

Touch Input and Gaze Correlation on Tablets

Pierre Weill-Tessier and Hans Gellersen

School of Computing and Communications,
Lancaster University, UK

{p.weill-tessier,h.gellersen}@lancaster.ac.uk

Abstract. Prior work has shown gaze behaviour correlates with mouse movement in conventional desktop interfaces. As many devices now employ direct touch instead of a mouse, we have conducted a first study analysing how gaze correlates with touch input. We collected touch and gaze data on a tablet from 24 participants, on search, shopping, and “link-following” tasks representative of typical online use of tablets. Our analysis shows a gaze fixation typically leads touch input by 337ms within 197 pixels. We observed a stronger correlation in the context of link-following, in comparison with touch on other HTML objects and virtual keyboard.

Keywords: Touch input, Eye tracking, Gaze, Hand-eye correlation, Tablet interaction

1 Introduction

When interacting with computing devices, manual input is highly connected with how users visually inspect UI content. The correlation between manual input (using the mouse as a proxy for the hand) and gaze has been of particular interest in many research efforts [3,5,10], to better understand visual attention across the variety of computing devices we use on a daily basis, and to propose new concepts that enhance the interaction. However, besides the increasing popularity of tactile devices, correlation between touch input and gaze has, to our knowledge, not been studied yet.

In this work, we investigate how tapping correlates with gaze on a tablet device. We conducted a study with 24 participants and collected data related to touch input, gaze and tapped targets. Our study focused on Internet tasks, as browsing is a typical task widely used as study context for measuring mouse-eye correlation [8] and commonly performed by tablet users.

Analysis of the data indicates the following results: (1) gaze precedes touch with similar spatial and temporal features as observed with the mouse, and (2) the distance kept between the gaze and the touch varies across users, and is influenced by the learning and anticipation effects of the tasks.

2 Related work

Before touch input modality became widespread, mouse served as a proxy for manual input on computers. Early work on gaze and mouse input correlation in Human-Computer Interaction can be found in [13], where Smith et al. studied the correlation in target selection tasks and found several patterns, and in [5], where Chen et al. found correlation patterns applied to web browsing tasks and an average correlation of 0.58. Liebling and Dumais [10] explored the correlation of eye and mouse in everyday computer work tasks. They confirmed the eyes lead the mouse, but nuanced this paradigm, indicating it occurs only two thirds of the time, as “[this] depends on the type of target and the familiarity with the application”. We contribute to the understanding of the correlation between manual input and gaze in a natural environment by studying touch input instead of the mouse.

Browsing Internet is a common activity on tablets. Literature related to Internet usage often focuses on search tasks and Search Engine Results Page (SERP). Search queries can be categorised as informational, navigational or transactional - navigational being less common than transactional, and even less than informational [4,12]. Search tasks usually consist of answering informational questions [7], which can be combined with navigational questions [8]. Another classic activity is on-line shopping, with or without instruction regarding the items to buy [2,6]. Instead of focusing our work on a particular type of Internet based activity, we selected several in order to reflect the tablet’s use in real life.

3 Study

3.1 Participants

We collected data from 24 participants (9 female, age $\mu=31.4$, $\sigma=11$). All of them were familiar with Internet browsing and experienced with tablets and touch devices (3.7 and 3.9 of average on a 5-point Likert scale). Some participants needed visual correction during the study (7 wore glasses, 3 wore contact lenses). All but one were right-handed.

3.2 Tasks

We prepared 3 tasks to cover different Internet related activities, while keeping similarities with other studies, and maintaining naturalness in the tablet’s usage.

The **search** task comprises 10 questions which participants were asked to answer, by finding the relevant information on Internet with the means of their choice. This task echoes a common tablet activity, as often found in research literature. We chose 5 *informational* and 5 *navigational* questions, inspired by similar research.

The choice of a **shopping** task is driven by the interaction with richer content website style¹ and with forms (therefore typing). We asked participants to

¹ We chose the website of a leading supermarket in the UK, Sainsbury’s.

simulate the purchase of at least 10 different items. This process required them to fill forms.

The **game** task (“Wikipedia game”) is thought of as a way to generate data from hyperlinks, in a non systematic manner, with deeper cognitive and reading demands. In Wikipedia, participants were asked to reach a specific article from a specific source article by only following internal hyperlinks. A basic description of the articles’ topic was provided for help. They could play 2 rounds.

3.3 Apparatus

We favoured tablets for the data collection over other touch devices because of their reasonable size, prevalence, and compatibility with eye trackers. We used a Microsoft Surface Pro 3 (2160×1440 pixels resolution). We chose the Tobii X2-60 eye-tracker (60Hz), designed for studies on smaller devices. It comes with a stand designed for interacting with the device without occluding the eye tracker. Figure 1 shows how the stand and eye-tracker were set for the data collection.

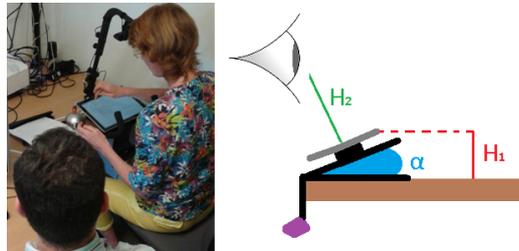


Fig. 1. Eye-tracker and stand configuration ($\alpha \approx 25^\circ$, $H_1 \approx 12$ cm and $H_2 \approx 40$ cm).

3.4 Data and Implementation

The data collection consisted of retrieving the following information: touch input, on-screen gaze position, and tapped object characteristics. Touch data is collected through different steps. Initially, we parse the display’s HID reports and send the touch time-stamp and coordinates to a Java application. The Java application interprets them as touch gestures, and records them onto log files. We wrote an application which retrieves gaze data samples from the eye-tracker, and writes in a log file their time-stamp, normalised on-screen position and validity code. The application also runs a 9-point calibration (22-pixel radius) before each task and a drift evaluation afterwards. We implemented a web browser (based upon Internet Explorer 11) in order to easily get feedback from it and offer a basic UI for all participants. The browser had a dimension of 1440×960 pixels², with a view port of 1440×914 pixels, topped by a navigation bar (Figure 2). The tapped objects have 3 distinct natures: HTML, browser or keyboard

² This is the dimension of the full screen in the non high DPI mode on the tablet.

element. HTML and browser targets are tracked via the browser which writes related information (time-stamp, nature, position and size) into a log file. We made a specific application to track keyboard elements and write related information (time-stamp and key code) into another log file. In a post-hoc step, we then merge these different log files with the taps log files into a single file, based on the time-stamps.



Fig. 2. Browser’s navigation bar part (truncated).

4 Results

4.1 General Results

We collected in total 574 675 touch data samples and 1 869 705 gaze data samples. We tracked 3 types of touch gesture actions: taps (72%), pans (28%) and zooms (<1%). The tablet was large enough for the participants to use comfortably (seldom zooms). The tapped object distribution is as follows: keyboard (68.9%), HTML (26.5%), browser (4.6%). Fixations are computed post-hoc with OGAMA³ on a spatial detection threshold of 22 pixels ($\sim 0.56^\circ$ of visual angle).

4.2 Similarity with Mouse Studies

We learnt from studies in psychology that, when pointing at objects, gaze leads the hand [1]. This is also verified in HCI studies with the mouse as manual input. We question how touch input compares with the mouse counterpart.

As we consider the **tapping** part of the whole target selection process, touch input can be assimilated with a punctual event in time and space. Similar to mouse/gaze studies, we check where gaze is located when a tap arises: the difference between tap and fixations positions around tap moment. We use the fixation’s start moment for the temporal dimension.

Figure 3(a) shows gaze approaches the selected target *before* the tap is actually performed. This estimation is obtained by using a Generalized Additive Model (cubic spline). We estimate gaze precedes the tap by 0.337s, within 197 pixels - similar to what is reported in gaze/mouse correlation studies (about 0.3s and 160 pixels [3]). The spatial difference can be explained by the decrease of pointing precision with finger, and a target size generally greater than in [3].

³ <http://www.ogama.net/>

4.3 Influence of Participants, Tasks and Targets on the Correlation

In the previous section, we describe a coarse estimation for *all* participants, tasks and target types. But do they influence the correlation?

We estimate the relationship between touch input and gaze for each participant, task and tapped object type with the same model as in section 4.2. In order to compare them, we report the descriptive statistics related to the **minimum point** given by the estimation model.

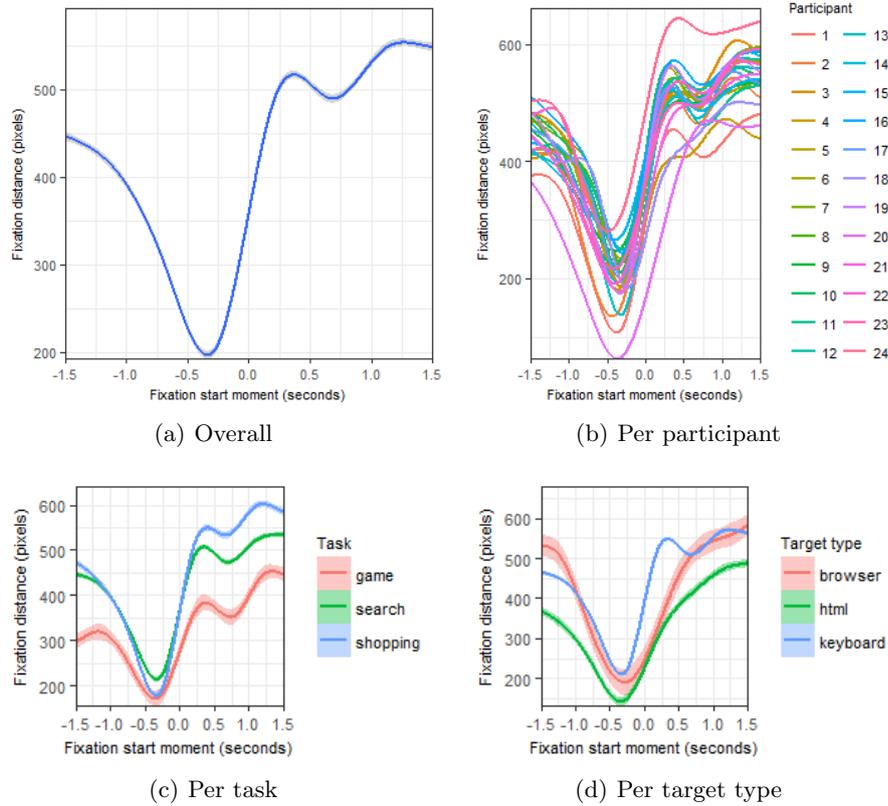


Fig. 3. Fixation start moment vs. fixation distance (relative to tap moment/position).

Estimations for each participant are plotted in Figure 3(b), and the minima's value reported in Table 1. For all participants, gaze leads the tap. However, disparity is observed between each participant. This difference is more important in space ($\Delta=218.8$ pixels, $\sigma=50.2$ pixels) than in time ($\Delta=249$ ms, $\sigma=57$ ms). We deduce that the distance within which a user keeps gaze away from the target is a personal feature, and influences the correlation.

Table 1. Fixation start moment vs. fixation distance (minima, per participant)

P ^(a)	S.M. ^(b)	Dist. ^(c)	P ^(a)	S.M. ^(b)	Dist. ^(c)	P ^(a)	S.M. ^(b)	Dist. ^(c)
#1	-377	106.9	#9	-402	246.4	#17	-325	200.1
#2	-432	135.1	#10	-333	209.9	#18	-24	183.6
#3	-324	185.8	#11	-411	221.5	#19	-339	173.9
#4	-291	184.3	#12	-304	249.2	#20	-368	62.2
#5	-365	181.3	#13	-323	137.1	#21	-347	176.9
#6	-314	234	#14	-363	192.6	#22	-405	191.2
#7	-274	206	#15	-408	265.5	#23	-417	216.1
#8	-313	226.8	#16	-305	243.4	#24	-489	281

^(a)Participant ^(b)Start moment (ms) ^(c)Distance (pixels)

Estimations for each task are plotted in Figure 3(c) and the minima’s value reported in Table 2 (left part). Again, we observe that gaze precedes touch, and that tasks influence the spatial dimension ($\Delta=43.4$ pixels, $\sigma=23.3$ pixels) rather than the temporal dimension ($\Delta=17$ ms, $\sigma=9$ ms). For the search task, the minimum has a greater Y-axis value than for the other two tasks. We can interpret this as a consequence of SERPs being systematically queried by the participants for that task. Users mainly follow the first link after scanning a few [9]. Thus, taps are possibly already “prepared” to be performed while users still scan the page, and then tap without a need to acquire the target again. The task nature has therefore a clear influence on the correlation.

Estimations for each tapped object type are plotted in Figure 3(d) and the minima’s value reported in Table 2 (right part). Gaze still precedes touch, and the temporal difference varies less than in space (respectively $\Delta=57$ ms, $\sigma=3$ ms; $\Delta=68.1$ pixels, $\sigma=34.94$ pixels). Fixations for the keyboard and browser objects seem to happen earlier before the touch, with a farther distance to the target. We explain this by the potential learning effect in typing and using the browser. Participants take less time “searching” the target, and do not need to visually focus on it as they know where it is. Thus the tapped target has a role in the correlation, depending on its likelihood to be known in advance.

Table 2. Fixation start moment vs. fixation distance (minima, per task/target type)

Task	S.M. ^(a)	Dist. ^(b)	Target type	S.M. ^(a)	Dist. ^(b)
Search	-335	214.3	Keyboard	-332	210
Shopping	-339	177.9	HTML	-343	141.9
Game	-352	170.9	Browser	-286	189.5

^(a)Start moment (ms) ^(b)Distance (pixels)

4.4 A particular fixation: $F_{Closest}$

We focus on the description of the specific fixation, $F_{Closest}$, that arises *before* the tap moment *at the closest* to the tap position. Describing $F_{Closest}$ confronts the estimations given earlier and may serve as a baseline for further studies.

For each tap, we retrieve $F_{Closest}$ in a window starting -0.6s before the tap moment⁴.

Figures 4(a) and 4(b) respectively show the histograms of $F_{Closest}$'s start moment (to the tap moment) and distance (to the tap position), and the associated quartiles values. The modes are -0.313s for the start moment and 34 pixels for the distance. No correlation was found for the HTML targets between their size and $F_{Closest}$'s distance/time.

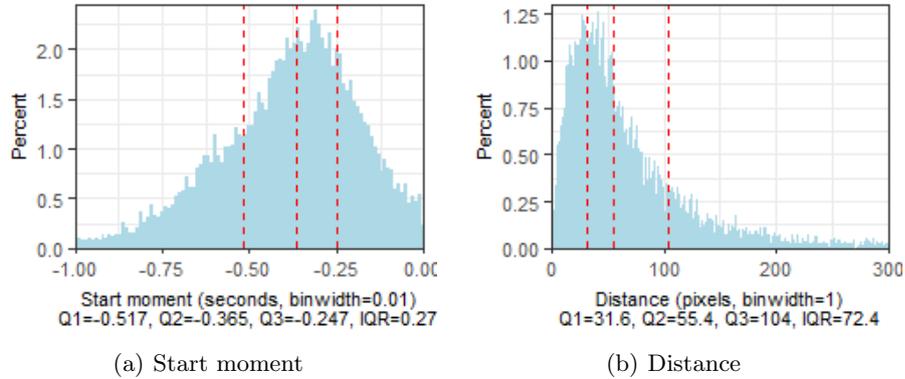


Fig. 4. $F_{Closest}$'s start moment / distance histograms and quartiles.

We study the spatial distribution of $F_{Closest}$ around the touch points, plotted in Figure 5(a). The standard deviation on the X-axis is 205.4 pixels, and 116.4 pixels on the Y-axis. Both mean and median positions show an offset which can be explained: web users tend to look more at the top-left part of the page [11].

Figure 5(b) represents $F_{Closest}$'s mean and median positions of each participant. Although they are spread around the global mean and median positions, they remain generally offset towards the top-left direction, most notably for the mean positions (92% of participants, against 67% for the median positions).

Figures 5(c) illustrates $F_{Closest}$'s mean and median positions for each task. For both search and shopping tasks, we notice the mean and median positions do not vary more than 8 pixels around the overall values in each direction. For the game task, we observe that $F_{Closest}$'s mean and median positions are closer to the tap point. In this task, the targets' position (mostly links) cannot be "learnt"

⁴ Value based on the minimum point reported in section 4.2 and the difference within participants in 4.3 ($-0.337 - 0.249 \approx -0.6$).

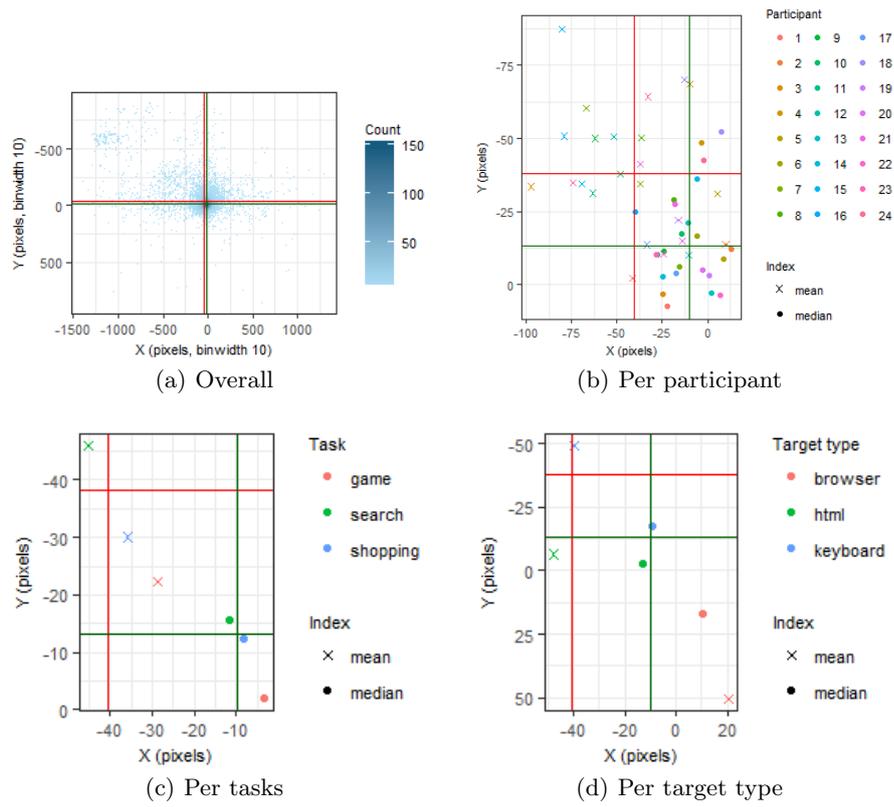


Fig. 5. $F_{Closest}$'s mean and median positions around tap position (red intersection: overall mean (-40.5,-38.1) pixels; green intersection: overall median (-9.8,-13.1) pixels).

nor anticipated, contrary to the other tasks. For the search task, learning effect of selecting the first link(s) in the SERP, or a tap anticipation as discussed in the section 4.3 can appear. For the shopping task, learning effect may come from the commercial website interaction. We suppose the learning effect and the tap anticipation brought by a task can influence the distance between gaze and tap: when there are none of these effects, gaze acquires targets with a closer distance.

$F_{Closest}$'s mean and median positions for each tapped object type are represented in Figure 5(d). There is an expected difference for the browser elements. Being situated in the top part of the screen, $F_{Closest}$'s mean and median positions are not likely to show an offset on the top-left side of the screen. Browser elements allow navigation and trigger changes in the view-port situated below, hence an opposite offset direction. The case of the HTML elements shows a very small vertical offset (less than 7 pixels for both mean and median values) indicating that gaze is more often vertically aligned with the targets. We can interpret this result as an effect of reading: most HTML targets are links ($\approx 40\%$) and text input fields ($\approx 26\%$).

5 Discussion

Our results give a coarse description of the correlation between touch and gaze in a natural context. Understanding this correlation can lead the way to improved gaze estimation methods, based on touch, notably, for eye tracking calibration. Even finer results were limited by our study design. We did not consider touch within the whole web page content, and therefore cannot understand how elements other than the tapped one influence the correlation. Also, preserving the naturalness of the tasks in the study led to an heterogeneous data set: participants tapped on different elements since they were free to browse at will to complete tasks.

6 Conclusion

We have devised a data collection in order to study the correlation between gaze and touch input on tablets. To reflect the naturalness of the tablet usage, and align our study with other works, we designed 3 Internet tasks as context. Our work confirms similar results found in previous gaze/mouse correlation studies: gaze precedes the touch by 0.337s, within 197 pixels. We described a specific fixation we coined $F_{Closest}$ (the fixation that is the closest from the target position, before the tap actually happens). We show that $F_{Closest}$'s median position can act as a gaze estimator, especially for a task which generates few learning and tap anticipation effects. Our study only focused on the tapping part of the gesture movement. Our future work will investigate a wider scope in order to cover the complete touch gestural mechanism.

References

1. Abrams, R.A., Meyer, D.E., Kornblum, S.: Eye-hand coordination: Oculomotor control in rapid aimed limb movements. (1990)
2. Atterer, R., Schmidt, A.: Tracking the Interaction of Users with AJAX Applications for Usability Testing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1347–1350. CHI '07, ACM, New York, NY, USA (2007), <http://doi.acm.org/10.1145/1240624.1240828>
3. Bieg, H.J., Chuang, L.L., Fleming, R.W., Reiterer, H., Bülthoff, H.H.: Eye and pointer coordination in search and selection tasks. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. pp. 89–92. ETRA '10, ACM, New York, NY, USA (2010), <http://doi.acm.org/10.1145/1743666.1743688>
4. Broder, A.: A Taxonomy of Web Search. SIGIR Forum 36(2), 3–10 (Sep 2002), <http://doi.acm.org/10.1145/792550.792552>
5. Chen, M.C., Anderson, J.R., Sohn, M.H.: What Can a Mouse Cursor Tell Us More?: Correlation of Eye/Mouse Movements on Web Browsing. In: CHI '01 Extended Abstracts on Human Factors in Computing Systems. pp. 281–282. CHI EA '01, ACM, New York, NY, USA (2001), <http://doi.acm.org/10.1145/634067.634234>
6. Chudá, D., Krátky, P.: Usage of Computer Mouse Characteristics for Identification in Web Browsing. In: Proceedings of the 15th International Conference on Computer Systems and Technologies. pp. 218–225. CompSysTech '14, ACM, New York, NY, USA (2014), <http://doi.acm.org/10.1145/2659532.2659645>
7. Guo, Q., Jin, H., Lagun, D., Yuan, S., Agichtein, E.: Towards Estimating Web Search Result Relevance from Touch Interactions on Mobile Devices. In: CHI '13 Extended Abstracts on Human Factors in Computing Systems. pp. 1821–1826. CHI EA '13, ACM, New York, NY, USA (2013), <http://doi.acm.org/10.1145/2468356.2468683>
8. Huang, J., White, R., Buscher, G.: User See, User Point: Gaze and Cursor Alignment in Web Search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1341–1350. CHI '12, ACM, New York, NY, USA (2012), <http://doi.acm.org/10.1145/2207676.2208591>
9. Jansen, B.J., Spink, A.: Web Mining. IGI Global (jan 2005), <http://www.igi-global.com/chapter/analysis-document-viewing-patterns-web/31146>
10. Liebling, D.J., Dumais, S.T.: Gaze and mouse coordination in everyday work. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. pp. 1141–1150. UbiComp '14 Adjunct, ACM, New York, NY, USA (2014), <http://doi.acm.org/10.1145/2638728.2641692>
11. Nielsen, J.: Horizontal attention leans left (2010), <https://www.nngroup.com/articles/horizontal-attention-leans-left/>
12. Rose, D.E., Levinson, D.: Understanding User Goals in Web Search. In: Proceedings of the 13th International Conference on World Wide Web. pp. 13–19. WWW '04, ACM, New York, NY, USA (2004), <http://doi.acm.org/10.1145/988672.988675>
13. Smith, B.A., Ho, J., Ark, W., Zhai, S.: Hand eye coordination patterns in target selection. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications. pp. 117–122. ETRA '00, ACM, New York, NY, USA (2000), <http://doi.acm.org/10.1145/355017.355041>