Comparing Dwell time, Pursuits and Gaze Gestures for Gaze Interaction on Handheld Mobile Devices

Omar Namnakani
o.namnakani.1@research.gla.ac.uk
University of Glasgow
Glasgow, United Kingdom

Yasmeen Abdrabou
yasmeen.essam@unibw.de
University of the Bundeswehr
Munich, Germany and
University of Glasgow, UK

Jonathan Grizou
jonathan.grizou@glasgow.ac.uk
School of Computing Science
University of Glasgow
Glasgow, United Kingdom

Augusto Esteves
augusto.esteves@tecnico.ulisboa.pt
ITI / LARSyS, Instituto Superior
Técnico, University of Lisbon
Lisbon, Portugal

Mohamed Khamis
mohamed.khamis@glasgow.ac.uk
University of Glasgow
Glasgow, United Kingdom

Figure 1: We evaluate the performance of the three widely used gaze-based interaction methods: Dwell time (A), Pursuits (B) and Gaze gestures (C), for target selections on handheld mobile devices while sitting (left) and while walking (right). All participants performed all selections using the three different techniques while sitting and while walking. The red arrow in (B) illustrates the direction in which a yellow dot stimuli was rotating around a selectable target. The red arrows in (C) indicate the directions in which the user could perform a gaze gesture. All arrows are for illustration and were not shown to participants.

Abstract

Gaze is promising for hands-free interaction on mobile devices. However, it is not clear how gaze interaction methods compare to each other in mobile settings. This paper presents the first experiment in a mobile setting that compares three of the most commonly used gaze interaction methods: Dwell time, Pursuits, and Gaze gestures. In our study, 24 participants selected one of 2, 4, 9, 12 and 32 targets via gaze while sitting and while walking. Results show that input using Pursuits is faster than Dwell time and Gaze gestures especially when there are many targets. Users prefer Pursuits when stationary, but prefer Dwell time when walking. While selection using Gaze gestures is more demanding and slower when there are many targets, it is suitable for contexts where accuracy is more important than speed. We conclude with guidelines for the design of gaze interaction on handheld mobile devices.

CHI '23, April 23–28, 2023, Hamburg, Germany

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; **Interaction techniques**.

Keywords

Eye Tracking, Smartphones, Tablets, Gaze-based Interaction, Smooth pursuit

ACM Reference Format:

Omar Namnakani, Yasmeen Abdrabou, Jonathan Grizou, Augusto Esteves, and Mohamed Khamis. 2023. Comparing Dwell time, Pursuits and Gaze Gestures for Gaze Interaction on Handheld Mobile Devices. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3544548.3580871

1 Introduction

With the recent advances in smartphone manufacturing, front-facing cameras are becoming more powerful and processors are becoming faster [24, 43], allowing smartphones to run eye tracking applications [24, 43]. Such advances allow a wide range of HCI applications. This includes empowering mobile users with novel gaze-based input methods that enable hands-free interaction. Examples of previously proposed applications of gaze-based interaction on mobile devices include gaze-based authentication [42, 46], transferring content between devices [76], gaze-based scrolling [73], and supporting other modalities using gaze [52, 71, 73].

Although a number of gaze interaction techniques, such as dwell time [21, 24], Pursuits [78], and gestures [21, 61] were deemed promising for mobile devices [15, 43], an empirical evaluation of how well they perform when used to interact with mobile devices in mobile settings is missing.

Most studies on gaze interaction were deployed in settings that are very different from those for daily gaze interaction on handheld mobile devices such as on interactive surfaces [81], desktop machines [69], public displays [47, 50], mobile devices held by a mount [24, 43], wearables [28] and head-mounted displays (HMDs) (e.g., for VR and AR) [27, 34, 49]. On the other hand, several prior works suggested that gaze interaction methods that do not require precise gaze estimates hold a lot of promise [38]. Examples of these include the use of eye behaviours such as gaze gestures [23] and smooth pursuit [78]. Both Pursuits and gestures were suggested to be suitable for mobile devices [43], and Pursuits was found to be suitable for selection while walking past public displays [47]. These positive results were likely due to the fact that these techniques are less reliant on calibration, and allow accurate interactions using inaccurate gaze estimates [43]. This is a substantial advantage in the context of mobile devices because calibration is likely to break frequently in the inevitably shaky mobile settings especially when the user moves or changes their posture. Nevertheless, considering that neither gestures nor Pursuits allow users to "point and select" as done by a mouse or a tapping a touchscreen, gaze interfaces that require accurate gaze estimation such as dwell time will continue to exist [30]. This underlines the importance of understanding the performance of the aforementioned gaze interaction techniques in mobile settings as a prerequisite to reaping the benefits of gaze on mobile devices.

While previous work evaluated similar techniques in non-mobile settings, or on modified mobile devices in unnatural settings (e.g., by wearing an HMD while interacting with a phone [34]), natural mobile settings feature unique challenges that make results from said previous studies inapplicable. This underlines the importance of investigating how well these techniques perform a) directly on mobile devices, b) while users are on the move, and c) with different number of selectable targets.

In this work, we report on the results of the first user study (N=24) to compare three widely used gaze interaction techniques namely 1) Dwell time, 2) Pursuits, and 3) Gaze gestures on smartphones while sitting and while walking, to select one of 2, 4, 9, 12 and 32 on-screen targets. We compare the methods in terms of selection time, error counts, timeouts, perceived cognitive load, and user preference. We found that while sitting, input using Pursuits (1.36 sec) is statistically significantly faster compared to Dwell time (2.33 sec) and Gaze gestures (5.17 sec). While gaze input is generally slower while walking, it is significantly slower when using Gaze gestures (6.68 sec) compared to Pursuits (2.14 sec) and Dwell time (2.76 sec). Despite being the slowest and most demanding, input using Gaze gestures is significantly more accurate than both other methods both while walking and while sitting, suggesting that it is more suited for situations in which accuracy is more important than speed. When asked about their preferences, our participants preferred Pursuits the most for stationary interaction and preferred Dwell time for interaction while walking.

Based on our results, we conclude with a set of guidelines for gaze interactions on handheld mobile devices. Our findings contribute to the field of gaze-aware interfaces and pave the road for adapting and deploying eye gaze interaction on handheld devices.

2 Related Work

Our work builds on previous work on gaze-based interaction and the opportunities and challenges of eye tracking on handheld mobile devices.

2.1 Gaze Interaction Techniques

Gaze-based interaction has long been studied by HCI researchers. It has been classified in prior work into two main categories: 1) Implicit gaze-based interaction, in which the interface adapts to the users' passive gaze behaviour. This approach is often used in attentive user interfaces [13, 20, 40] and in security applications [5, 14, 59, 62] especially biometric authentication [74]; and 2) Explicit gaze-based interaction, where users deliberately move their eyes to provide direct input. Our work focuses on explicit gaze-based interaction on handheld mobile devices.

The most common technique for explicit gaze input is Dwell Time [38]. In Dwell time, a brief delay is required to differentiate between casual viewing and gaze input [24, 30, 38, 57], thereby coping with the so-called Midas touch problem in which users make unintentional selections as they perceive potential targets. Dwell time has been used in gaze-only interactions [3, 54, 56, 70] and in multimodal interaction to support other modalities such as touch [63, 71], input using a stylus [64] or hand gestures [3, 65]. Dwell time requires accurate gaze estimates, hence the accuracy of the technique is highly dependent on calibration, and the mobile

nature of handhelds makes it likely that calibration would break often [43].

An alternative to Dwell time is to use interaction techniques that do not require accurate gaze estimates. In this regard, several interaction techniques were proposed, such as Pursuits [78]. The Pursuits technique relies on the Smooth Pursuit eye movements which is a distinctive form of eye movement that occurs when the eyes follow a moving object and human eyes cannot generate such movements without external stimuli to follow [28, 78]. This means that interfaces that use Pursuits need to show users moving stimuli. By measuring how well the user's eye movements match the trajectory of the moving stimuli, the system determines which object is being gazed at. This is a significant distinction from interfaces that employ absolute point of gaze [78]. Fatigue and confusion due to the constant movements of objects and being able to create distinct trajectories for all objects are among the challenges to enabling Pursuits on mobile devices [28, 78]. Pursuits has been used to interact with different devices such as smart watches [28], public displays [50, 78, 79], and in VR [49] and AR [29] environments. Input using Gaze gestures is also a widely used alternative in the literature [21, 23, 25, 26]. The technique requires users to perform coarse gestures in a certain direction [82] or perform a series of gestures with their eyes [23, 69]. One of the important advantages of gestures is that they do not require a screen's real state [21], and by using more gestures, a greater number of commands may be issued. On the downside, increasing the number of gestures introduces some complexity and comes with problems, as complex gestures may be difficult to recall cognitively, and they may be challenging to initiate and perform physically [61]. Gaze gestures found applications in gaming [36, 37, 61], authentication [4, 44, 46, 48, 69] and also as a generic input method for mobile devices [41]. The difference between Gaze gestures and Pursuits is that Gaze gestures, in recent implementations, are performed from memory rather than by following a stimulus. This requires spatial rather than temporal synchronisation and requires a learning phase for the user to know how to perform the gesture. Pursuits, on the other hand, has a tight spatial and temporal coupling with visual stimulus, meaning that you can only issue the commands by following the moving objects within the time window; manipulating the commands out of sync with the animation will not trigger a response [18, 23, 28, 78]. In terms of eye movements, gaze gestures are more similar to saccades, whereas Pursuits requires smooth pursuit eye movements. Both Pursuits and Gaze gestures allow for calibration-free interaction and also reduce the chances of unintentional selection as they require users to perform specific eye movements [10, 30].

Other types of gaze interaction techniques include the use of eye vergences [51, 53, 75], which relies on the simultaneous movements of both eyes in opposition to one another when looking at closeby targets [75]. Although eye vergence showed great potential in solving the Midas Touch problem [7] and outperformed Dwell time in selection speed [51], it is perceived to be uncomfortable and hard to perform regularly. It also shifts user's focus away from the screen, making it challenging to perceive the content while providing input [51]. A further promising gaze interaction technique relies on the Vestibulo-ocular reflex (VOR) which is a reflex action that occurs when a human fixates on a target and moves their head [11, 19]. VOR was proposed for detecting head gestures based on eye movements [67], and was used for improving gaze-based selection in VR when targets are occluded [58]. Prior work on gaze interaction also explored Optokinetic Nystagmus [39], which is a combination of saccadic and smooth pursuits eye movements, and was used to detect the image of interest in image scrolling application [39].

2.2 Opportunities and Challenges of Eye Tracking on Handheld Mobile Devices

While eye tracking on handheld mobile devices has been studied for more than 20 years in Mobile HCI, it is only recently that we started to see an uptake of real-time gaze estimation on off-the-shelf handheld mobile devices outside the lab. The recent integration of front-facing depth cameras in handheld mobile devices, and their improved processing power and camera resolutions, are transforming mobile eye tracking. In their review of gaze-enabled handheld mobile devices, Khamis et al. [43] argued that this brings a myriad of opportunities, such as gaze-based interaction on the move and in-the-wild analysis of gaze behaviour on mobile devices. Applications of this include improving interaction on mobile devices [16, 20, 21], security applications [42] including authentication [48] and privacy protection [9, 72, 83], as well as in-the-wild gaze behaviour analysis [5, 6, 80].

On the downside, eye tracking on mobile devices comes with a unique set of challenges. Unlike wearable eye trackers and stationary remote cameras, handheld mobile devices track the user's eyes using a front-facing camera, whose view might be occluded by the user's clothing or their hands [35, 45]. Khamis et al. collected a dataset of 25,726 photos taken from front-facing cameras of smartphones in the wild [45]. They found that users hold their phones in different ways and that it is not always the case the face is sufficiently visible to the camera for estimating gaze. Huang et al. also collected a dataset of photos taken from tablets but in controlled settings [35] and found that the full face of users was visible in only 30.8% of the photos. Additionally, the environment is often shaky due to the user's movements (e.g., interacting while walking) and the dynamic environment (e.g., in a bus on a bumpy road). This is further worsened by the user's holding posture, which may not necessarily result in their face being visible to the camera [45]. All of the factors above impact gaze estimation accuracy, which may in turn impact the user experience when using gaze for interaction.

2.3 Summary and Contribution Over Prior Work

Our work focuses on application-independent gaze-based interaction, where users actively move their eyes to select one of many on-screen targets when stationary and when on the move. While previous work studied gaze interaction and eye tracking on handheld mobile devices, an investigation of how gaze interaction techniques perform on said devices in mobile settings is missing. A key novelty of our contribution is that we consider mobile settings where users are interacting while walking. While gaze interaction while on the move was studied on public displays [47] and on HMDs [49], the results from prior studies are not transferable to the context of handheld mobile devices, where users' holding postures and

the shaky environments play a big role in the quality of collected gaze data.

3 Gaze Interaction Techniques: Concepts and Our Implementations

In this work, we focus on comparing three of the most widely used gaze interaction methods: 1) Dwell time, 2) Pursuits, and 3) Gaze gestures. Below we describe our implementations. Note that none of the implementations visualised where the user is looking, as this will likely distract users, especially when the gaze estimates are inaccurate.

3.1 Dwell Time

Dwell time selection is performed by fixating on a target for a period of time [24]. In order to perform a selection using Dwell time: the user would gaze and fixate at the object to be selected for a period of time to indicate attention. Literature on dwell times reports values from 150 ms to 1500 ms [27]. Based on prior work and pilot tests with 5 participants using 500 ms, 800 ms, and 1000 ms, we adopted a dwell time of 800 ms [17, 30]. Thus, in our implementation, a selection is executed when a target has been gazed at for a minimum of 800 ms by calculating the mean fixation point within the specified time window. This improves our approach's robustness against noise, because if a few gaze points land outside the target, they will not impact selection negatively. In our implementation, this is the only gaze interaction technique that is preceded by a calibration process.

3.2 Pursuits

Pursuits selection can be performed by matching the eyes' trajectory with the relative trajectory of the object of interest [78]. To provide input using Pursuits, the user would: 1) observe the target that needs to be selected, then 2) follow the object orbiting around the target briefly with their eyes. This will result in selecting the target and triggering its functionality. In our implementation, we use the Pearson correlation coefficient to determine how similar the user's eye movements are to the moving stimuli or target. This is the most common implementation in prior work [22, 27, 28, 49, 77, 78], which is calculated as follows:

$$corr_x = \frac{E[(Eye_x - \bar{Eye_x})(Targ_x - \bar{Targ_x})]}{\sigma_{Eye_x}\sigma_{Targ_x}} \quad (1)$$

Where $\bar{Eye_x}$, σ_{Eye_x}, $\bar{Targ_x}$, and σ_{Targ_x} are the means and standard deviations of the horizontal eye and moving targets positions respectively. The coefficient is also calculated for the vertical positions the same way, denoted as $corr_y$. Our system calculates Pearson's correlation, $corr_x$ and $corr_y$ every one second (30 samples). The choice of 1,000 ms was motivated by prior work [27, 28] and pilot testing with 2 participants. As done in prior work [28], as long as the correlation of the smallest of the two is greater than 0.8, the moving target whose movement most closely matches the user's eye movement is considered to be the target being observed. If the system does not detect a selection within the time window, the correlation is computed for every new sample collected over a sliding window of 30 samples. 120°/sec is the constant speed of all targets.

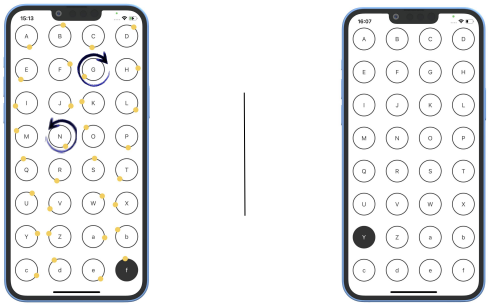


Figure 2: Targets were designed to look like app icons found on the home screens of Android and iOS devices. All the targets presented in the interface were selectable and the targets participants had to select were coloured in black. For all three interaction techniques, the interface has the same arrangement with the letters inscribed alphabetically in the circles, left to right, top to bottom. Left: Interface for Pursuits with dynamic targets – note that arrows are for illustration and were not shown to participants. Right: Dwell time interface. When a user fixates on a target for at least 800 ms, the background of the target changes colour.

By having half of the targets move in opposite directions (clockwise and counterclockwise) and separating their initial positions by 360/n (with n equal to the number of targets displayed on-screen), we increased the distance between the targets presented to minimise acquisition errors [28]. Calibration is not required for our implementation, and thus no calibration data is used for Pursuits selections.

3.3 Gaze Gestures

Gaze gestures are essentially eye movements that follow particular patterns in a sequential time order to issue commands or perform selections. Our implementation of Gaze gestures follows that of Look to Speak by Google [2]. For the user to provide input, they: 1) locate the target on the screen, either on the left or right side of the screen (see Figure 3), 2) perform a gesture (right or left), to select the side that has the desired target. 3) Each time the user selects a side, the number of targets will narrow down until they select the desired target. Gesture selection is achieved by performing a single right or left gesture within a 1000 ms (30 samples) time window, which was also chosen based on prior work [21, 61] and pilot tests with 2 participants. Single gestures are argued to be efficient, easy to learn, and require less effort compared to complex gestures [60], and horizontal gestures were found to be faster compared to vertical gestures [61]. For detecting gestures, we use Pearson correlation coefficient as part of a template matching algorithm to match the user's scan path against the right and left gestures template paths [69]. With a correlation value above a threshold of 0.8, the template path of the gesture that is similar to the candidate path is determined as the one followed by the user. To differentiate between normal eye movement and gaze gestures, the gesture is completed when the user gazes on either side outside the screen's bounds as they move their gaze towards them [33, 60]. To detect that the user's

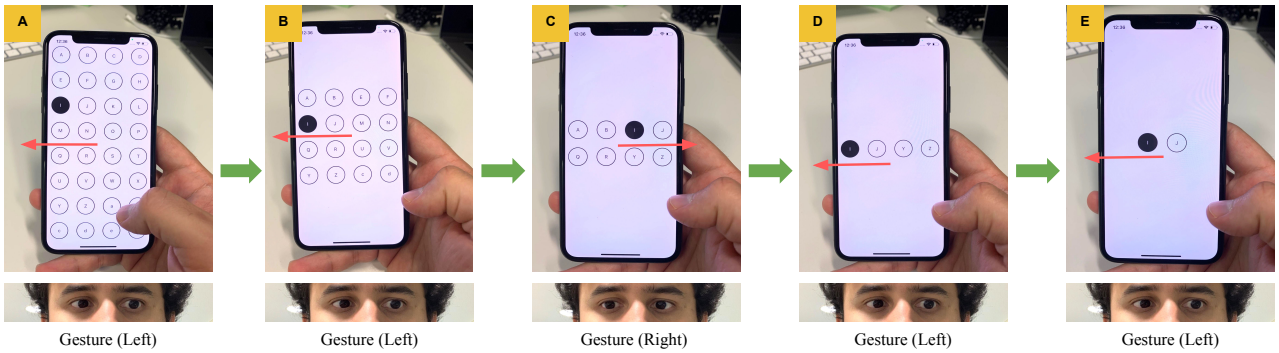


Figure 3: Making selections using gaze gestures in our implementation requires a number of steps, which increases as the number of targets increase. For example, to select one of 32 targets, the participant has to perform a gaze gesture in the direction where the target is. In this example, the target is labelled "I" (left in A). The targets on the left side are then redistributed to allow the user to narrow down their selection further.

gaze left the screen to complete the gesture, and based on pilot testing with 2 participants, we implemented a caution area (mobile device's length × 20 pt width in the iOS coordinate system), near the screen border on both the left and right edges of the screen [33]. If the last horizontal gaze point in the current window lands in the caution area, they are considered as outside the screen [33]. Because we only detect left and right gestures, calibration errors on the y-axis did not impact selection accuracy. Calibration errors on the x-axis would have an impact only if they offset the last gaze point in the gesture by more than 2 × 20 pt + mobile device's width in the opposite direction. A calibration error that large is unlikely and was indeed not observed throughout our experiment. Gestures ending inside the screen area are not registered. Calibration is not required; thus, no calibration data is used in our implementation of gaze gestures.

4 Evaluation

Our experiment was conducted in a quiet room without windows and under constant lighting condition. Our aim was to compare the performance and perception of three gaze interaction techniques for handheld mobile devices: Dwell time, Pursuits and Gaze gestures. To this end, we studied the effect of the number of targets (2, 4, 9, 12, 32) on the performance of each technique in two different states: while sitting and while walking. Participants were free to hold the mobile device the way they would do naturally. Markers were put on the floor to indicate how the participants should move in the walking state (see Figure 1), and participants were instructed to walk as they would naturally and not to pause walking while completing the tasks. Throughout the experiment, the experimenter monitored the participants and ensured they did not stop when performing selection while walking. Participants were seated in front of a desk for the sitting state.

4.1 Participants

We recruited 24 participants (15 male, 9 female) with an average age of 28.88 years old ($range$: 20−41, SD = 5.65), and an average height of 1.72m ($range$: 1.58m − 1.95m, SD = 8.89). Participants indicated that they have little to no previous experience (M = 1.08, SD = 1.50) in eye tracking (0: no experience; 5: Very experienced). Out of the 24 participants, 1 participant suffered from astigmatism, 5 suffered from farsightedness, 5 suffered from nearsightedness and 1 participant suffered from both nearsightedness and astigmatism, while 1 more participant suffered from nearsightedness, farsightedness and astigmatism. Participants reported that they sometimes (M = 3.13, SD = 1.19) use their phones while walking (0: never; 5: always). 3 participants wore glasses during the experiment. The experiment took an hour and participants were compensated by an e-shop voucher. This experiment received ethical approval from our institution.

4.2 Apparatus

For the experiment, we used an iPhone X with iOS version number 15.3.1. The device is equipped with a super Retina HD display that has a resolution of 2436x1125 pixels and a front-facing camera with 7 MP sensor, f/2.2-aperture lens, and 32mm-equivalent focal length. The SeeSo SDK library (link) was used for Eye tracking. With 30 frames per second, SeeSo uses the RGB images from the front-facing camera of the phone for real-time gaze point estimation. All calculations were done locally on the device. A path with a total distance of 5.20 m was marked on the floor using tape to guide the participants when performing the tasks while walking.

As for the user interface, to keep all conditions visually identical, targets on the screen were arranged in a way to represent typical app icons on the android and iOS home screens (see Figure 2). Targets were centered to maintain consistency among all the techniques. We chose this design because in gaze gestures, targets need to be centered before being split into the right and left sides of the screen to narrow down the selection. As for the target size, a size of (65 pt; 195 pixels) was used. This is equivalent to 1.7° visual angle since the calculated average distance of participants to the screen was 41.5 cm. This was motivated by the fact that Apple recommends using a size of 60 pt for apps icons [1] which maps to (180 pixels) in iPhone X, the device we used for the experiment. We added an extra 5 pt to the recommended size to account for the area in which the moving stimuli is displayed in the Pursuits condition.

4.3 Study Design

The study was designed as a repeated measures experiment. There were two independent variables:

IV1 **Gaze Interaction Techniques**: we covered three conditions: Dwell time, Pursuits and Gaze gestures.
IV2 **Number of Targets**: We covered five different number of targets: 2, 4, 9, 12 and 32. We were interested in exploring the impact of the number of targets on the selection techniques. We chose a maximum of 32 targets as this is the highest number of targets that can be fit on the screen considering Apple recommended size for apps icons [1].

4.4 Procedure

Upon arrival, participants were asked to sign a consent form, complete a demographic questionnaire and rate their previous experience with eye tracking. An information sheet about the experiment was also provided. Afterwards, participants were introduced to the experiment, its goal, the tasks, and also the metrics the system collects (selection time, error counts and timeouts). A Latin Square design was used to counter-balance the order of conditions. The study was split into two parts: a part for the walking state, and another for the sitting state. Half of the participants started with the walking state, followed by the sitting state, while the other half started with the sitting state and concluded with the walking state. In each part, participants went through three blocks; one block per gaze interaction technique. In each block, participants had to select one of 2, 4, 9, 12 and 32 targets in a counter-balanced order using Latin square. **Before each block**, participants were explained how to perform the selection and were allowed to complete three trial runs to familiarise themselves with the technique. These runs were excluded from the analysis. All the targets presented in the interface were selectable. The target participants had to select was coloured in black (see Figure 2). Participants underwent a calibration phase before the dwell time block, but not before the other two techniques, as it is not required. Using SeeSo's calibration process, the calibration was performed by presenting participants with five different targets at known points on the screen to gaze at for few seconds, one at a time, to establish a mapping between the screen's coordinates and the optical axes of the eyes. Calibration data was discarded after the dwell time block. **During the block**, for each selection task with different number of targets, participants were first presented with a screen with instructions. Upon tapping the start button, the task started. Each selection task ended after a 20-second timeout has elapsed or when participants selected the correct target with a displayed message confirming that, whichever is earlier. **Following each block**, participants' perceived workload was collected using the NASA-TLX scale (link), and then qualitative feedback was collected through 5-point Likert scale and open-ended questions. **After completing the three blocks**, we asked participants to rank the selection techniques based on their preference and performance. Each participant performed 2 states x 3 techniques x 5 number of targets = 30 trials. The study lasted for approximately an hour for each participant.

4.5 Limitations

In our study, participants did not need to examine selectable objects in search for the target as it was already highlighted and can be clearly distinguished. We made this decision as our focus is to evaluate the interaction methods, whereas the time it takes to find the target is out of our scope. However, this also means the performance of the input techniques may differ in day-to-day scenarios in terms of selection time, as users will need to search for the target, and in terms of errors, as users may accidentally select a target while searching. This is particularly the case for Dwell time, where Midas touch selections, while the user is examining the interface, can be disruptive. Midas touch can be overcome by requiring longer dwell durations, but this, in turn, results in longer interaction time.

Another limitation is that the effects caused by the different targets shapes and sizes were not analysed. As mentioned earlier, we aimed to keep all conditions visually identical to avoid potential biases.

Finally, in the walking part of the study, participants had no obstacles to avoid while performing selections. Our aim was to characterise the performance of interaction techniques while participants are walking and pave the way towards on the go gaze interaction. However, during everyday life, people face many obstacles and distractions which might impact their gaze behaviours when interacting with their mobile devices, which in turn might affect the results. This can be addressed in a future study

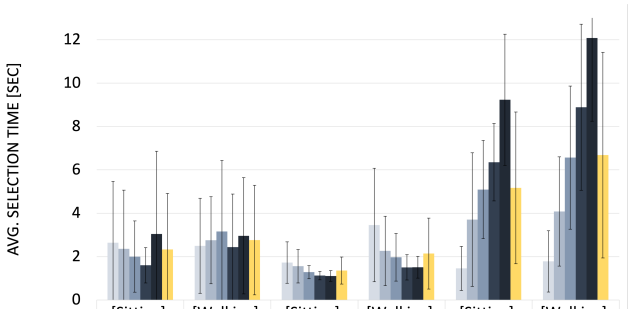
5 Results

We measured selection time, error counts, timeouts, user perceived cognitive load using NASA TLX [32], and user preference. We excluded data points that matched the following criteria as outliers: selection times where a timeout occurred (i.e., exceeded 20 seconds without a successful selection [31]), selection times and error counts where gaze tracking failed (e.g., due to the participant's face not being visible) and values that were deemed as outliers when inspecting a box plot visualisation prior to any exclusion [68]. These outliers were replaced with the mean of the rest of the values for that condition. Out of 720 values for each measure, 5 data points were excluded from selection time and 7 data points from error counts. For selection time, error counts and the perceived cognitive workload, we used two-way repeated measures ANOVA tests. We used Greenhouse-Geisser-correction in cases where Mauchly's test indicated a violation of sphericity. P-values for post-hoc tests were corrected using Bonferroni correction to account for multiple comparisons. In case of an interaction effect, separate one-way repeated measures ANOVA tests were run too to distinguish the impact of each condition.

5.1 Selection Time

We define selection time as the time from the moment the targets appeared until the moment the correct target was selected. The descriptive statistics are summarised in figure 4.

5.1.1 Sitting state The statistical tests revealed a significant main effect for input techniques $F_{1.452,\ 33.395} = 67.814$, $P < .001$ and number of targets $F_{2.754,\ 63.335} = 18.561$, $P < .001$ on selection time, with an interaction effect between the two $F_{4.326,\ 99.506} = 19.355$,



	Dwell time		Pursuits		Gaze gestures	
	[Sitting]	[Walking]	[Sitting]	[Walking]	[Sitting]	[Walking]
2 targets	2.64	2.50	1.72	3.46	1.46	1.78
4 targets	2.36	2.76	1.56	2.26	3.71	4.08
9 targets	2.00	3.16	1.28	1.97	5.09	6.57
12 targets	1.60	2.44	1.13	1.50	6.36	8.89
32 targets	3.05	2.96	1.11	1.51	9.23	12.08
Average	2.33	2.76	1.36	2.14	5.17	6.68

Figure 4: Participants select faster using Pursuits regardless of the number of targets in both sitting and walking states. Selection time for Gaze gestures is linearly affected by the number of targets. The error bars represent the standard deviation.

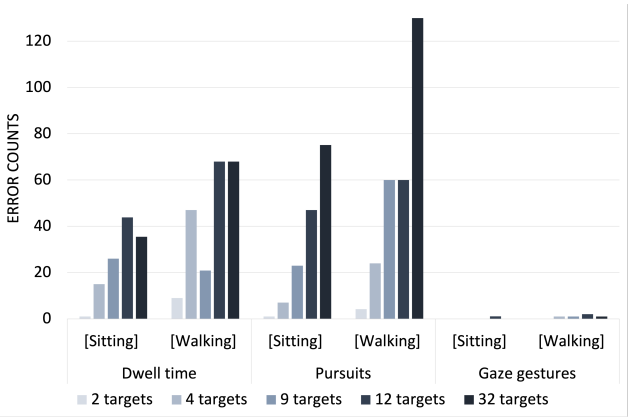


Figure 5: Number of attempts before a successful selection. Errors are less in the sitting state across all three input techniques. The graphs show that Gaze gestures is less error-prone and immune to the impact of the number of targets. Errors increase with more targets when using Dwell time and Pursuits

$P < .001$. Significant differences were found between all pairs: between Dwell time and Pursuits ($P = .002$), between Dwell time and Gaze gestures ($P < .001$), and between Pursuits and Gaze Gestures ($P < .001$). The mean values were 2.33 sec for Dwell time ($SD = 2.58\,sec$), 1.36 sec for Pursuits ($SD = .62\,sec$), and 5.17 sec for Gaze Gestures ($SD = 3.50\,sec$). Due to the interaction effect, we investigated in-depth the effect of the techniques on selection time with respect to each number of targets. Post-hoc analysis showed significant differences between multiple pairs as detailed in table 1.

5.1.2 Walking state We found a significant main effect for techniques $F_{2,\,46} = 96.896$, $P < .001$, and number of targets $F_{4,\,92} = 14.728$, $P < .001$ on selection time. There was also statistically significant two-way interaction between the two $F_{8,\,184} = 27.988$, $P < .001$. Significant differences ($P < .001$) were found between Gaze gestures ($M = 6.68\,sec, SD = 4.74\,sec$) and both Dwell time ($M = 2.76\,sec, SD = 2.53\,sec$) and Pursuits ($M = 2.14\,sec, SD = 1.64\,sec$). Post-hoc analysis showed significant differences between multiple pairs as detailed in table 1.

Observation 1: In terms of selection time, selection using Pursuits performs better than Dwell time and Gestures in both the sitting and walking states.

5.1.3 Sitting vs Walking The results show that regardless of the participants' states during selection, Pursuits was the fastest for four or more targets, followed by Dwell time. On the other hand, Gaze gestures was the fastest when there were two targets to choose from, followed by Pursuits in the sitting state, and followed by Dwell time in the walking state (see Figure 4). To support the quantitative results, participants responded to 'Performing selections using this technique is fast' on a 5-point Likert scale (1=strongly disagree; 5=strongly agree), they perceived Pursuits as the fastest technique when sitting ($Med = 4.5, IQR = 4 - 5$) while the three input techniques were perceived equally fast when walking (see Section 5.5).

5.2 Error Counts

We define error count as the number of attempts before a successful selection, as erroneous selections were not accepted and did not cause the trials to end. Figure 5 summarises the descriptive statistics.

5.2.1 Sitting state We tested for effects on the error counts. We found a significant main effect of techniques $F_{1.227,28.217} = 14.447$, $P < .001$, significant main effect of number of targets $F_{2.332,53.626} = 8.737$, $P < .001$, and an interaction between the two $F_{3.301,75.925} = 4.416$, $P = .005$. Significant differences were found between Dwell time and Gaze gestures ($P < .005$) and between Pursuits and Gaze gestures ($P < .001$). The mean values were 1.01 for Dwell time ($SD = 2.62$), 1.28 for Pursuits ($SD = 1.80$), and .01 for Gaze gestures ($SD = .09$). When analysing the effect of techniques on error counts, pairwise comparisons revealed significant differences between multiple pairs as detailed in table 2. For two targets, we found no evidence of significant differences between techniques as error counts were low for all techniques with one wrong selection for Dwell time and Pursuits and no wrong selection for Gaze gestures. Regardless of the number of targets in the sitting state, 60% of trials were completed with no errors using Dwell time, 42% using Pursuits and 94% using Gaze gestures.

5.2.2 Walking state Significant main effects were found for techniques $F_{1.419,32.640} = 16.446$, $P < .001$ and number of targets $F_{2.703,62.172} = 14.649$, $P < .001$ on error counts. There was an interaction between the techniques and the number of targets $F_{3.737,85.948} = 5.740$, $P < .001$. Significance differences ($P < .001$) were found between Dwell time and Gaze gestures and between Pursuits and Gaze gestures. The mean values were 1.77 for Dwell time ($SD = 3.27$), 2.32

Table 1: In both Sitting (Left) and walking (Right) states, the selection time for Gaze gestures is significantly longer than that of Dwell time and Pursuits as the number of targets increases. Pursuits, on the other hand, is significantly faster than Dwell time and Gaze gestures when more targets are displayed. Means are given in brackets.

Significant differences in Selection Time [Pairwise comparisons]					
Sitting state			**Walking state**		
Number of targets: 2		**p<**	**Number of targets: 2**		**p<**
			Pursuits (3.46 sec)	Gestures (1.78 sec)	.05
Number of targets: 4		**p<**	**Number of targets: 4**		**p<**
Pursuits (1.56 sec)	Gestures (3.71 sec)	.05	Dwell time (2.76 sec)	Gestures (4.08 sec)	.05
			Pursuits (2.26 sec)	Gestures (4.08 sec)	.05
Number of targets: 9		**p<**	**Number of targets: 9**		**p<**
Dwell time (2.00 sec)	Gestures (5.09 sec)	.001	Dwell time (3.16 sec)	Gestures (6.57 sec)	.01
Pursuits (1.25 sec)	Gestures (5.09 sec)	.001	Pursuits (1.97 sec)	Gestures (6.57 sec)	.001
Number of targets: 12		**p<**	**Number of targets: 12**		**p<**
Dwell time (1.60 sec)	Pursuits (1.13 sec)	.05	Dwell time (2.44 sec)	Gestures (8.89 sec)	.001
Dwell time (1.60 sec)	Gestures (6.36 sec)	.001	Pursuits (1.50 sec)	Gestures (8.89 sec)	.001
Pursuits (1.13 sec)	Gestures (6.36 sec)	.001			
Number of targets: 32		**p<**	**Number of targets: 32**		**p<**
Dwell time (3.05 sec)	Gestures (9.23 sec)	.001	Dwell time (2.96,sec)	Pursuits (1.51 sec)	.05
Pursuits (1.11 sec)	Gestures (9.23 sec)	.001	Dwell time (2.96 sec)	Gestures (12.08 sec)	.001
			Pursuits (1.51 sec)	Gestures (12.08 sec)	.001

for Pursuits (SD = 3.24), and .04 for Gaze gestures (SD = .20). Post-hoc analysis showed significant differences between multiple pairs as detailed in table 2. Similar to the sitting state for two targets, tests showed no evidence of significant differences between techniques on error counts. In the walking state and regardless of the number of targets, 49% of trials were completed with no errors using Dwell time, 38% using Pursuits, and 92% using Gaze gestures.

5.2.3 Sitting vs Walking The results show that participants made more errors in the walking state compared to sitting. On the other hand, Gaze gestures performed well in both states and regardless of the number of targets, as it produces fewer input errors. Participants' feedback (5-point Likert scale; 1=strongly disagree; 5= strongly agree) also supports this argument as participants perceived Gaze gestures to be more accurate than Dwell time and Pursuits in both sitting ($Med = 5, IQR = 4 - 5$) and walking ($Med = 4, IQR = 3 - 4.25$) states ,when asked if they found the technique accurate (see Section 5.5). The results also show that Dwell time was more accurate than Pursuits as the number of targets increased in both the sitting and the walking states (see Figure 5).

Observation 2: While gaze gestures require longer selection times when there are many targets, they are highly accurate and thus reliable in contexts where accuracy is more important than speed

5.3 Timeout

We counted how many times a timeout occurred, calculated as a percentage of the total number of trials per condition. A timeout is considered if a participant failed to select the correct targets within 20 seconds.

We observed that timeout occurred more when performing a selection using Dwell time compared to Pursuits and Gaze gestures

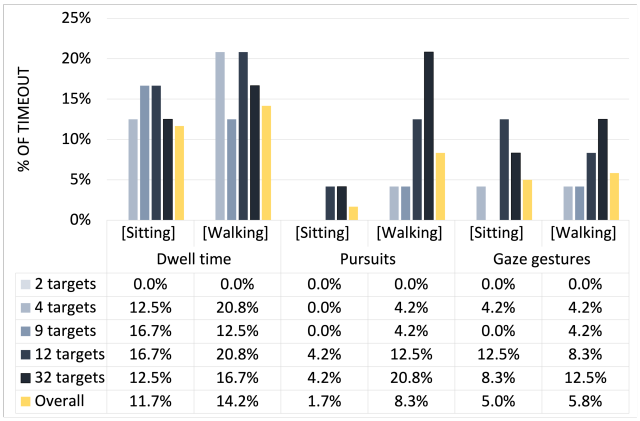


Figure 6: Overall, more participants failed to select the correct targets within 20 seconds in the walking state compared to the sitting state. More timeout occurred when performing selection using Dwell time compared to Pursuits and Gaze gestures in both sitting and walking states.

in both sitting and walking states (see Figure 6). Timeout could happen due to the participant's face not being visible as a result of participants' changing their holding postures. For Dwell time, since it also requires accurate gaze estimation and relies on calibration [24], the inevitable shaky mobile settings cause calibration to break frequently and thus, results in participants failing to perform a selection within the time limit.

Observation 3: As expected, Dwell time relies on calibration because it requires accurate gaze estimation. The inevitably shaky mobile settings typically result in calibration to break often. This

Table 2: Performing a selection using Gaze gestures is less error-prone compared to Dwell time and Pursuits in both Sitting (Left) and walking (Right) states. We found no significant differences between Dwell time and Pursuits in terms of error counts in most conditions, means in brackets.

Significant differences in Error counts [Pairwise comparisons]					
Sitting state			**Walking state**		
Number of targets: 4		**p<**	**Number of targets: 4**		**p<**
Dwell time (.625)	Gestures (.000)	.05	Dwell time (1.958)	Gestures (.042)	.05
Pursuits (.292)	Gestures (.000)	.05	Pursuits (1.000)	Gestures (.042)	.05
Number of targets: 9		**p<**	**Number of targets: 9**		**p<**
Pursuits (.958)	Gestures (.000)	.05	Dwell time (.870)	Pursuits (2.500)	.05
			Dwell time (.870)	Gestures (.042)	.01
			Pursuits (2.500)	Gestures (.042)	.001
Number of targets: 12		**p<**	**Number of targets: 12**		**p<**
Dwell time (1.826)	Gestures (.043)	.05	Dwell time (2.833)	Gestures (.083)	.05
Pursuits (1.958)	Gestures (.043)	.001	Pursuits (2.500)	Gestures (.083)	.01
Number of targets: 32		**p<**	**Number of targets: 32**		**p<**
Pursuits (3.130)	Gestures (.000)	.001	Dwell time (2.833)	Gestures (.042)	.005
			Pursuits (5.417)	Gestures (.042)	.001

prevents participants from selecting the correct target within the time limit.

5.4 Perceived Cognitive Load

Figure 7 shows the mean scores for NASA-TLX dimensions in both sitting and walking states. The scores are out of 100. The lower the scores, the lower the workload.

5.4.1 Sitting state The statistical tests revealed significant differences in the overall NASA-TLX score when calculating the mean across all six dimensions between the input techniques, $F_{1.410,32.427} = 4.387$, $P = .032$. Significant differences ($P < .05$) were found between Pursuits ($M = 9.76, SD = 8.86$) and Gaze gestures ($M = 19.72, SD = 18.59$). We ran multiple repeated measures ANOVAs for each NASA-TLX dimension. A significant main effect was found for input techniques on mental demand $F_{1.599,36.782} = 5.905$, $P = .009$, physical demand $F_{1.575,36.229} = 4.646$, $P = .023$, and effort $F_{2,46} = 5.891$, $P = .005$. Post-hoc pairwise comparison revealed a significant difference between Pursuits and Gaze gestures ($P < .01$) in mental demand, physical demands, and effort.

Observation 4: The difference in users' perceived workload was significant between Pursuits and Gaze gestures in the sitting state. Participants self-reported lower mental demand, lower physical demand, and less effort and frustration when providing input with Pursuits compared to Dwell time and Gaze gestures.

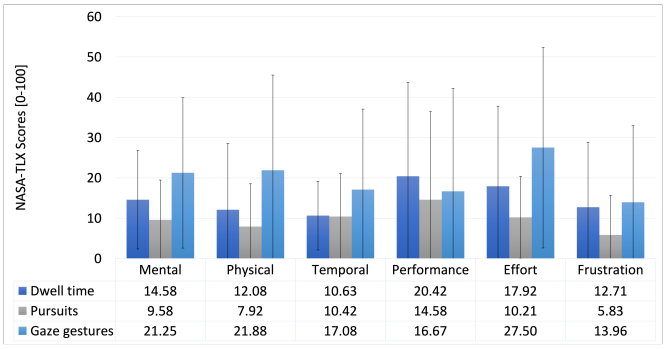
5.4.2 Walking state We found no statistically significant differences in the overall NASA-TLX score between the input techniques $F_{2,46} = 5.891$, $P = .097$. The mean values were 18.96 for Dwell time (SD = 14.18), 19.44 for Pursuits ($SD = 16.90$), and 27.01 for Gaze gestures ($SD = 16.80$). Multiple repeated measures ANOVAs were run to investigate if there is an effect of input techniques on each NASA-TLX dimension. Significant main effects were found for input techniques on physical $F_{2,46} = 8.414$, $P < .001$, and temporal demand $F_{2,46} = 3.712$, $P = .032$. Post-hoc pairwise comparison revealed a significant difference between Dwell time and Gaze gesture ($P < .01$) and between Pursuits and Gaze gestures ($P < .05$) in physical demand. the not significantly different ($P > .05$).

5.5 Qualitative Feedback

Participants responded to Likert scale questions (1-Strongly disagree to 5-Strongly agree) and open-ended questions to reflect on the different input techniques. As this data is non-parametric, we used Friedman tests to check for significance and Wilcoxon Signed Rank Test for posthoc pairwise comparisons with Bonferroni correction applied on the significance level to account for multiple comparisons. Overall, participants enjoyed the hands-free nature of gaze selections.

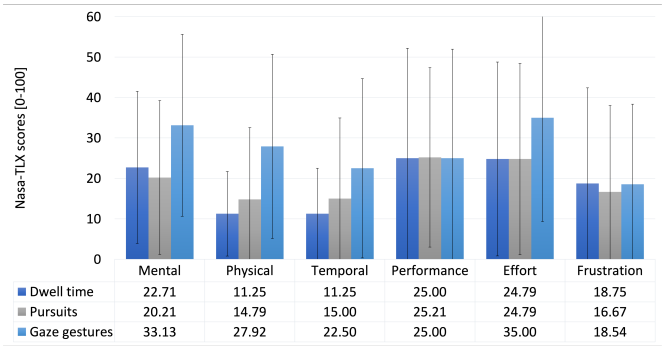
5.5.1 Sitting state As shown in Figure 8a, participants found Pursuits to be faster ($Med = 4.5, IQR = 4 - 5$) and less eye tiring ($Med = 2, IQR = 1 - 3$) compared to Dwell time and Gaze gestures. Although using Gaze gestures was the least favourable, participants indicated it was the most accurate ($Med = 5, IQR = 4 - 5$). Participants perceived Pursuits significantly easier to use $\chi^2(2) = 6.904$, $P = .032$, where pairwise comparisons showed significant differences between Pursuits ($Med = 5, IQR = 4 - 5$) and both Dwell time ($Med = 4, IQR = 3 - 5$) and Gaze gestures ($Med = 4, IQR = 3 - 5$). Gaze gestures was perceived as the least natural to use ($Med = 3, IQR = 2 - 3.25$). This was also found to be statistically significant $\chi^2(2) = 10.962$, $P < .005$, where comparisons revealed significant differences between Gaze gestures and Pursuits ($Med = 4, IQR = 3 - 4.25$), $P = .003$. This indicates that Pursuits is perceived as more natural.

Participants preferred Pursuits and Dwell time over Gaze gestures when selecting one of 9, 12, and 32 targets. While with less number of targets, no technique has preference over the others (see Figure 8c). However, some participants disliked Pursuits with more



	Mental	Physical	Temporal	Performance	Effort	Frustration
Dwell time	14.58	12.08	10.63	20.42	17.92	12.71
Pursuits	9.58	7.92	10.42	14.58	10.21	5.83
Gaze gestures	21.25	21.88	17.08	16.67	27.50	13.96

(a) [Sitting state]



	Mental	Physical	Temporal	Performance	Effort	Frustration
Dwell time	22.71	11.25	11.25	25.00	24.79	18.75
Pursuits	20.21	14.79	15.00	25.21	24.79	16.67
Gaze gestures	33.13	27.92	22.50	25.00	35.00	18.54

(b) [Walking state]

Figure 7: The mean Task Load index score of participants as indicated in the NASA TLX questionnaire. The error bars represent the standard deviation.

targets as they found the moving stimuli distracting. Although the use of gaze gestures was not the most preferred, P13 preferred it because of its reliability regardless of the number of targets *"it was easy to select irrespective of the number of targets but a bit of eye movement was needed"*.

We asked the participants about their perception of the input techniques. Four participants perceived Dwell time and Pursuits to be *fast* while nine participants reported Gaze gestures as such. *"It's [Gaze gestures] quite fast when you get the hang of it"* (P21). One participant mentioned that Pursuits *"Feels very interactive"*. Four participants reported Pursuits to be accurate. On the downside, calibration in Dwell time is an issue that was reported by 5 participants. P11 criticised the techniques *"It forbids me from following my natural instinct of moving my eyes"*, and 4 participants reported that Dwell time requires a lot of time for selection (P1, P3, P8, P15). Few participants reported that Pursuits become stressful to the eyes as the number of targets increase. Few participants disliked the fact that Gaze gestures require lots of eye movements and that they had to look towards the edge of the screen to complete the gestures.

Observation 5: In the sitting state: participants found Pursuits significantly easier to use, faster, and less tiring compared to Dwell time and Gaze gestures. Both Pursuits and Dwell time were perceived to be more natural to use, easier to learn, and can be used daily.

Ranking the techniques: At the end of the study, we asked the participants to rank their preferences for the input techniques. Raw scores were replaced by their weight factor; an input technique gains 3 points if ranked first, 2 points if ranked seconds, and 1 point if ranked last, and then weights are summed up to compute weighted scores. Regardless of the number of targets, Pursuits was the most preferred one (Score = 56), followed by Dwell time (Score = 49), and then Gaze gestures (Score = 39). This matches the results from the perceived selection time (see Figure 8a) and the measured selection time (see Figure 4). When asked to rank their preference based on the number of targets, Pursuits was also the most preferred technique for 2, 4 and 9 targets while Dwell time was the most preferred with 12 and 32 targets. Participants who preferred Pursuits attributed that to the speed and ease of use, *"Pursuits technique was the best responsive one"* (P3). Some ranked Dwell time first because they found it intuitive, easier to learn and fast. P22 preferred Pursuits because it *"didn't require calibrating anything"*. P11 ranked Gaze gestures first because *"felt more in control"* while those who ranked it last found it stressful and straining. The ranking results are summarised in Table 3.

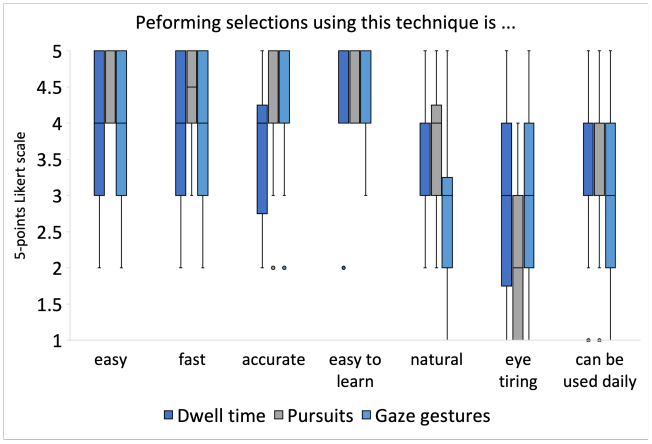
Observation 6: In the sitting state: Participants ranked Pursuits first over other techniques with 2,4 and 9 targets while Dwell time was the most preferred with 12 and 32 targets

5.5.2 Walking state Through the questionnaire, participants perceived both Pursuits and Dwell time as easier to learn and more natural compared to Gaze gestures (see Figure 8b). We can also see that Dwell time was perceived as less eye-tiring and preferred to be used daily. Friedman revealed statistically significant differences between the techniques in terms of being natural to use, $\chi^2(2) = 11.760$, $P = .003$, where pairwise comparisons showed that Dwell time ($Med = 4, IQR = 4 - 5$) and Gaze gestures ($Med = 3, IQR = 2 - 4$) differs significantly, $P < .001$.

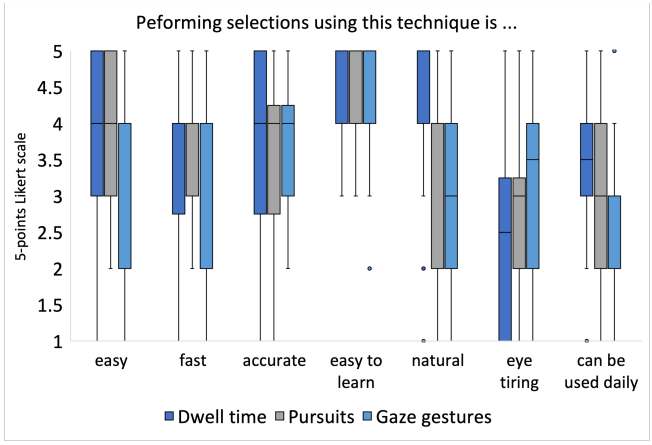
Similar to the sitting state, participants indicated that Gaze gestures was the most eye-tiring ($Med = 3.5, IQR = 2 - 4$). This was also shown to be significantly different, $\chi^2(2) = 6.441$, $P < .05$. However, post-hoc analysis did not reveal significant differences between the pairs. Six participants reported that Gaze gestures was eye tiring because it was hard to end the gesture off-screen.

Participants preferred to use Dwell time and Pursuits over Gaze gestures as the number of targets increase except for 32 targets where they preferred Dwell time (see Figure 8d). However, no significant differences in users' preference were found between techniques.

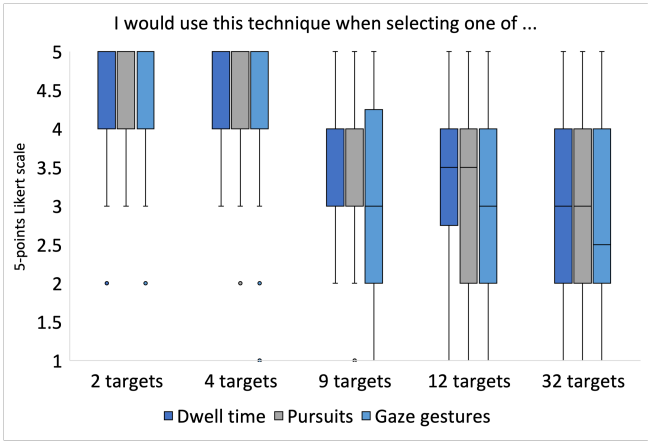
Observation 7: In the walking state: Participants ranked Dwell time as more preferred input method over Pursuits and Gaze gestures.



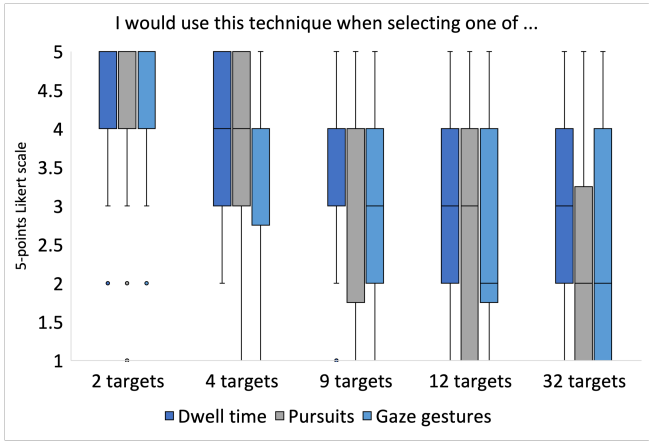
(a) Usability aspects in the sitting state



(b) Usability aspects in the walking state



(c) Preferences for techniques in the sitting state



(d) Preferences for techniques in the walking state

Figure 8: Participants rated aspects of the three input techniques on 5-point Likert Scales (1=Strongly Disagree;5=Strongly Agree) in both states: Sitting in (a) and (c) and walking in (b) and (d). The box represents the 25th and 75th percentiles and the line dividing the box represents the median responses.

We asked participants what they liked and disliked about each technique. For Dwell time they stated, it is easy (9 participants), works better while walking (2 participants), takes longer than while sitting (5 participants), and causes some frustration (4 participants). P23 associated that with movements *"if I move while walking this alters the accuracy of the technique and creates frustration"*. Pursuits was reported to be easy to use or learn (8 participants), accurate (2 participants) and intuitive (2 participants) but suffer from inaccuracy as the number of targets increase as reported (3 participants). Four participants mentioned that performing selection using Pursuits becomes more demanding and requires much focus while walking. Gaze gestures were perceived to be easy to learn or use (5 participants) and fast (8 participants). P22 mentioned that Gaze gestures *"involved a slightly bigger learning curve"* (P22), while P7 noted that it *"was quick and also I get a chance to look ahead while walking"*. On the downside, Gaze gestures was reported to be tiring to eyes by 9 participants.

Observation 8: In the walking state, both Pursuits and Dwell time are perceived to be easier to learn and more natural compared to Gaze gestures. In terms of the natural to use aspect, the difference was significant between Dwell time and Gaze gestures.

Ranking the techniques: We asked participants to rank their preference for the input techniques when performing selection in the walking state. Regardless of the number of targets, Dwell time was the most preferred one (Score = 55), while Gaze gestures was ranked as the least preferred (Score = 38). Pursuits was ranked second (Score = 51). Similarly, while considering the number of targets, Dwell time was the most preferred too. This result matches the preference for techniques with various targets (see Figure 8d). Results are summarised in Table 3. One participant (P2) ranked Dwell time first because the context is walking and needing to focus

Table 3: Participants had different preferences for techniques for sitting and walking states. When asked to rank their preference, the weighted score showed that Pursuits was the most preferred for few targets while Dwell time was preferred with more targets in sitting state (to left). For the walking state, Dwell time was the most preferred technique.

Sitting			
# of targets	Ranking	Technique	Weighted score
2	1	Pursuits	52
	2	Dwell time	47
	3	Gaze gestures	45
4	1	Pursuits	59
	2	Dwell time	45
	3	Gaze gestures	40
9	1	Pursuits	54
	2	Dwell time	51
	3	Gaze gestures	39
12	1	Dwell time	53
	2	Pursuits	47
	3	Gaze gestures	44
32	1	Dwell time	53
	2	Gaze gestures	46
	3	Pursuits	45

Walking			
# of targets	Ranking	Technique	Weighted score
2	1	Dwell time	52
	2	Pursuits	46
	2	Gaze gestures	46
4	1	Dwell time	53
	2	Pursuits	48
	3	Gaze gestures	43
9	1	Dwell time	52
	2	Pursuits	50
	3	Gaze gestures	42
12	1	Dwell time	51
	1	Pursuits	51
	3	Gaze gestures	42
32	1	Dwell time	52
	2	Pursuits	48
	3	Gaze gestures	44

on the way. Most participants who ranked Dwell time first found it easier to use and more natural, *"Dwell time felt more natural"* (P13), while those who ranked Pursuits first mostly mentioned performance, *"I think I based on my experience and performance. With Pursuits, I wasn't frustrated and could select most perfectly if I remember"*(P17). Accuracy was the main drive behind some participants to rank Gaze gestures the first, *"Gaze gestures even though unnatural would give you the result you want"* (P23).

Observation 9: Participants had different preferences for techniques based on sitting and walking states: Pursuits ranked as the preferred input technique in sitting and Dwell time was ranked first in walking.

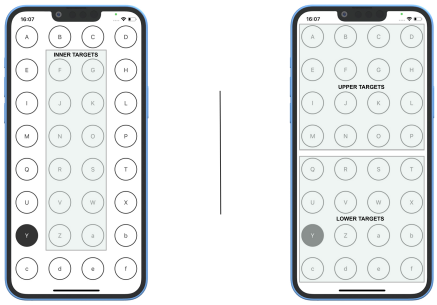


Figure 9: Left: Inner targets are highlighted while all the other targets are considered as borders. Right: Upper and lower targets are highlighted when 32 targets are presented. Targets are highlighted in green for illustration purposes, and in each trial of the study, the targets to be selected were random.

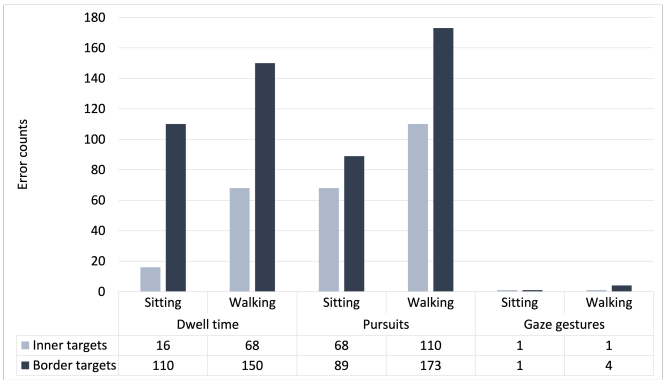
6 Discussion and Future Work

In our implementation of the three commonly used gaze-based interaction selection methods, Pursuits was found to be faster than Dwell time and Gaze gestures in both the sitting and walking states, especially when more targets are displayed. Pursuits was highly ranked by participants as the most preferred technique to be used when stationary regardless of the number of targets. Dwell time, on the other hand, although slower, was preferred when walking.

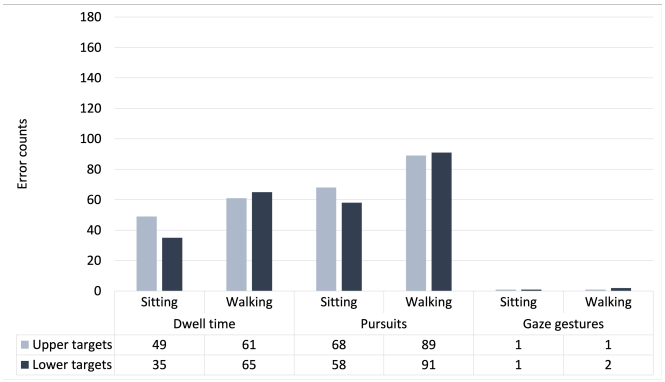
In addition to the aforementioned novel insights, our results also confirm that a number of findings from prior work hold for gaze interaction on mobile devices. Prior work that compared gaze input on head-mounted displays also reported that participants found that Dwell time requires less exertion than motion matching [27]. While only a few participants ranked Gaze gestures as their most preferred input method in our study, our results, as was suggested in the literature [24], showed that Gaze gestures are highly accurate.

6.1 Challenges of Gaze Interaction

6.1.1 Gaze Interaction while on the Move Overall, all selection techniques take longer time for selection and are less accurate while walking. This is due to the shaky environment, which in turns impacts the quality of the face images that are analysed to estimate the gaze point. Most affected by this was Dwell time as it relies on calibration and accurate gaze estimation. Pursuits and Gaze gestures are also affected by this, albeit to a lesser degree. Even though they both do not require calibration or accurate gaze estimates, walking results in noise that reduces the chance of matching eye movements to trajectories and templates. Some participants reported that it was challenging to focus on their way as they walk while making selections at the same time. Future implementations of gaze input methods need to account for more distractions when providing input. A promising direction to compensate for this is to explore ways to save the states of the users' input (e.g., store the gaze points when looking away and resume the same input when the user looks at the same target again [30]). Some of our findings align with prior work on gaze interaction with public and head-mounted displays while on the move. For example, prior work on gaze interaction with public displays also showed that selections using Pursuits are slower while walking than while stationary [47]. In the context of



(a) Inner vs border targets



(b) Upper vs lower targets

Figure 10: For each trial, we allow participants to continue selecting until the correct target is selected or a timeout occurs. Overall, selecting border targets is more error-prone compared to inner targets (a). When comparing upper and lower targets when selecting one of 9, 12, and 32 targets, no particular trend was observed (b).

head-mounted displays, Pursuits was found to be more demanding when making selections while walking [49].

6.1.2 Face and Eye Visibility In our study, we detected several cases where participants failed to complete the selection tasks because their faces were not visible to the camera and therefore, resulting in gaze data loss. Assuming users' full faces are always visible when estimating gaze is not well suited for mobile settings [43]. Prior work suggested that users hold their phones in different ways and that hand postures are different across smartphones [35, 45]. Since most gaze estimation algorithms require full-face images for eye tracking to be reliable, the face and eye visibility issue become prominent [43]. Future research directions are to explore gaze estimation algorithms that rely on the eyes only for the mobile context [12, 35] or to investigate ways to inform users when tracking is lost or guide them to the best holding posture to maximise eye tracking accuracy. Similar concepts were proposed in the literature for guiding users in front of public displays to the ideal interaction position [8]; studying the applicability of these methods in the context of mobile devices is promising.

6.1.3 Calibration in Mobile Settings Gyroscope data collected during the experiment revealed variations in participants' holding posture. We also noticed frequent changes in gyroscope data in the same session, suggesting a lack of uniformity in phone-holding posture throughout interaction sessions. These changes affect the performance of gaze input methods especially Dwell time as it relies on accurate gaze estimates. Even if held in the same way, the inevitable shaking of the device could make the calibration data obsolete and would make re-calibration necessary [10].

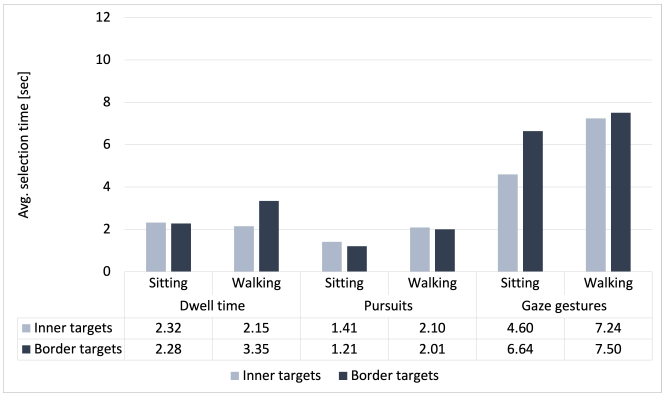
Some proposals were made to mitigate such issues. One approach is to leverage the devices' inertial sensors such as gyroscope and accelerometer to decide what frames or images to use for gaze estimation [55, 66]. A possible direction for future work is to automatically compensate for the changed posture using the internal sensing data or camera to update calibration parameters without the need for re-calibration. Another approach is to use gaze input methods that do not require calibration such as Pursuits and Gaze gestures.

6.2 Impact of Target and Camera Positions

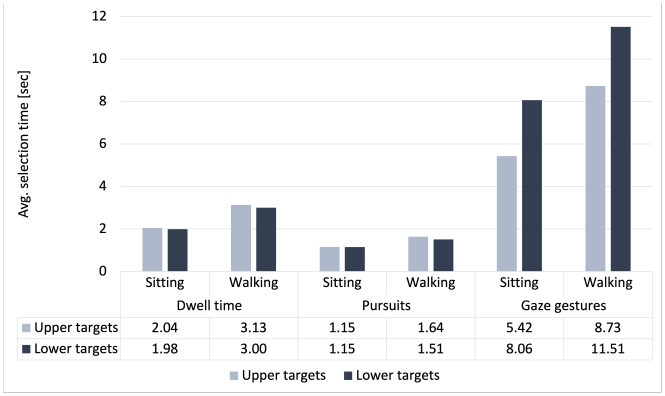
We noticed interesting differences in the number of errors when performing selection between targets located on the border of the screen and inner targets (see Figure 9 and 10). Selecting border targets using Dwell time is more error-prone compared to inner ones in both the sitting and walking states, where participants made a total of 84 errors when selecting one of the inner targets, while they made 260 errors when selecting from the border targets. In terms of selection time, the average selection time for Dwell time in walking state when selecting inner targets was 2.15 *sec* while it was 3.35 *sec* when selecting border targets (see Figure 11a). As mentioned earlier, this could be as result of inaccurate gaze estimation due to mobile settings. Five participants noticed how sensitive Dwell time is to shaky environments and that calibration data might just deteriorate over time. The negative effect of border targets on selection time was also observed when using Gaze gestures while sitting (see Figure 11a). Prior work comparing selection mechanism for gaze input techniques on head-mounted displays also showed an increase in selection time on the corner over the center targets when using Dwell time while the effect of target location was reduced when using motion matching [27]. Additional research is needed to assess the impact of target positions on gaze input in mobile settings.

On the other hand, when comparing targets located at the upper part of the screen and the lower part when selecting one of 9, 12, or 32 targets, the selection time increased on the lower targets using Gaze gestures in both the sitting and walking states (see Figure 9 and 11b).

We found that the tracking accuracy is worse towards the left edge of the screen compared to the right edge where in 119 trials, participants made 265 errors whereas, with targets on the right edge, participants made 201 errors in 165 trials. This could be due to the camera's position on the smartphone, which is on the center-right of the top of the device in our experiment. This warrants future work to investigate how the camera's position influences gaze selection.



(a) Inner vs border targets



(b) Upper vs lower targets

Figure 11: Regardless of the number of targets, selection time increased for border compared to inner targets using Dwell time while walking and also when using Gaze gestures while sitting (a). When comparing upper and lower targets when selecting one of 9, 12, or 32 targets, the selection time increased using Gaze gestures in both the sitting and walking states (b).

6.3 Guidelines for Gaze Interaction on Mobile Devices

Based on our results, we recommend the following:

[1] Use Pursuits if users are expected to use the phone while stationary and there are < 9 targets. Gaze gestures is also suitable when there are few targets (e.g., 2 targets). However, it requires a longer learning curve.
[2] Dwell time should be used when there are > 9 targets while stationary and while on the go. While Pursuits performs well in terms of selection time, it is demanding and distracting when there are many targets.
[3] Use Gaze gestures when accuracy is more important than speed in both sitting and walking states.
[4] Allow users to opt for alternative techniques depending on the context and number of targets.

7 Conclusion

In this work, we compared three of the most commonly used gaze interaction methods in mobile settings; while sitting and while walking: Dwell time, Pursuits, and Gaze gestures using quantitative and qualitative measures. We found that input using Pursuits is faster than Dwell time and Gaze gestures. When there are many targets, Pursuits is particularly faster but also more distracting to users. Users prefer Pursuits when stationary, but prefer Dwell time when walking. While selection using Gaze gestures is more demanding and slower when there are many targets, it is suitable for contexts where accuracy is more important than speed. Based on the analysis of our results and on prior work, We concluded with guidelines for the design of gaze interaction on handheld mobile devices.

Acknowledgments

This work has been funded by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK EPSRC under grant number EP/S035362/1, and an EPSRC New Investigator award (EP/V008870/1), the Islamic university of Madinah, the Royal Society of Edinburgh (RSE) grant number 1931, and the Fundação para a Ciência e a Tecnologia and LARSyS (grant UIDB/50009/2020).

References

[1] 2022. Human Interface Guideline - Apple. Webpage. https://developer.apple.com/design/human-interface-guidelines/foundations/app-icons/ accessed 15 September 2022.
[2] 2022. Look to Speak. Webpage. https://experiments.withgoogle.com/looktospeak accessed 13 September 2022.
[3] Yasmeen Abdrabou, Mohamed Khamis, Rana Mohamed Eisa, Sherif Ismail, and Amrl Elmougy. 2019. Just Gaze and Wave: Exploring the Use of Gaze and Gestures for Shoulder-Surfing Resilient Authentication. 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 29, 10 pages. https://doi.org/10.1145/3314111.3319837
[4] Yasmeen Abdrabou, Ken Pfeuffer, Mohamed Khamis, and Florian Alt. 2020. GazeLockPatterns: Comparing Authentication Using Gaze and Touch for Entering Lock Patterns. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 29, 6 pages. https://doi.org/10.1145/3379156.3391371
[5] Yasmeen Abdrabou, Johannes Schütte, Ahmed Shams, Ken Pfeuffer, Daniel Buschek, Mohamed Khamis, and Florian Alt. 2022. "Your Eyes Tell You Have Used This Password Before": Identifying Password Reuse from Gaze and Keystroke Dynamics. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 400, 16 pages. https://doi.org/10.1145/3491102.3517531
[6] Yasmeen Abdrabou, Ahmed Shams, Mohamed Omar Mantawy, Anam Ahmad Khan, Mohamed Khamis, Florian Alt, and Yomna Abdelrahman. 2021. GazeMeter: Exploring the Usage of Gaze Behaviour to Enhance Password Assessments. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 9, 12 pages. https://doi.org/10.1145/3448017.3457384
[7] Sunggeun Ahn, Jeongmin Son, Sangyoon Lee, and Geehyuk Lee. 2020. VergeIt: Gaze Interaction for a Binocular Head-Worn Display Using Modulated Disparity Vergence Eye Movement. 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, 1–7. https://doi.org/10.1145/3334480.3382908
[8] Florian Alt, Andreas Bulling, Gino Gravanis, and Daniel Buschek. 2015. GravitySpot: Guiding Users in Front of Public Displays Using On-Screen Visual Cues. 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, 47–56. https://doi.org/10.1145/2807442.2807490

[9] Mihai Bâce, Alia Saad, Mohamed Khamis, Stefan Schneegass, and Andreas Bulling. 2022. PrivacyScout: Assessing Vulnerability to Shoulder Surfing on Mobile Devices. *Proc. Priv. Enhancing Technol.* 2022, 3 (2022), 650–669. https://doi.org/10.56553/popets-2022-0090

[10] Michael Barz, Florian Daiber, Daniel Sonntag, and Andreas Bulling. 2018. Error-Aware Gaze-Based Interfaces for Robust Mobile Gaze Interaction. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, Article 24, 10 pages. https://doi.org/10.1145/3204493.3204536

[11] Mathieu Beraneck, François M. Lambert, and Soroush G. Sadeghi. 2014. Chapter 15 - Functional Development of the Vestibular System: Sensorimotor Pathways for Stabilization of Gaze and Posture. In *Development of Auditory and Vestibular Systems*, Raymond Romand and Isabel Varela-Nieto (Eds.). Academic Press, San Diego, 449–487. https://doi.org/10.1016/B978-0-12-408088-1.00015-4

[12] Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus H Gross. 2014. High-quality capture of eyes. *ACM Trans. Graph.* 33, 6 (2014), 223–1.

[13] Andreas Bulling. 2016. Pervasive attentive user interfaces. *Computer* 49, 01 (2016), 94–98.

[14] Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2012. Increasing the Security of Gaze-Based Cued-Recall Graphical Passwords Using Saliency Masks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 3011–3020. https://doi.org/10.1145/2207676.2208712

[15] Andreas Bulling and Hans Gellersen. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing* 9, 4 (October 2010), 8–12. https://doi.org/10.1109/MPRV.2010.86

[16] Andreas Bulling and Hans Gellersen. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing* 9, 4 (2010), 8–12. https://doi.org/10.1109/MPRV.2010.86

[17] Xiuli Chen, Aditya Acharya, Antti Oulasvirta, and Andrew Howes. 2021. An Adaptive Model of Gaze-Based Selection. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 288, 11 pages. https://doi.org/10.1145/3411764.3445177

[18] Dietlind Helene Cymek, Antje Christine Venjakob, Stefan Ruff, Otto Hans-Martin Lutz, Simon Hofmann, and Matthias Roetting. 2014. Entering PIN codes by smooth pursuit eye movements. *Journal of Eye Movement Research* 7, 4 (May 2014). https://doi.org/10.16910/jemr.7.4.1

[19] Murtaza Dhuliawala, Juyoung Lee, Junichi Shimizu, Andreas Bulling, Kai Kunze, Thad Starner, and Woontack Woo. 2016. Smooth Eye Movement Interaction Using EOG Glasses. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (Tokyo, Japan) *(ICMI '16)*. Association for Computing Machinery, New York, NY, USA, 307–311. https://doi.org/10.1145/2993148.2993181

[20] Connor Dickie, Roel Vertegaal, Changuk Sohn, and Daniel Cheng. 2005. EyeLook: Using Attention to Facilitate Mobile Media Consumption. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (Seattle, WA, USA) *(UIST '05)*. Association for Computing Machinery, New York, NY, USA, 103–106. https://doi.org/10.1145/1095034.1095050

[21] Heiko Drewes, Alexander De Luca, and Albrecht Schmidt. 2007. Eye-Gaze Interaction for Mobile Phones. In *Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology* (Singapore) *(Mobility '07)*. Association for Computing Machinery, New York, NY, USA, 364–371. https://doi.org/10.1145/1378063.1378122

[22] Heiko Drewes, Mohamed Khamis, and Florian Alt. 2019. DialPlates: Enabling Pursuits-Based User Interfaces with Large Target Numbers. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia* (Pisa, Italy) *(MUM '19)*. Association for Computing Machinery, New York, NY, USA, Article 10, 10 pages. https://doi.org/10.1145/3365610.3365626

[23] Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the computer using gaze gestures. In *Ifip conference on human-computer interaction*. Springer, 475–488.

[24] Morten Lund Dybdal, Javier San Agustin, and John Paulin Hansen. 2012. Gaze Input for Mobile Devices by Dwell and Gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 225–228. https://doi.org/10.1145/2168556.2168601

[25] Hanene Elleuch, Ali Wali, and Adel M. Alimi. 2014. Smart Tablet Monitoring by a Real-Time Head Movement and Eye Gestures Recognition System. In *2014 International Conference on Future Internet of Things and Cloud*. 393–398. https://doi.org/10.1109/FiCloud.2014.70

[26] Hanene Elleuch, Ali Wali, Anis Samet, and Adel M Alimi. 2016. A real-time eye gesture recognition system based on fuzzy inference system for mobile devices monitoring. In *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 172–180.

[27] Augusto Esteves, Yonghwan Shin, and Ian Oakley. 2020. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies* 139 (2020), 102414.

[28] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. 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, 457–466. https://doi.org/10.1145/2807442.2807499

[29] Augusto Esteves, David Verweij, Liza Suraiya, Rasel Islam, Youryang Lee, and Ian Oakley. 2017. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 167–178. https://doi.org/10.1145/3126594.3126616

[30] Misahael Fernandez, Florian Mathis, and Mohamed Khamis. 2020. *GazeWheels: Comparing Dwell-Time Feedback and Methods for Gaze Input*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3419249.3420122

[31] Denzil Ferreira, Jorge Goncalves, Vassilis Kostakos, Louise Barkhuus, and Anind K. Dey. 2014. Contextual Experience Sampling of Mobile Application Micro-Usage. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 91–100. https://doi.org/10.1145/2628363.2628367

[32] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[33] Henna Heikkilä and Kari-Jouko Räihä. 2012. Simple Gaze Gestures and the Closure of the Eyes as an Interaction Technique. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 147–154. https://doi.org/10.1145/2168556.2168579

[34] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. 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/10.1145/3290605.3300855

[35] Qiong Huang, Ashok Veeraraghavan, and Ashutosh Sabharwal. 2017. TabletGaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications* 28, 5 (2017), 445–461.

[36] Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze Gestures or Dwell-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, 229–232. https://doi.org/10.1145/2168556.2168602

[37] Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. 2010. Designing Gaze Gestures for Gaming: An Investigation of Performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 323–330. https://doi.org/10.1145/1743666.1743740

[38] Robert J. K. 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* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[39] Shahram Jalaliniya and Diako Mardanbegi. 2016. EyeGrip: Detecting Targets in a Series of Uni-Directional Moving Objects Using Optokinetic Nystagmus Eye Movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5801–5811. https://doi.org/10.1145/2858036.2858584

[40] Zhiping Jiang, Jinsong Han, Chen Qian, Wei Xi, Kun Zhao, Han Ding, Shaojie Tang, Jizhong Zhao, and Panlong Yang. 2016. VADS: Visual attention detection with a smartphone. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2016.7524398

[41] Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. 2014. Gaze Gestures and Haptic Feedback in Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 435–438. https://doi.org/10.1145/2556288.2557040

[42] Christina Katsini, Yasmeen Abdrabou, George Raptis, Mohamed Khamis, and Florian Alt. 2020. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions.. In *Proceedings of the 38th Annual ACM Conference on Human Factors in Computing Systems* (Honolulu, Hawaii, USA) *(CHI '20)*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3313831.3376840

[43] Mohamed Khamis, Florian Alt, and Andreas Bulling. 2018. The past, present, and future of gaze-enabled handheld mobile devices: Survey and lessons learned. In *Proceedings of the 20th International Conference on Human-Computer Interaction*

with Mobile Devices and Services. 1–17.

[44] Mohamed Khamis, Florian Alt, Mariam Hassib, Emanuel von Zezschwitz, Regina Hasholzner, and Andreas Bulling. 2016. GazeTouchPass: Multimodal Authentication Using Gaze and Touch on Mobile Devices. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 2156–2164. https://doi.org/10.1145/2851581.2892314

[45] Mohamed Khamis, Anita Baier, Niels Henze, Florian Alt, and Andreas Bulling. 2018. Understanding Face and Eye Visibility in Front-Facing Cameras of Smartphones Used in the Wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173854

[46] Mohamed Khamis, Mariam Hassib, Emanuel von Zezschwitz, Andreas Bulling, and Florian Alt. 2017. GazeTouchPIN: Protecting Sensitive Data on Mobile Devices Using Secure Multimodal Authentication. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI 2017)*. ACM, New York, NY, USA, 446–450. https://doi.org/10.1145/3136755.3136809

[47] Mohamed Khamis, Alexander Klimczak, Martin Reiss, Florian Alt, and Andreas Bulling. 2017. EyeScout: Active Eye Tracking for Position and Movement Independent Gaze Interaction with Large Public Displays. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software & Technology* (Quebec City, QC, Canada) *(UIST '17)*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3126594.3126630

[48] Mohamed Khamis, Karola Marky, Andreas Bulling, and Florian Alt. 2022. User-centred multimodal authentication: securing handheld mobile devices using gaze and touch input. *Behaviour & Information Technology* 41, 10 (2022), 2047–2069. https://doi.org/10.1080/0144929X.2022.2069597 arXiv:https://doi.org/10.1080/0144929X.2022.2069597

[49] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018. VRpursuits: Interaction in Virtual Reality Using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces* (Castiglione della Pescaia, Grosseto, Italy) *(AVI '18)*. Association for Computing Machinery, New York, NY, USA, Article 18, 8 pages. https://doi.org/10.1145/3206505.3206522

[50] Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. 2016. TextPursuits: Using Text for Pursuits-Based Interaction and Calibration on Public Displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 274–285. https://doi.org/10.1145/2971648.2971679

[51] Dominik Kirst and Andreas Bulling. 2016. On the Verge: Voluntary Convergences for Accurate and Precise Timing of Gaze Input. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1519–1525. https://doi.org/10.1145/2851581.2892307

[52] Andy Kong, Karan Ahuja, Mayank Goel, and Chris Harrison. 2021. EyeMU Interactions: Gaze+ IMU Gestures on Mobile Devices. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 577–585.

[53] Shinya Kudo, Hiroyuki Okabe, Taku Hachisu, Michi Sato, Shogo Fukushima, and Hiroyuki Kajimoto. 2013. Input Method Using Divergence Eye Movement. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 1335–1340. https://doi.org/10.1145/2468356.2468594

[54] Manu Kumar, Tal Garfinkel, Dan Boneh, and Terry Winograd. 2007. Reducing Shoulder-Surfing by Using Gaze-Based Password Entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania, USA) *(SOUPS '07)*. Association for Computing Machinery, New York, NY, USA, 13–19. https://doi.org/10.1145/1280680.1280683

[55] Zhenjiang Li, Mo Li, Prasant Mohapatra, Jinsong Han, and Shuaiyu Chen. 2017. iType: Using eye gaze to enhance typing privacy. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2017.8057233

[56] 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* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. https://doi.org/10.1145/1518701.1518758

[57] Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*. Springer, 39–65.

[58] Diako Mardanbegi, Tobias Langlotz, and Hans Gellersen. 2019. Resolving Target Ambiguity in 3D Gaze Interaction through VOR Depth Estimation. 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/10.1145/3290605.3300842

[59] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. 2015. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2953–2962. https://doi.org/10.1145/2702123.2702539

[60] Emilie Mollenbach, John Paulin Hansen, Martin Lillholm, and Alastair G. Gale. 2009. Single Stroke Gaze Gestures. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 4555–4560. https://doi.org/10.1145/1520340.1520699

[61] Emilie Møllenbach, Martin Lillholm, Alastair Gail, and John Paulin Hansen. 2010. Single Gaze Gestures. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 177–180. https://doi.org/10.1145/1743666.1743710

[62] Kevin Pfeffel, Philipp Ulsamer, and Nicholas H. Müller. 2019. Where the User Does Look When Reading Phishing Mails – An Eye-Tracking Study. In *Learning and Collaboration Technologies. Designing Learning Experiences*, Panayiotis Zaphiris and Andri Ioannou (Eds.). Springer International Publishing, Cham, 277–287.

[63] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. GazeTouch: 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* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[64] 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* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 373–383. https://doi.org/10.1145/2807442.2807460

[65] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[66] Carmelo Pino and Isaak Kavasidis. 2012. Improving mobile device interaction by eye tracking analysis. In *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 1199–1202.

[67] Thammathip Piumsomboon, Gun Lee, Robert W. Lindeman, and Mark Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 36–39. https://doi.org/10.1109/3DUI.2017.7893315

[68] Helen C Purchase. 2012. *Experimental human-computer interaction: a practical guide with visual examples*. Cambridge University Press.

[69] Vijay Rajanna, Adil Hamid Malla, Rahul Ashok Bhagat, and Tracy Hammond. 2018. DyGazePass: A gaze gesture-based dynamic authentication system to counter shoulder surfing and video analysis attacks. In *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*. 1–8. https://doi.org/10.1109/ISBA.2018.8311458

[70] Radiah Rivu, Yasmeen Abdrabou, Ken Pfeuffer, Augusto Esteves, Stefanie Meitner, and Florian Alt. 2020. StARe: Gaze-Assisted Face-to-Face Communication in Augmented Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 14, 5 pages. https://doi.org/10.1145/3379157.3388930

[71] 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

[72] Alia Saad, Dina Hisham Elkafrawy, Slim Abdennadher, and Stefan Schneegass. 2020. Are They Actually Looking? Identifying Smartphones Shoulder Surfing Through Gaze Estimation. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 42, 3 pages. https://doi.org/10.1145/3379157.3391422

[73] Selina Sharmin, Oleg Špakov, and Kari-Jouko Räihä. 2013. Reading On-Screen Text with Gaze-Based Auto-Scrolling. In *Proceedings of the 2013 Conference on Eye Tracking South Africa* (Cape Town, South Africa) *(ETSA '13)*. Association for Computing Machinery, New York, NY, USA, 24–31. https://doi.org/10.1145/2509315.2509319

[74] Chen Song, Aosen Wang, Kui Ren, and Wenyao Xu. 2016. EyeVeri: A secure and usable approach for smartphone user authentication. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2016.7524367

[75] FM Toates. 1974. Vergence eye movements. *Documenta Ophthalmologica* 37, 1 (1974), 153–214.

[76] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-device Gaze-supported Point-to-point Content Transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. ACM, New York, NY, USA, 19–26. https://doi.org/10.1145/2578153.2578155

[77] Eduardo Velloso, Marcus Carter, Joshua Newn, Augusto Esteves, Christopher Clarke, and Hans Gellersen. 2017. Motion Correlation: Selecting Objects by Matching Their Movement. *ACM Trans. Comput.-Hum. Interact.* 24, 3, Article 22 (apr 2017), 35 pages. https://doi.org/10.1145/3064937

[78] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 439–448. https://doi.org/10.1145/2493432.2493477

[79] Mélodie Vidal, Ken Pfeuffer, Andreas Bulling, and Hans W. Gellersen. 2013. Pursuits: Eye-Based Interaction with Moving Targets. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 3147–3150. https://doi.org/10.1145/2468356.2479632

[80] Erroll Wood and Andreas Bulling. 2014. EyeTab: Model-Based Gaze Estimation on Unmodified Tablet Computers. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 207–210. https://doi.org/10.1145/2578153.2578185

[81] Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21, 1 (2017), 173–186.

[82] Hui Zheng and Vivian Genaro Motti. 2018. Assisting Students with Intellectual and Developmental Disabilities in Inclusive Education with Smartwatches. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173924

[83] Huiyuan Zhou, Vinicius Ferreira, Thamara Alves, Kirstie Hawkey, and Derek Reilly. 2015. Somebody Is Peeking! A Proximity and Privacy Aware Tablet Interface. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 1971–1976. https://doi.org/10.1145/2702613.2732726