Gaze-based Mode-Switching to Enhance Interaction with Menus on Tablets

Yanfei Hu Fleischhauer
LMU Munich
Munich, Germany
yanfei.hu@campus.lmu.de

Hemant Surale
University of Waterloo
Waterloo, Canada
hemant.surale@gmail.com

Florian Alt
florian.alt@unibw.de
Bundeswehr University Munich
Munich, Germany

Ken Pfeuffer
Aarhus University
Aarhus, Denmark
Bundeswehr University Munich
Munich, Germany
ken@cs.au.dk

ABSTRACT

In design work, a common task is the interaction with menus to change the drawing mode. Done frequently, this can become a tedious and fatiguing task, especially for tablets where users physically employ a stylus or finger touch. As our eyes are naturally involved in visual search and acquisition of desired menu items, we propose gaze to shortcut the physical movement. We investigate gaze-based mode-switching for menus in tablets by a novel mode-switching methodology, assessing a gaze-only (dwell-time) and multimodal (gaze and tap) technique, compared to hand-based interaction. The results suggest that users can efficiently alternate between manual and eye input when interacting with the menu; both gaze-based techniques have lower physical demand and individual speed-error trade-offs. This led to a novel technique that substantially reduces time by unifying mode-selection and mode-application. Our work points to new roles for our eyes to efficiently short-cut menu actions during the workflow.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Pointing**; **Touch screens**.

KEYWORDS

gaze, touch, pen, mode switching, menu interface, tablet, mobile device

ACM Reference Format:
Yanfei Hu Fleischhauer, Hemant Surale, Florian Alt, and Ken Pfeuffer. 2023. Gaze-based Mode-Switching to Enhance Interaction with Menus on Tablets. In *2023 Symposium on Eye Tracking Research and Applications (ETRA '23), May 30-June 2, 2023, Tubingen, Germany*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3588015.3588409

1 INTRODUCTION

Direct manipulation using pens and touch-operated devices is prevalent in Computer-Aided Design (CAD) [Hinckley et al. 2010; Kim et al. 2019; Pfeuffer et al. 2021; Srinivasan et al. 2020]. In such productivity-focused apps, switching between modes is a common operation, to modify the design using various commands in a menu. A mode-switching process has three steps: (1) a visual search to capture the desired menu item, (2) a hand movement to point at and select the desired menu item, and (3) a hand movement to return to the main workspace. Frequent mode-switching requires significant hand movement and accrues to physical fatigue. Also, users must abandon their current position on the screen before engaging in the menu, which might divert attention from the main task and decrease productivity.

We investigate gaze input to enhance mode-switching for selecting menu items using a modern eye-tracking solution for pen- and touch-operated devices. The user can glance over the desired menu item to switch the application mode at any time during their typical workflow. The potential gain is to skip many lengthy physical roundtrips of a hand to select a menu item from a menu bar near the tablet's bezel, rendering CAD tools for tablets easier to use.

We focus on mode-switching in the pragmatic use of user interfaces (UI), where users operate the UI with pen and touch but at times can engage their eyes to switch between multiple modes. As such, this involves the cognitive and motor control costs for users to effectively interweave hand-operated and eye-operated tasks. Prior studies in gaze HCI have compared gaze-based interaction techniques to manual baselines in isolated selection tasks [Cheng et al. 2017; Kytö et al. 2018; Pfeuffer and Gellersen 2016; Sibert and Jacob 2000; Wagner et al. 2023; Ware and Mikaelian 1986; Zhai et al. 1999], without considering the cost of mode-switching during the live workflow. In this paper, we employ the *Subtraction method* [Dillon et al. 1990] that allows us to precisely isolate the mode-switching cost within the application's workflow. This has been demonstrated in prior studies of hand-controlled input devices [Li et al. 2005; Surale 2020; Surale et al. 2017], and we extend it to evaluate mode-switching performance between three modes.

We present an empirical investigation of a main (N=18) and a follow-up experiment (N=6) of gaze-based mode-switching in

---

ETRA '23, May 30-June 2, 2023, Tubingen, Germany

ACM ISBN 979-8-4007-0150-4/23/05…$15.00
https://doi.org/10.1145/3588015.3588409

tablets. Two established gaze-based interaction techniques are evaluated. *DWELL* where the user fixates on the menu item for 0.5 s [Jacob 1990], and multimodal *GAZETAP* where the user looks at the target and taps by hand at the current physical position to select it (e.g., as in [Pfeuffer et al. 2014; Stellmach and Dachselt 2012]). In the follow-up experiment, we evaluate an extended *GAZETAP** technique that unifies mode-switch and -application. Our findings are:

- Both the *GAZETAP* and *DWELL* techniques successfully eliminate the physical effort, as all users are adept at integrating gaze for quick access to menu mode-switches, during hand-controlled drawing tasks in all study trials.
- Reduction of physical movement comes at a trade-off to higher menu selection time. Times are (from short to long): 1003 ms (manual baseline), 1236 ms (*DWELL*), 1246 ms (*GAZETAP*).
- Hand-fatigue was rated lower for both gaze techniques, whereas eye fatigue was reported higher for *DWELL*.
- Most users prefer a gaze-based technique (9 votes for *DWELL*, 7 *GAZETAP*, 2 hand baseline).
- Our follow-up experiment showed that *GAZETAP** brings a major time save (519 ms), indicating a new potential to significantly speed up mode-switching actions.

The contributions are: (1) the first investigation exploring the mode-switching phenomenon between manual and gaze input, (2) empirical evidence of users being efficient at switching between gaze and hand modalities, gaze being useful to save manual effort, and users prefer gaze-based methods over the most familiar baseline methods, and (3) a new technique that unifies mode-switch and application for major time saves, indicating promising potential for techniques to become useful in design-focused application.

2 RELATED WORK

Although user interface layouts have undergone innumerable enhancements over time, the division between the main program area and the auxiliary menu items suggests an attention dilemma: "*A user's focus of attention must constantly change from some point on the artwork to a UI widget at the edge of the screen and then refocus on the artwork again*" (Kurtenbach et al. [Kurtenbach et al. 1997]). Early methods to address this include spring-loaded modes [Li et al. 2005], Toolglass [Bier et al. 1993], Marking Menus [Kurtenbach and Buxton 1994], context menus [con 2021], and hybrid menus with multiple modes that can be invoked at the cursor position to avoid roundtrips [Kurtenbach et al. 1997; Lepinski et al. 2010]. Further, the menu UI can be designed spatially near to the hand that holds a tablet to reduce mode-switching costs [Pfeuffer et al. 2017; Zhang et al. 2019]. While these methods reduce the round-trips by quick access to the most frequent commands, our focus is on the conventional menus that are still frequently used to access other commands placed away from the central workspace, especially in design applications.

Target acquisition typically necessitates the user first looking at the target before engaging cursor control, especially when the user does not have information about the location of the target [Jacob 1991]. This represents a unique opportunity to support, automate, and extend the UI with gaze controls [Pfeuffer et al. 2014; Zhai et al. 1999]. Gaze as a pointing device has been investigated as an alternative to manual pointing mechanisms of a mouse [Lutteroth et al. 2015; Sibert and Jacob 2000; Zhai et al. 1999] touchscreen input [Pfeuffer et al. 2016; Pfeuffer and Gellersen 2016; Stellmach and Dachselt 2012], 3D hand gestures [Lystbæk et al. 2022a,b; Wagner et al. 2023], or head gesture [Sidenmark et al. 2020], and leading researchers in principle see it as a natural, convenient, and fast input medium since the 80s-90s [Jacob 1991; Sibert and Jacob 2000; Ware and Mikaelian 1986; Zhai et al. 1999]. Multi-modal techniques were developed that combine multiple modalities such as gaze, hand, and head for selection tasks [Kytö et al. 2018; Pfeuffer and Gellersen 2016; Sidenmark et al. 2023; Zhai et al. 1999]. These are considered as stand-alone techniques, and evaluated in contrast to fully manual or eye based techniques, without considering switching costs.

However, an important question is how gaze can complement the manual UI – to be potentially integrated in the many interactions we do today. A subset of gaze HCI work proposed gaze for context switching between large areas of interest such as displays and applications [Bolt 1981; Isokoski 2000; Morimoto and Amir 2010; Salvucci and Anderson 2000; Tula et al. 2012]. More fine grained control mechanisms have been explored by Pfeuffer et al., where gaze is used as complement to hand-based UI controls [Pfeuffer et al. 2014, 2015; Pfeuffer and Gellersen 2016]. Gaze-Touch [Pfeuffer et al. 2014] for example integrates within the touch UI, to enable users to look up to a specific menu item and touch-tap anywhere on the screen to select it – a temporal activation of gaze controls. This principle has been extended to both pen and touch devices and a variety of UI components and tools for CAD [Pfeuffer et al. 2015]. Rivu et al. investigated how eye-gaze can be used only for specific UI elements, such as a button [Rivu et al. 2019] and text fields [Rivu et al. 2020]. Elmadjian and Morimoto's GazeBar [Elmadjian and Morimoto 2021] is an advanced menu where the user's gaze input selects and navigates the menu, to be used complementary to mouse-based interaction in the main part of the application. Although prior research has underlined the potential of using gaze input to enhance frequent mode-switching, there is a lack of empirical evidence to validate its efficacy when coupled with manual input.

To assess the performance improvements when alternating between gaze and manual inputs, a requisite is to capture mode-switching cost. Modes are "a functioning arrangement or condition" [MacKenzie 2012], and mode switching is necessary to access them. The 'Subtraction Method' is a measurement of the mode-switching time [Dillon et al. 1990; Donders 1969]. Researchers have adopted it to investigate the mode-switching phenomenon for mouse and trackball [Kabbash et al. 1994], pen [Hinckley et al. 2006; Li et al. 2005], touch [Surale et al. 2017], and mid-air virtual reality UI [Surale et al. 2019]. Most of these past works shared a common goal — they investigated the performance of input techniques when alternating between only two modes (e.g., alternating two line colors). Notably, Dillon et al. [Dillon et al. 1990] pointed out that the smooth integration of the selection methods (mode-switching techniques) aligned with the user's workflow is more important than the selection performance of the individual method. So, to our knowledge, we are the first to use the subtraction method to the practical application of mode-switching workflow, which requires using several modes, and to the evaluation of multi-modal gaze technique in comparison to manual baseline techniques.

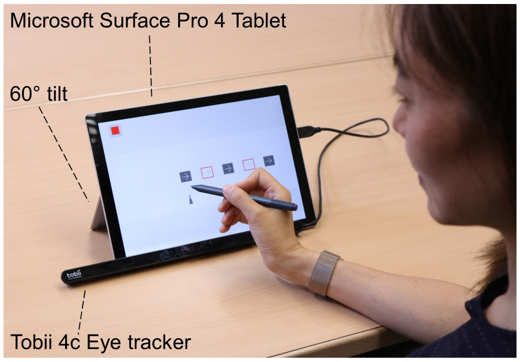

Figure 1: Study setup with a tablet, and eye tracker on a table.

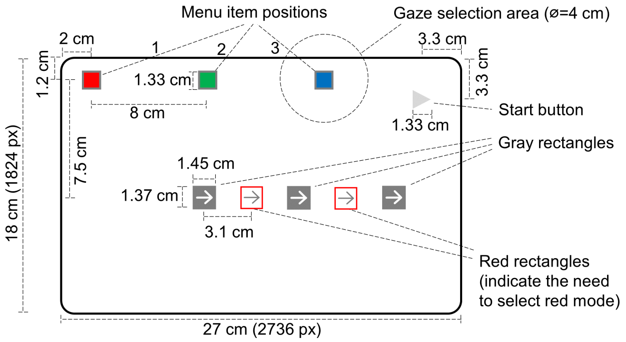

Figure 2: Task layout, including crossing targets (mid) and mode-switches (top)

3 FIRST EXPERIMENT: GAZE-BASED MODE-SWITCHING

We evaluate gaze-based mode-switching techniques in comparison with manual baselines on tablets (Figure 1).

3.1 Experimental Protocol and Task

Our experimental task is derived from Dillon et al. [Dillon et al. 1990] and Surale et al. [Surale et al. 2017]. The "Subtraction Method" isolates the menu selection time by subtracting the mean time of performing a series of tasks using a single or no mode from the mean time taken to perform the same series of tasks with the intermediate menu selection task.

Our experimental task is crossing rectangles. The users alternate between two types of blocks—a baseline block without a menu selection and a compound block with two menu item selections. Both block types include five rectangle crossings. In a compound block, crossing rectangles 2 and 4 involves selecting a corresponding menu item before crossing. Crossing rectangles 1, 3, and 5 does not require selecting a menu item. Figure 2 shows the rectangles and menu item placements. The user crosses the rectangles from left to right. An arrow at the rectangle shows the crossing direction for each rectangle. Note that these three menu items function as a representation of multiple modes in real UIs.

Figure 3 illustrates an example compound block operated by a pen. First, the user starts the block by crossing the first gray rectangle, which disappears after a successful crossing. Then, the first cycle begins. The cycle involves selecting the mode, crossing the second crossing target, and the third crossing target. Afterward, the second cycle begins for the 4th and 5th targets, which are similar in procedure to the first cycle. After a block is finished, pressing a start button on the right side of the tablet screen will start the next block. The menu items and crossing targets are colour-coded to convey the menu item position when the user sees the crossing target.

3.2 Study Design and Independent Variables

We conduct a within-subjects repeated-measures study. The main independent variable is *technique* where we compare three techniques. The first technique is ***DWELL***, where a time threshold allows to confirm the selection of the looked target [Bolt 1981; Jacob 1990; Sibert and Jacob 2000]. Prior work used time thresholds in the range of 150ms to 2000ms [Bernardos et al. 2016; Jacob 1990; Majaranta et al. 2009; Miniotas et al. 2006; Sibert and Jacob 2000], of which we chose 500ms as a good fit to view and select targets. Visual feedback indicates interaction states, of idle mode (Figure 4-1), when viewed with the eyes (2), and when 500 ms are over and selection confirmed (3).

The second technique is ***GAZETAP***, as a multimodal technique for selection with gaze and confirmation by pen or touch tap [Pfeuffer et al. 2014, 2015; Stellmach and Dachselt 2012]. The user interacts manually on the main canvas (Figure 5-1), then looks up and fixates on the desired item in the menu (2), visually indicated by item border highlighting. A tap with the manual input device confirms the selection (3), and then users return to their main workspace continue with the manual work.

Third, a **fully-hand-based baseline** where users move the finger or pen to the menu item for selection, which represents the default behaviour in contemporary application.

Notably, all three techniques are repeated in two variants: (a) one where users use finger touch for all manual tasks, and (b) one where users employ a stylus. The order of the 3×2 conditions is counterbalanced by a Latin square. At each condition, users performed all blocks. Block type was counterbalanced and menu item positions was randomised.

Further factors vary the task environment (Figure 2). Menu item targets (red, green, blue) cover three *positions* at the top of the UI area. i.e., a typical placement of menus, toolbars, and ribbons. Crossing *directions* (up, down, left, right) are consistent per block and randomized across *blocks*. In sum there were: 3 techniques × 2 devices × 3 positions × 4 directions × 2 block types (1 baseline, 1 compound) × 5 rectangle crossing = 720 rectangle crossings per participant.

3.3 Setup and Apparatus

The software was developed using Java with Processing (v3.0)[1] on Microsoft Surface Pro 4 tablet (12.3", 2736×1824px, 267pixels-per-inch) with a Surface Pro Pen and a Tobii 4c eye tracker placed at the bottom of it (Figure 1). The participants were seated about 60cm in front of the screen throughout the study. At the beginning of

---

[1] Processing, URL: https://processing.org/, accessed 3/16/23

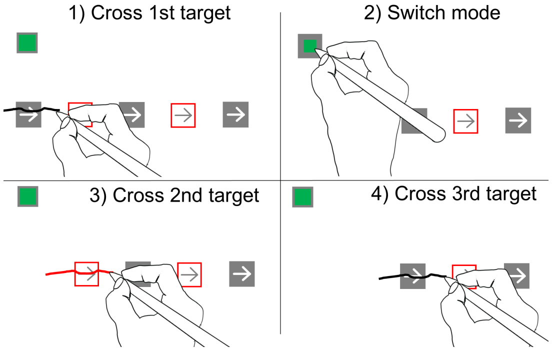

Figure 3: Compound block example for crossings 1-3 (4-5 are similar to 2-3).

the study, we calibrated each participant once with the eye tracker resulting in a successful calibration without errors for all users without canceling any. The touch or pen-down events can select the menu item in all conditions except for the dwell technique, where the temporal thresholds would invoke selection. In line with prior recommendations [Feit et al. 2017], we set a sizeable 4×4cm invisible selection area around the target for the gaze techniques to tackle eye-tracker inaccuracies. Our own post-hoc analysis where we measured the distance of gaze to the menu item position at the time of its selection, showed the gaze techniques have an accuracy of about 1 cm (1cm ≈0.954° of visual angle at a screen-eye distance of 60 cm). To address potential data loss from occluding the eye tracker by the hands, we considered the device's three eye tracking sensors on the left, center, and right parts. No data is available when the center sensor is occluded. We adapted our task design by moving the rectangle crossing to the right side of the tablet screen (Figure 1). While all the participants were right-handed, the task orientation could switch to accommodate the handedness.

3.4 Evaluation Metrics

Menu selection time is calculated by subtracting the mean duration of two time cycles of a baseline block from the mean duration of two time cycles of a compound block, as per the Subtraction Method [Li et al. 2005; Surale et al. 2017]. The first cycle starts after the first crossing (i.e., the pen/finger is lifted) and ends after the third. The second cycle ranges from the 3rd and ends after the 5th crossing. We analyze **three error types** [Li et al. 2005; Surale et al. 2017]: *Crossing error*, when participants do not correctly cross the crossing target, e.g., the stroke is drawn in the wrong direction or starts/ends inside of the target; *Out of target error*, if the drawn stroke did not intersect with the crossing target; *Mode error*, when crossing the target without selecting the mode before (only possible in compound block). Participants provided ratings for each technique, between 1 (strongly disagree) and 5 (strongly agree), on a **usability questionnaire**, which included six statements: *´´The task was [easy to learn | easy to use | accurate | fast | comfortable for the eyes | comfortable for the hands]"*.

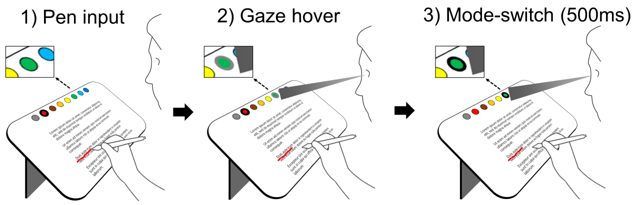

Figure 4: Dwell-Time technique.

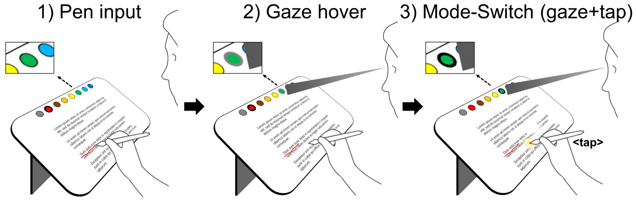

Figure 5: Gaze-Tap technique.

3.5 Procedure

Following a briefing, each participant completed consent and demographics forms. The participant completed training of around 30 seconds that spanned between three to five trials to reach the necessary proficiency level, of clearly understanding the efficient usage, before each technique.The learning effects analysis where we analysed time across blocks via an ANOVA confirmed the training was adequate as no significant results were reported. In total, each participant completed 24 blocks in a session. Despite recommending breaks between blocks, most participants finished the session without them and no fatigue was reported for the overall experiment. Participants answered a custom usability questionnaire after finishing each technique and submitted a ranking in the final questionnaire after completing all techniques. A brief interview, at last, concluded the study. The entire study took approximately 45 to 50 minutes.

3.6 Participants

We recruited 18 paid participants (9 female, 9 male, all right-handed) using the university mailing lists and approaching potential participants within the institute. They were aged between 20 and 56 (M=27.97, SD=8.16). On a scale between 1 (no experience) to 5 (expert), participants rated themselves as moderately experienced with tablets (M=3.77, SD=1.1), and less experienced with gaze (M=2.37, SD=1.45) and a pen (M=1.9, SD=1.18) input.

3.7 Results

A trial was treated as an outlier if the task duration was more than 3SD from the mean (of task completion time for each condition for each participant). Crossings that belong to the same cycle as an outlier were also removed, leading to the removal of 2.6% of the total crossings (1.9% to 3.4% per technique). For menu selection time analysis, error tasks were removed together with the tasks that belong to the same cycle (5.2%). No learning effects were indicated by analysing technique × block interaction on menu selection time and error rate, therefore all blocks are used in subsequent analysis.

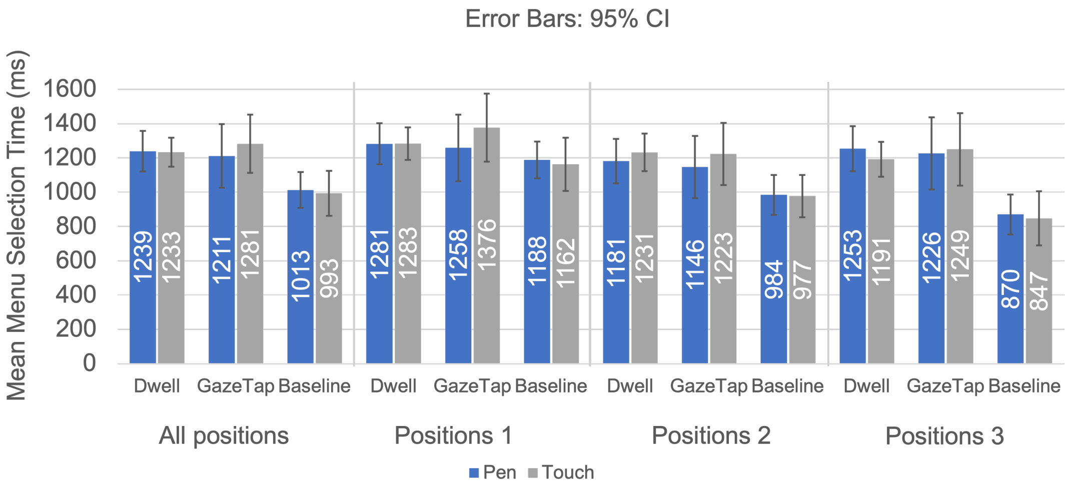

Figure 6: Menu selection time results.

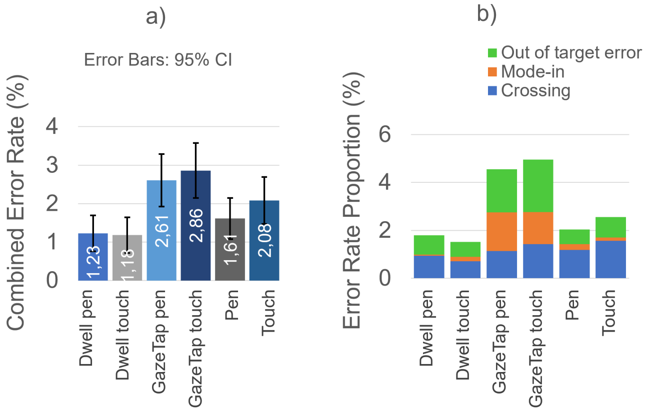

Figure 7: Error rate results.

Table 1: Results on the subjective feedback on six usability categories (mean and SD).

Rating	dwellPen	DwellTouch	GazeTapPen	GazeTapTouch	Pen	Touch
Easy to learn	4.4±0.8	4.4±0.9	3.9±0.9	3.8±1.1	4.8±0.4	4.8±0.4
Easy of Use	4.3 ±0.9	4.3±0.8	3.7±1.0	3.4±0.9	4.5±0.8	4.2±0.9
Accuracy	4.2±1.0	4.1±0.8	3.7±1.0	3.7±0.9	4.3±0.8	3.9±1.0
Speed	4.0±1.0	4.1±0.8	3.4±1.0	3.4±1.0	4.1±1.0	4.0±0.9
Eye Comfort	3.0±1.0	3.3±1.3	3.1±1.0	3.1±1.2	3.9±0.8	3.9±0.8
Hand Comfort	4.2±0.9	3.2±1.2	3.7±0.9	2.9±1.0	3.1±1.1	2.3±1.0
Combined	**4.0±0.7**	**3.9±0.7**	**3.6±0.6**	**3.4±0.6**	**4.1±0.5**	**3.9±0.5**

We performed repeated measures analysis of variance (ANOVA) with Greenhouse-Geisser corrections and post-hoc pairwise comparisons (Bonferroni corrected). Shapiro-Wilk tests confirmed the normality of menu selection times. Friedman tests with Wilcoxon Signed Rank tests (Bonferroni corrected) are used for the error rates and the Likert scale ratings.

Figure 6 shows results on **menu selection time**. For technique ($F^{1.715}_{29.161}$=12.032, p<.001), we find that the baseline was significantly faster than *Dwell* (p<.001) and *GazeTap* (p<.006).

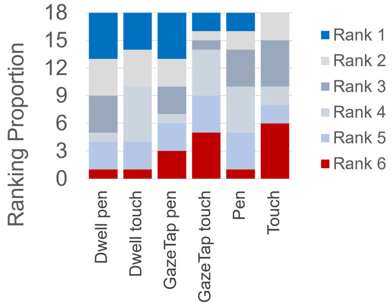

Figure 8: Ranking results.

Figure 7 shows results on **errors**. For *combined error* ($\chi^2(2)$ = 18.657, p<.001), we find that *Dwell* had significantly less errors than *GazeTap* (Z = -3.37, p<.003). For *crossing error* ($\chi^2(2)$ = 6.426, p<.041), *Dwell* had significantly fewer errors than the manual baseline (Z = -2.516, p = .036). And, for *out of target error* ($\chi^2(2)$ = 10.531, p<.006), there were less errors with *Dwell* (Z = -3, p<.009) and the baseline (Z = -2.987, p<.009) than *GazeTap*.

Table 1 shows **usability ratings**. *Ease of use* ($\chi^2(2)$ = 20.5, p<.001) and *speed* ($\chi^2(2)$ = 20.48, p<.001) were rated higher for *Dwell* (p = .002) and the manual baseline (p<.001) than *GazeTap*. For *learning* ($\chi^2(2)$ = 11.9, p<.003), users found the manual baseline (p<.004) was significantly easier to learn than *GazeTap*. *Eye fatigue* ($\chi^2(2)$ = 8.3, p<.016) for the manual baseline (p = .023) was significantly lower than *Dwell*, whereas *hand fatigue* ($\chi^2(2)$ = 18.2, p<.001) for *Dwell* (p<.001) and *GazeTap* (p = .039) was significantly lower than the manual baseline.

Figure 8 shows the **rankings**. Most-preferred ranking sorted by frequency were: *Dwell* & pen (5), *GazeTap* & pen (5), *Dwell* & touch (4), *GazeTap* & touch (2), pen (2), touch (0). The user feedback indicated that *Dwell* was preferred for ease of use and intuitiveness, P16 noted, *"Gaze dwell is very intuitive, as you automatically look at the colour square"*. Those who preferred hands-only expressed

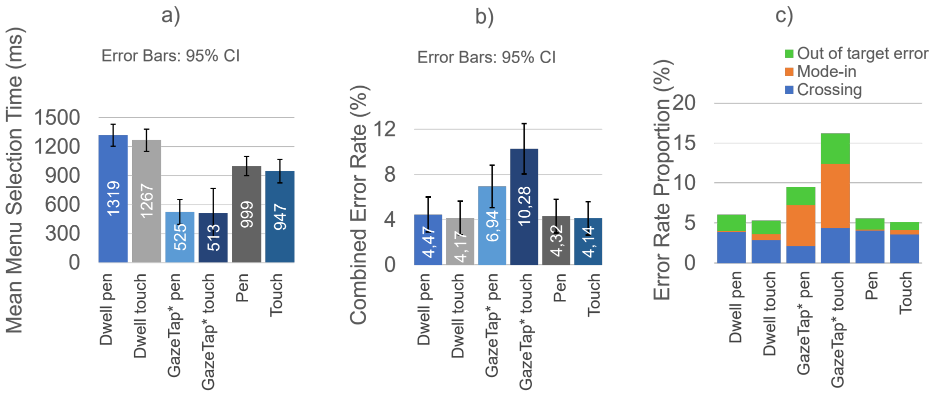

Figure 9: Time (a) and error (b, c) results of the second experiment (N=6) using an advanced *GAZETAP* technique (marked *) that unifies mode-switch and line-drawing tasks.

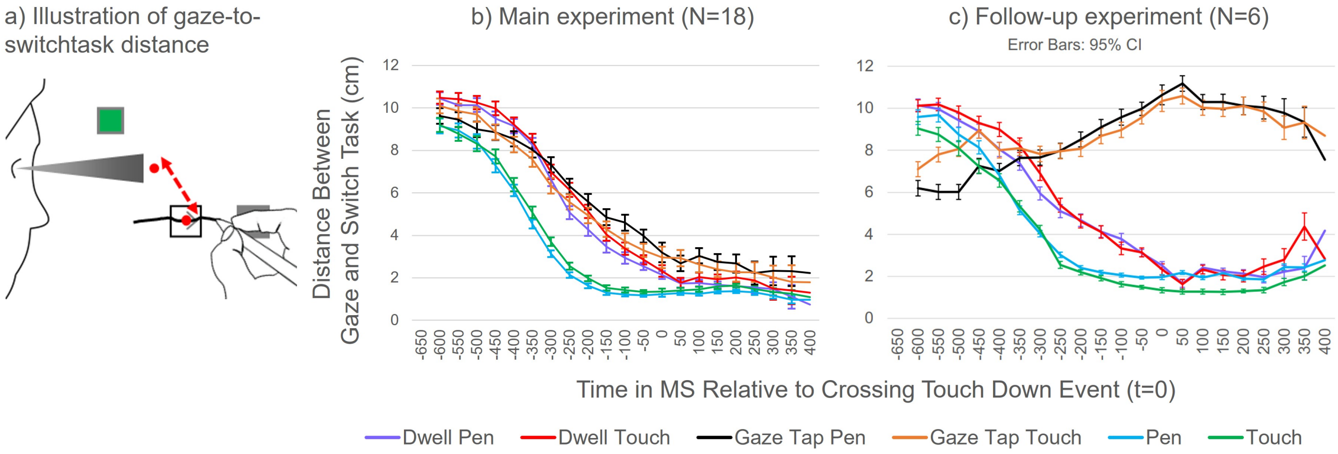

Figure 10: Spatio-temporal plots based on gaze-to-switchtask distance (a) showing that in contrast to the first study (b), the second study with the enhanced GazeTap* technique (c) leads to users draw the line eyes-free (as still focusing on the menu).

concerns regarding the synchronisation, *"I thought it is difficult to switch between two modalities"* (P9). Participants who preferred GAZETAP noted physical movement, comfort (e.g., P17: *"Changing the colour with the eyes and confirming with a touch makes it much more efficient with little hand movement, which makes it more comfortable, especially at the end of the task."*, as well as speed in contrast to DWELL (P17: *"When the colour is chosen only with the gaze, it takes more time, since it needs longer time to confirm it with the eye at 0.5 ms."*).

3.8 Summary

The experiment demonstrated that participants are proficient at carrying out brief gaze interactions between hand movements. Eliminating the physical roundtrip is the primary distinction between gaze and manual input; this can simplify input interactions. However, there is a trade-off between time and physical effort. *DWELL* and *GAZETAP* resulted in a longer switching time, but most participants preferred these techniques over the baseline at the end of the study. Contrasting both gaze-based interaction techniques, *DWELL* is more user-friendly and causes fewer mistakes than *GAZETAP*, and compared to the baseline techniques it causes greater eye fatigue. User preferences were divided between the two. Nonetheless, the result highlights the potential of gaze-based mode-switching techniques and an interesting question, can we make them time efficient?

4 FOLLOW-UP EXPERIMENT: UNIFYING MODE-SWITCHING AND APPLICATION

Led by the first experiment's results, we designed and implemented a novel technique to address the problem of a longer menu selection time for *GAZETAP*. We focus on *GAZETAP*, because enhancements to *DWELL* such as lower time thresholds and their time/error trade-offs are extensively studied. Our new technique, *GAZETAP**, extends gazetap. The operation involves two steps: (1) as in *GAZETAP*, users look at the menu item and then touch-down at the present screen location to switch the mode; (2) the new aspect is that users can immediately continue to touch-drag to draw a line without having to lift a pen or finger. As a result, it makes use of the input phrase chunking technique described by Buxton [Buxton 1995], in which the mode selection task from the menu and the line drawing task are combined into a single cognitively coherent phrase.

We conducted a follow-up experiment with 6 individuals to get first insights into the new technique. We use the same study design as the main study, with the difference that we use *GAZETAP** instead of *GAZETAP*. Results (Figure 9) show an advantage in menu selection time ($F^{1.42}_{7.1}$=47,p<.001) by 46.6% over the manual baseline and 59.8% over *DWELL* (both p<.001). This presents a vastly improved temporal performance and a highly efficient way to switch modes. As in the first experiment, the data indicated a trade-off with the error-rate ($\chi^2(2)$ = 6.3, p < .042), but no post-hoc significant differences were reported.

A unique aspect of *GAZETAP** is that users can, but do not necessarily have to, start to draw the crossing line eyes-free (as the eyes are fixated on the menu), which may contribute to the particular speed/error trade-off. To better understand this, we conducted a spatio-temporal analysis of the user's gaze behaviour over time (Figure 10). For the temporal dimension, we compute the gaze-to-switch task distance (a) across time intervals of [-650, 400] ms around the moment of touch-down event (t=0) of crossing a line. As a single gaze point can be erratic, we use the mean gaze point in the $[-25ms, 25ms]$ window around touch down. We find that in the first experiment (Figure 10b), the gaze is closely located to the crossing target during the start of line drawing with *GAZETAP*; consistent with the other techniques. With *GAZETAP** (c), the distance is much further apart. This indicates that users performed the task eyes-free, likely contributing to the distinct speed-error trade-off. As such, this technique might lend itself more for rapid and short ink tasks, or where the initiation of the stroke is less reliant on specific positioning.

5 LIMITATIONS

Our work has a few limitations that must be considered when interpreting our findings. Our studies were performed in a laboratory environment, and mobile tablet use cases can show a different picture, e.g., the hands might be more unstable and gaze-based shortcuts may become more practical, whereas technical eye-tracking precision might be affected in less controlled settings. We also tested a particular demographic background as students and employees of the local university, a target group; however, there are many others like elder users, where the remedy of physical effort might have more significant utility. Further, the second study was a quick follow-up study to demonstrate the speed benefits of a novel technique. While the results are encouraging, it has limited power due to a smaller sample size. At last, our techniques work with menus with a particular target size that is typical for eye-based interfaces but relatively large for standard menus, which could be addressed through nested menus [Pfeuffer et al. 2015].

6 CONCLUSION

This work explores a UI that offers gaze interaction to use in alternation with the hands of a user. Our main insight is users can efficiently alternate between manual and gaze input to perform rapid mode switches and menu selections. Users overall prefer to use a gaze-based mode-switch, which comes at individual speed-error trade-offs as the user shifts between hand and eye-based inputs. Our work paves the way for future UIs that incorporate gaze sensing without replacing the manual controls, in order to render menu actions by touch and stylus more efficient. Beyond that, we consider our work as first step of a grander vision where hand-controlled UIs are advanced through intelligent integration of our eyes, to generally render many frequent, repetitive, and tedious manual tasks easier and with it simplify the UI controls.

REFERENCES

2021. Context Menu. https://en.wikipedia.org/wiki/Context_menu.

Ana M Bernardos, David Gómez, and José R Casar. 2016. A comparison of head pose and deictic pointing interaction methods for smart environments. *International Journal of Human-Computer Interaction* 32, 4 (2016), 325–351.

Eric A Bier, Maureen C Stone, Ken Pier, William Buxton, and Tony D DeRose. 1993. Toolglass and magic lenses: the see-through interface. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques.* 73–80.

Richard A Bolt. 1981. Gaze-orchestrated dynamic windows. *ACM SIGGRAPH Computer Graphics* 15, 3 (1981), 109–119.

William Buxton. 1995. Chunking and phrasing and the design of human-computer dialogues. In *Readings in Human–Computer Interaction.* Elsevier, 494–499.

Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D Wilson. 2017. Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* 3718–3728.

Richard F Dillon, Jeff D Edey, and Jo W Tombaugh. 1990. Measuring the true cost of command selection: techniques and results. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 19–26.

Franciscus Cornelis Donders. 1969. On the speed of mental processes. *Acta psychologica* 30 (1969), 412–431.

Carlos Elmadjian and Carlos H Morimoto. 2021. GazeBar: Exploiting the Midas Touch in Gaze Interaction. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21).* Association for Computing Machinery, New York, NY, USA, Article 248, 7 pages. https://doi.org/10.1145/3411763.3451703

Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 Chi conference on human factors in computing systems.* 1118–1130.

Ken Hinckley, Francois Guimbretiere, Patrick Baudisch, Raman Sarin, Maneesh Agrawala, and Ed Cutrell. 2006. The springboard: multiple modes in one spring-loaded control. In *Proceedings of the SIGCHI conference on Human Factors in computing systems.* 181–190.

Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddington, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. 2010. Pen+ touch= new tools. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology.* 27–36.

Poika Isokoski. 2000. Text Input Methods for Eye Trackers Using Off-Screen Targets. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00).* Association for Computing Machinery, New York, NY, USA, 15–21. https://doi.org/10.1145/355017.355020

Robert JK Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 11–18.

Robert JK Jacob. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)* 9, 2 (1991), 152–169.

Paul Kabbash, William Buxton, and Abigail Sellen. 1994. Two-handed input in a compound task. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* 417–423.

Nam Wook Kim, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael J. McGuffin, and Hanspeter Pfister. 2019. DataToon: Drawing Dynamic Network Comics With Pen + Touch Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300335

Gordon Kurtenbach and William Buxton. 1994. User learning and performance with marking menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 258–264.

Gordon Kurtenbach, George Fitzmaurice, Thomas Baudel, and Bill Buxton. 1997. The design of a GUI paradigm based on tablets, two-hands, and transparency. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems.* 35–42.

Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. 2018. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in*

Computing Systems. 1–14.

G Julian Lepinski, Tovi Grossman, and George Fitzmaurice. 2010. The design and evaluation of multitouch marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2233–2242.

Yang Li, Ken Hinckley, Zhiwei Guan, and James A Landay. 2005. Experimental analysis of mode switching techniques in pen-based user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 461–470.

Christof Lutteroth, Moiz Penkar, and Gerald Weber. 2015. Gaze vs. Mouse: A Fast and Accurate Gaze-Only Click Alternative. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 385–394. https://doi.org/10.1145/2807442.2807461

Mathias N. Lystbæk, Ken Pfeuffer, Jens Emil Sloth Grønbæk, and Hans Gellersen. 2022a. Exploring Gaze for Assisting Freehand Selection-Based Text Entry in AR. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 141 (may 2022), 16 pages. https://doi.org/10.1145/3530882

Mathias N. Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022b. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 145 (may 2022), 18 pages. https://doi.org/10.1145/3530886

I Scott MacKenzie. 2012. *Human-computer interaction: An empirical research perspective.* Newnes.

Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 357–360.

Darius Miniotas, Oleg Špakov, Ivan Tugoy, and I Scott MacKenzie. 2006. Speech-augmented eye gaze interaction with small closely spaced targets. In *Proceedings of the 2006 symposium on Eye tracking research & applications*. 67–72.

Carlos H. Morimoto and Arnon Amir. 2010. Context Switching for Fast Key Selection in Text Entry Applications. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 271–274. https://doi.org/10.1145/1743666.1743730

Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 509–518.

Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 373–383.

Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2016. Partially-indirect bimanual input with gaze, pen, and touch for pan, zoom, and ink interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2845–2856.

Ken Pfeuffer, Abdullatif Dinc, Jan Obernolte, Radiah Rivu, Yasmeen Abdrabou, Franziska Shelter, Yomna Abdelrahman, and Florian Alt. 2021. Bi-3D: Bi-Manual Pen-and-Touch Interaction for 3D Manipulation on Tablets. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 149–161. https://doi.org/10.1145/3472749.3474741

Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan). Association for Computing Machinery, New York, NY, USA, 301–311.

Ken Pfeuffer, Ken Hinckley, Michel Pahud, and Bill Buxton. 2017. Thumb+ Pen Interaction on Tablets.. In *CHI*. 3254–3266.

Radiah Rivu, Yasmeen Abdrabou, Ken Pfeuffer, Mariam Hassib, and Florian Alt. 2020. Gaze'N'Touch: Enhancing Text Selection on Mobile Devices Using Gaze. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382802

Sheikh Rivu, Yasmeen Abdrabou, Thomas Mayer, Ken Pfeuffer, and Florian Alt. 2019. GazeButton: Enhancing Buttons with Eye Gaze Interactions. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research &; Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 73, 7 pages. https://doi.org/10.1145/3317956.3318154

Dario D Salvucci and John R Anderson. 2000. Intelligent gaze-added interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 273–280.

Linda E Sibert and Robert JK Jacob. 2000. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 281–288.

Ludwig Sidenmark, Christopher Clarke, Joshua Newn, Mathias N. Lystbæk, Ken Pfeuffer, and Hans Gellersen. 2023. Vergence Matching: Gaze Selection in 3D based on Modulation of Target Distance from the Eyes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580685

Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. 2020. BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/3379155.3391312

Arjun Srinivasan, Bongshin Lee, Nathalie Henry Riche, Steven M. Drucker, and Ken Hinckley. 2020. InChorus: Designing Consistent Multimodal Interactions for Data Visualization on Tablet Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376782

Sophie Stellmach and Raimund Dachselt. 2012. Look & touch: gaze-supported target acquisition. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2981–2990.

Hemant Surale. 2020. Barehand Mode Switching in Touch and Mid-Air Interfaces. (2020).

Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2017. Experimental analysis of mode switching techniques in touch-based user interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3267–3280.

Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental analysis of barehand mid-air mode-switching techniques in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

Antonio Diaz Tula, Filipe M. S. de Campos, and Carlos H. Morimoto. 2012. Dynamic Context Switching for Gaze Based Interaction. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 353–356. https://doi.org/10.1145/2168556.2168635

Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581423

Colin Ware and Harutune H Mikaelian. 1986. An evaluation of an eye tracker as a device for computer input2. In *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*. 183–188.

Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 246–253.

Yang Zhang, Michel Pahud, Christian Holz, Haijun Xia, Gierad Laput, Michael McGuffin, Xiao Tu, Andrew Mittereder, Fei Su, William Buxton, et al. 2019. Sensing posture-aware pen+ touch interaction on tablets. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.