer's tests¹, accuracy ranges from 0.4° to 1.0° visual angle and precision varies from 0.32° to 0.97° depending on viewing angle and lighting [18]. At a viewing distance of 50 cm from a typical 1920 × 1080 resolution screen, one degree of visual angle is approximately 21 pixels which is about the size of a small control like a checkbox.

After per-user calibration, the Tobii device sends input (including timestamp and x and y gaze coordinates) to the PC. Since gaze data is inherently noisy due to both systematic noise and measurement errors, we smoothed the gaze points using Stampe's two stage filter [17] and a two sample weighted average. In addition to the gaze input, our software records each mouse event (x and y screen position) via the UI event preview mechanism provided Windows. The program registers to preview global input events prior to forwarding them downstream to the active application. The Windows input processing stack adds only about 1 ms of latency to the system.

If the mouse event is a click, the system also screenshots a window of 200 × 200 pixels centered on the click point as well as metadata about the click target such as process name, control type (e.g. *Button*, *ListItem*, *Scrollbar*), control caption, and target size. The program obtains these data using the Windows Accessibility API. We record all signals through Event Tracing for Windows (ETW), an extremely low-latency real-time binary tracing framework that also gives us

¹ Per the manufacturer, Tobii REX measurements are similar to those of the X2-30 device, reported here.

high precision timestamps generated from the system clock. The median difference between timestamps reported by the tracker and timestamps in the log is 0.03 ms. The median inter-sample interval for mouse events is 8.0 ms (125 Hz).

Figure 1 shows an example screenshot for a click event. Gaze and mouse trajectories for one second preceding a mouse click are shown in blue and red, respectively. Open symbols show the beginning of the trajectories and filled symbols the end; the yellow diamond marker shows the click point. In this example, the participant clicked on the "Send" button in Microsoft Outlook, an email client. The gaze and mouse started off about 200 px away from each other, took different trajectories, but converged at the time and location of the click.

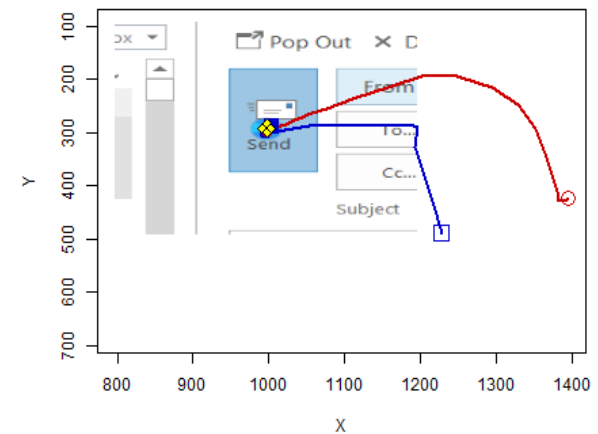


Figure 1. Gaze (blue/boxes) and mouse (red/circles) paths 1000 ms prior to clicking the *Send* button in Microsoft Outlook. Open symbols (\square , \circ) show path beginnings.

| Control | Percent |
|------------|---------|
| Document * | 12.1 |
| Button * | 11.9 |
| Edit * | 10.3 |
| Pane * | 10.0 |
| Custom * | 7.7 |
| DataItem | 6.5 |
| MenuItem * | 5.1 |
| Tab * | 5.1 |
| TreeItem * | 5.0 |
| DataGrid | 4.1 |

Table 1. Ten most frequently clicked control types. (*) Indicates five or more subjects.

Data Collection

We recruited twelve participants (10 male) from a large technology company with vision at or corrected to normal; half wore corrective lenses. Each participant sat at their own desk, seated 50-60 cm from the display. Participant displays ranged from 1920 × 1080 px to 2560 × 1600 px, with the latter being a Dell 24" (61 cm) diagonal display, used by half the participants. Although each participant used multiple displays, we connected the eye tracker to the primary display. We affixed the tracker to the center of the bottom bezel, then calibrated the tracker with the manufacturer's procedure. All participants used a mouse except for one who used a trackball. We asked each participant to use their PC normally and left the room for 30 to 50 minutes, after which we returned to terminate the logging system and uninstall the hardware. Because of system problems, one participant's data was captured poorly, leaving us with data from 11 participants.

Results

Overview of the data

In total we obtained 378 minutes of recorded data, which included 485,763 gaze points, 442,071 mouse track points, and 3,681 mouse clicks over 32 different

classes of targets. Table 1 shows the percentage of clicks on the ten most frequently clicked control types, which together account for 78.0% of clicks. Items marked with an asterisk (*) appear for 5 or more subjects. Average control sizes ranged from 821 px² for labeled *Checkboxes* to 3451 px² for *MenuItems* to 2.38 megapixels for *Panes*. Thus, our data contain a diverse set of target types spanning four orders of magnitude of size.


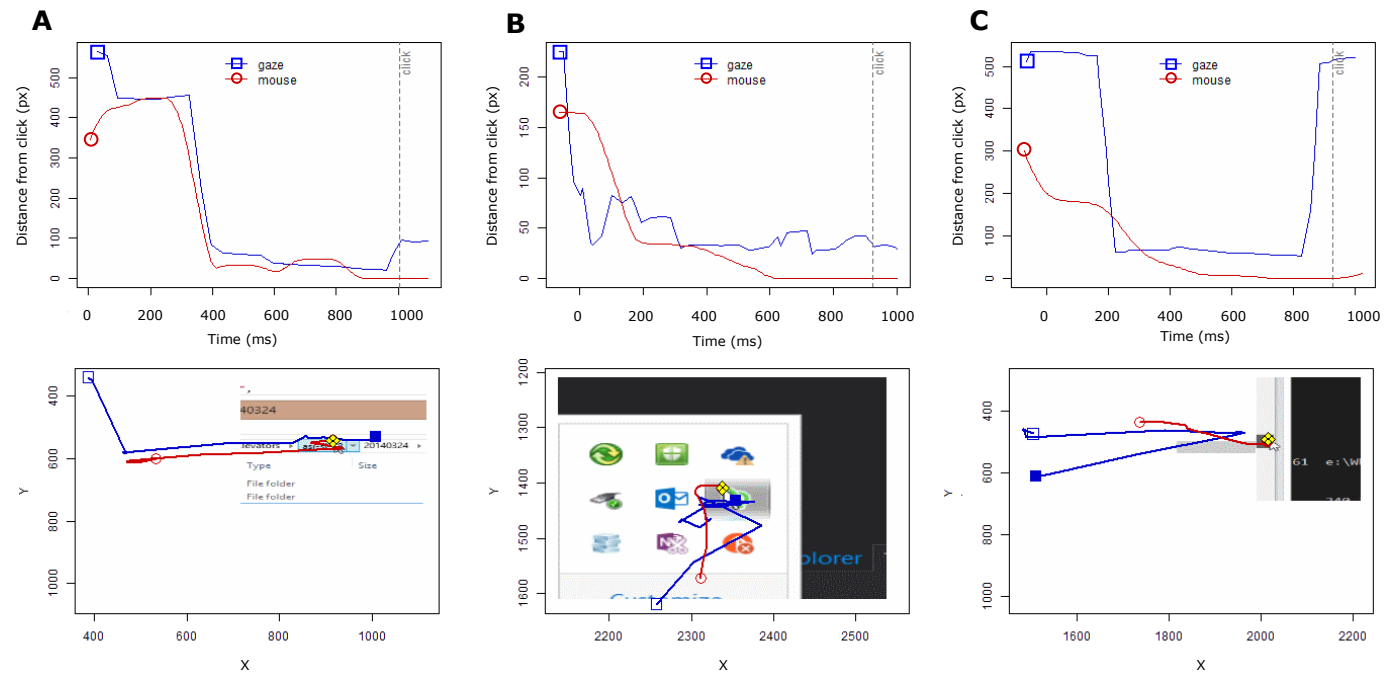
Figure 2 shows three example traces of the data that we collected. The top row shows the distance between the eyes and cursor for 1000 ms before and 200 ms after the click. The bottom row shows screen shot of the clicked region, as described earlier in Figure 1. The columns illustrate three different patterns of eye mouse coordination. On the left column (A), the eyes and mouse move in a coordinated fashion to select a split button control. In the center (B), gaze leads the mouse by about 150 ms in selecting the  icon from the Windows 8 notification area. In the right column (C), the mouse leads gaze by 250 ms in selecting the scroll bar thumb; gaze is present on this interface element for 600 ms but leaves the target 100 ms *before* the click occurs.

Figure 2. Mouse (red/circles) and gaze (blue/boxes) distances from click point over time (top row) and paths (bottom row) for three events. From left to right, (A) synchronized movement, (B) eye leading, and (C) mouse leading behaviors. In the far right, the eyes leave the target prior to the click. Open (\square , \circ) and closed (\blacksquare , \bullet) symbols show path beginnings and endings, respectively.



Determining gaze lead or lag

Clicks are useful to delimit the recorded movement streams. They are clear signals of action intent, and coupled with the target metadata, give insight into the underlying task. Prior work has shown that the eyes tend to lead the mouse in acquiring the target, but other gaze-mouse relationships are also seen during target acquisition. Characterizing the relationship is straightforward in controlled laboratory settings where discrete pointing trials are used and the cursor and eyes are aligned at the start of each trial. In the wild, task boundaries are not specified, and the alignment varies over time with gaze often starting and ending in places far from the click point. We begin by examining several different measures of eye mouse coordination in our naturalistic setting.

TIME BEFORE CLICK

Using the click as the point of interest, we measure the alignment of the cursor and gaze at different times prior to the click. Table 2 shows the percentile differences between the cursor and gaze as a function of time for five time intervals before the click. Not surprisingly, the distances tend to be larger 1000 ms before the click. The distances are lowest from 100 ms to 250 ms before the click, and then increase again at the click time. The median distance (50th percentile) between the two is largest 1000 ms before the click (171 px), decreases to its minimum at 250 ms before the click (74 px), and increases to 89 px at the time of the click. The somewhat larger distance at the click point may be due to the eyes leaving the target before the click, as shown in Figure 2C, top right.

| Time to click (ms) | Mouse-gaze distance (px) percentiles | | |
|--------------------|--------------------------------------|------------------|------------------|
| | 25 th | 50 th | 75 th |
| 1000 | 69.9 | 171.5 | 433.0 |
| 500 | 49.8 | 96.3 | 217.8 |
| 250 | 41.9 | 73.6 | 172.4 |
| 100 | 40.8 | 76.5 | 194.6 |
| 0 | 44.1 | 89.2 | 253.2 |

Table 2. Summary of mouse-gaze distance by time to click

CLICK POINT

To measure the coordination between the cursor and gaze we examine when gaze leads or lags the cursor in acquiring the target. We do this by measuring the time at which the gaze arrives near the clicked point. Because of inaccuracies in measuring gaze location, gaze may never reach the precise point of the click, so we define “reaching the click point” as gaze appearing within 50 pixels of the click point. We evaluate the lead or lag of the eyes by finding the difference between the time at which the eyes and mouse first come within 50 pixels of the click. Figure 3 (black line) shows the distribution of these time differences. The median arrival time is -137 ms before click and the mass of the tail is skewed toward the left, with the 25th and 75th percentiles at -422 ms and 5 ms, respectively. Gaze preceded the mouse to the click only 63.5% of the time. The extent to which gaze precedes the cursor varied by control type. Gaze reached within 50 pixels of the click point before the cursor did just 31.7% of the time for *TitleBar*, 49.5% for *Button* and 85.7% for *List*. It is worth noting that applications vary in how they use the control types. For example, in the Outlook mail

client the message list, which spans the height of the application and a substantial portion of its width, is an instance of the *List* control.

ENTRY INTO TARGET REGION

Since targets vary widely in size, we also examined the lag between when the eyes and cursor first acquire the target (entering the bounding box). Figure 3 (gray line) shows the distribution which is quite similar to the previous method. The median time is -99.9 ms, and the mass of the tail is skewed toward the left, with the 25th and 75th percentiles at -424 ms and 20 ms, respectively. 64.2% of the time the eyes precede the mouse entering the click area. Although the distributions are very similar, the set of controls that tend to be first entered by gaze differs. For the *List* control gaze enters the target before the mouse 28.6% of the

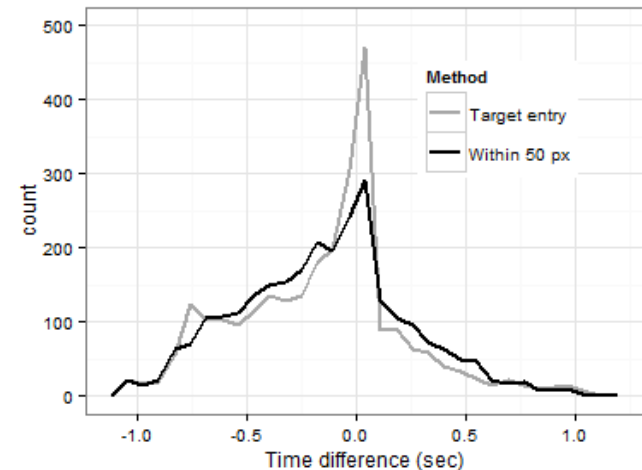


Figure 3. Distribution of differences in mouse and gaze time offsets for two different methods. Negative times indicate the gaze leads the mouse.

time (vs. 85.7% for the 50-pixel method). For the *Button* control the eyes lead in 69.6% of cases (vs. 49.5% for the 50-pixel). These differences likely have a variety of causes; for example, interaction sequences are often spatially clustered (such as within a dialog box), so the 50-pixel method is likely to have false positives since the gaze is already nearby in this case.

First saccade

Finally we examine the alignment of gaze and mouse before and after the first saccade toward the click target. Examples of saccades can be seen in Figure 2 at times 350, 0, and 200 ms in the three columns. (There is also a saccade away from the click point in the right column.) Saccades are an interesting measure because they may provide an actionable signal of change of focus, something that interaction designers could use to adapt the experience. We used a modest threshold of 200 px/sec to detect the first saccade before each click, which we found in 67.2% of cases. Table 4 shows the median distances to the click point around the time of the saccade. The saccade occurs at median time 816 ms prior to the click. Upon saccade completion, the median distance from mouse to click point only decreases by 2.2% while the median distance between gaze and click point decreases by 19.3%.

Additional observations

Controlled laboratory studies examining the coordination between gaze and cursor during pointing and visual search tasks provide important insights about the underlying perceptual and control mechanisms. However, in more natural settings, task boundaries are not clearly delineated, familiarity with applications and controls varies, and observed interaction patterns are more complex and nuanced.

TRANSITION TO NEXT ACTION

In our naturalistic observations we see clearly that target acquisition is one step of a typically more complex task. For example, we find gaze often leaves the target area, moving along to the next task before the click is completed. Figure 2C shows an example of this in which the gaze leaves the target area 100 ms before the click occurs. To estimate how often this occurs we identify cases where the mean distance from gaze to click after last saccade is higher than the distance before that saccade. In other words, the last saccade moves away from the click point. This occurred 7.7% of the time. We believe that this underestimates the frequency and that more sophisticated analysis of the gaze trajectories will identify more cases.

One implication of this finding is that interaction techniques like cursor warping or target expansion could interfere with completing the current click if the technique is invoked too quickly to gaze events alone. For example, techniques like MAGIC are based on using saccades alone or saccades coupled with mouse actuation to warp the cursor to the gaze position. Given our observations, warping techniques will also need to take into account gaze stability in a window around mouse movement rather than just sampling the gaze position prior to warping. Furthermore, if we are able to predict the type or size of target control with high accuracy, we could use this to further refine the warp based on empirical evidence.

INFLUENCE OF PREVIOUS EXPERIENCE

Previous experience with applications can influence the coordination of gaze and cursor in interesting ways that have not been previously observed in the laboratory. In mail applications such as Outlook, for example, the

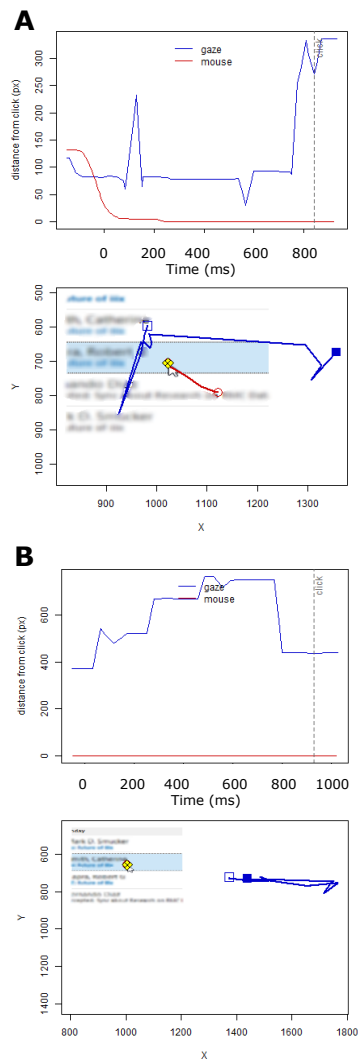


Figure 4. Two consecutive clicks in a message list. (A) Mouse and gaze are mostly stable. (B) The mouse remains parked while the eye saccades.

screen is divided into three regions: a folder list, a message list, and a reading pane. When the cursor selects a message from the list, the corresponding text is displayed in the reading pane. At left, Figure 4 shows an example of two consecutive clicks in the message list. The cursor is “parked” in the message list and doesn’t move very much. On the top, the mouse is quite close to the selected item, and the eyes move in from the reading pane to verify the item about to be selected and leave to go to the reading pane before the actual click happens. On the bottom, the gaze and mouse are decoupled because the reading continues and the cursor simply moves up one item in the list. Similar behavior is also seen sometimes when the mouse has acquired the scroll bar, and successive clicks on the bar occur as the eyes remain in the reading area without any additional visual control.

Prior knowledge of the locations and characteristics of frequently clicked targets can also influence gaze-mouse coordination. For example, the Windows “start button” is usually located in the bottom left corner of the screen. Although the target is relatively small in size (64×64 px) and would ordinarily require gaze to acquire accurately, we observed ballistic mouse movements that “jam” the mouse into the bottom left corner, because the OS does not allow the cursor run off the screen. The gaze was not required to acquire this point before clicking because of the prior knowledge of the constraints.

Discussion and Future Work

In this paper, we present a system to simultaneously record gaze and mouse behavior in natural interactions with desktop applications. We analyze the coordination of gaze and mouse behavior before and after clicks.

Compared to previous observations from laboratory studies, we find more complex and nuanced patterns. Using a variety of measures, we show that the eye leads the mouse click only about two thirds of the time, and that this depends on type of target and familiarity with the application. Design suggestions based on results from earlier laboratory studies will need to be enriched to accommodate more realistic interaction patterns observed in the wild.

For example, interaction techniques such as MAGIC cursor acceleration make the assumption that gaze leads the mouse, and that saccades (in combination with dwell time or mouse activation) can be used to warp the mouse to the point of gaze. Given the variability we observe in the coordination of gaze and mouse, including gaze leaving the target before a click 7.7% of the time, warping techniques will also need to take into account gaze patterns over time rather than at an individual point. Further, if we can predict the type or size of target control with high accuracy, we could use this to further refine the warping technique based on empirical evidence.

Similarly, Hornof and Halverson [9] showed how eye trackers can be continuously calibrated using “required fixation locations.” Expanding this technique to open world interaction has tremendous potential, but will depend on confidently knowing when the mouse and gaze are aligned spatially and temporally. Developing techniques to accurately predict when gaze and the mouse are aligned will be needed to enable this and other techniques.

This paper shows that most of the time, gaze and mouse behave as expected given existing literature.

However, about one third of the time, they behave differently. We believe that data collection and analysis in more realistic settings is a key direction to pursue in developing robust gaze-enhanced interaction techniques.

References

- [1] Aaltonen, A., Hyrskykari, A., and Raiha, K.-J. 101 spots, or how do users read menus? *Proc. CHI 1998*.
- [2] Bieg, H.-J., Reiterer, H., and Bülthoff, H.H. Eye and pointer coordination in search and selection tasks. *Proc. ETRA 2010*.
- [3] Byrne, M.D., Anderson, J.R., Douglass, S., and Matessa, M. Eye tracking the visual search of click-down menus. *Proc. CHI 1999*.
- [4] Chen, M.-C., Anderson, J.R., and Sohn, M.-H. What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. *Proc. CHI 2001 Extended Abstracts*.
- [5] Diaz, F., White, R., Liebling, D., and Buscher, G. Robust models of mouse movement on dynamic web search results pages. *Proc. CIKM 2013*.
- [6] Duchowski, A. *Eye Tracking Methodology: Theory and Practice*. London: Springer-Verlag, 2007.
- [7] Helsen, W.F., Elliott, D., Starkes, J.L., and Ricker, K.L. Temporal and spatial coupling of point of gaze and hand movement in aiming. *Journal of Motor Behavior* 30 (3), 1998, 249-259.
- [8] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van der Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 2011.
- [9] Hornof, A., and Halverson, T. Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments & Computers* 34 (4), 2002, 592-604.
- [10] Huang, J., White, R., and Buscher, G. User see, user point: Gaze and cursor alignment in web search. *Proc. CHI 2012*.
- [11] Jacob, R.J.K. What you look at is what you get: Eye movement-based interaction techniques. *Proc. CHI 1990*.
- [12] Kumar, M., and Winograd, T. Gaze-enhanced scrolling techniques. *Proc. UIST 2003*.
- [13] Levoy, M., and Whittaker, R. Gaze-directed volume rendering. *Proc. SIGGRAPH 1990*.
- [14] Matejka, J., Grossman, T., and Fitzmaurice, G. Patina: Dynamic heatmaps for visualizing application usage. *Proc. CHI 2013*.
- [15] Rodden, K., Fu, X., Aula, A., and Spiro, I. Eye-mouse coordination patterns on web search results pages. *Proc. CHI 2008 Extended Abstracts*.
- [16] Smith, B., Ho, J., Ark, W., and Zhai, S. Hand eye coordination patterns in target selection. *Proc. ETRA 2000*.
- [17] Stampe, D.M. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments & Computers* 25 (2), 1993, 137-142.
- [18] Tobii Technology AB. Accuracy and precision Test report X2-30 fw 1.0.1. 28 May 2013.
- [19] Vrzakova, H., and Bednarik, R. EyeCloud: Cloud Computing for pervasive eye-tracking. *Proc. PETMEI 2013*.
- [20] Vrzakova, H., and Bednarik, R. That's not normal: A detailed analysis of Midas touch in gaze-based problem-solving. *Proc. CHI 2013 Extended Abstracts*.
- [21] Wooding, D.S. Fixation maps: Quantifying eye-movement traces. *Proc. ETRA 2002*.
- [22] Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. *Proc. CHI 1999*.