GazeBlend: Exploring Paired Gaze-Based Input Techniques for Navigation and Selection Tasks on Mobile Devices

Omar Namnakani[a,*], Yasmeen Abdrabou[b], Jonathan Grizou[a,c], and Mohamed Khamis[a]

[a]School of Computing Science, University of Glasgow, Scotland, UK, [b]Chair for Human-Centered Technologies for Learning, Technical University of Munich, Germany, [c]GrizAI, UK

ARTICLE HISTORY
Compiled February 19, 2026

ABSTRACT
The potential of gaze for hands-free mobile interaction is increasingly evident. While each gaze input technique presents distinct advantages and limitations, a combination can amplify strengths and mitigate challenges. We report on the results of a user study (N=24), in which we compared the usability and performance of pairing three popular gaze input techniques: Dwell time, Pursuits, and Gaze Gestures, for navigation and selection tasks while sitting and walking. Results show that pairing Gestures for navigation with either Dwell time or Pursuits for selection improves task completion time and rate compared to using either individually. We discuss the implications of pairing gaze input techniques, such as how Pursuits may negatively impact other techniques, likely due to the visual clutter it adds, how integrating Gestures for navigation reduces the chances of unintentional selections, and the impact of motor activity on performance. Our findings provide insights for effective gaze-enabled interfaces.

KEYWORDS
Gaze-based Interactions, Mobile Devices, Dwell Time, Pursuits, Gaze Gestures

1. Introduction

Gaze interaction on mobile devices is a growing research area exploring eye movements to control smartphones and tablets (Khamis, Alt, & Bulling, 2018a). It has the potential to become integral to daily device use (Drewes & Schmidt, 2007; Møllenbach, Hansen, & Lillholm, 2013). Advances in front-facing cameras and mobile processing power are revolutionising eye-tracking technologies (Huang, Veeraraghavan, & Sabharwal, 2017; Huynh, Balan, & Ko, 2022; Khamis, Alt, & Bulling, 2018a; Lei, He, Khamis, & Ye, 2023), making gaze a promising input modality, especially when touch interaction is hindered due to contextual factors, such as when walking, carrying objects, in cold temperatures, or when the screen is wet (Goncalves et al., 2017; Kane, Wobbrock, & Smith, 2008; Sarsenbayeva, van Berkel, Luo, Kostakos, & Goncalves, 2017).

Prior research has examined gaze for hands-free mobile interaction (Drewes, De Luca, & Schmidt, 2007; Kong, Ahuja, Goel, & Harrison, 2021a; Lei, Wang, et

*Corresponding Author Omar Namnakani (omar.namnakani@gmail.com)

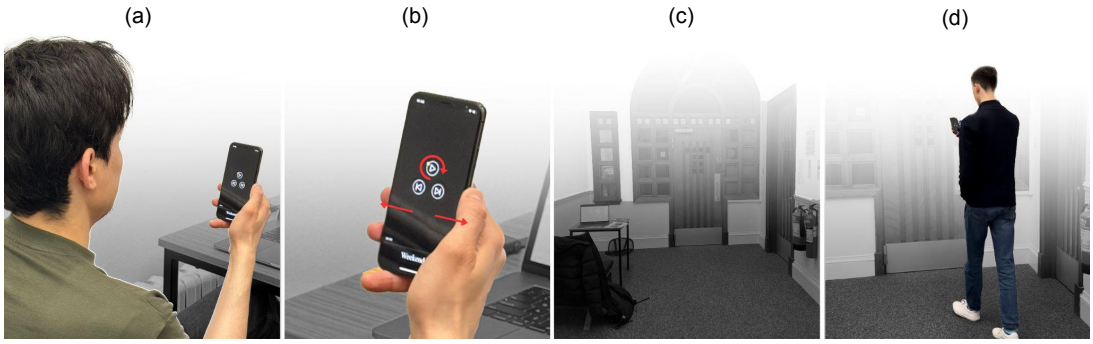

Figure 1. This work introduces and evaluates pairing gaze-based techniques: Dwell time, Pursuits, and Gestures for navigation and selection tasks on mobile devices, evaluating user perceptions within a single interface. We tested our techniques under two motor activities: while seated (a) and walking (d). (b) shows a participant navigating soundtracks via Gaze Gestures and selecting via Pursuits (arrows for illustration), and finally (c) shows the study location.

al., 2023; Namnakani, Abdrabou, Grizou, Esteves, & Khamis, 2023), exploring techniques based on distinct eye movements, such as Dwell time (Jacob, 1990; Ware & Mikaelian, 1986), Pursuits (Esteves, Velloso, Bulling, & Gellersen, 2015; Vidal, Bulling, & Gellersen, 2013; Vidal, Pfeuffer, Bulling, & Gellersen, 2013), and Gestures (Hyrskykari, Istance, & Vickers, 2012; Mollenbach, Hansen, Lillholm, & Gale, 2009; Møllenbach et al., 2013). The techniques have been evaluated in mobile contexts, such as walking (Drewes et al., 2007; Lei, Wang, et al., 2023; Namnakani et al., 2023). Each technique offers distinct advantages and limitations, with user preference and performance varying by context and task (Lei, Wang, et al., 2023; Namnakani et al., 2023). Dwell time, based on fixation, is well-known but demands accurate gaze estimation (Fernandez, Mathis, & Khamis, 2020; Jacob, 1990; Khamis, Alt, & Bulling, 2018a; Khamis, Buschek, Thieron, Alt, & Bulling, 2018; Khamis, Trotter, et al., 2018) and limits the number of displayed targets to compensate for tracking inaccuracy (Møllenbach et al., 2013). Techniques that rely on relative eye movements, such as Pursuits and Gestures, are promising for mobile contexts (Khamis, Alt, & Bulling, 2018a). Pursuits can activate small targets (Esteves et al., 2015) but require dynamic interfaces (Namnakani et al., 2023; Vidal, Bulling, & Gellersen, 2013; Vidal, Pfeuffer, et al., 2013). Gestures may be fast (Hyrskykari et al., 2012) and work with fewer targets (Namnakani et al., 2023), but lack visual cues (Møllenbach et al., 2013), rely on memorisation (Møllenbach, Lillholm, Gail, & Hansen, 2010), and require more gestures as commands increase, complicating interaction (Møllenbach et al., 2010).

Acknowledging each technique's advantages and limitations, we study how combining multiple gaze-based input techniques within the same interface can be leveraged to improve usability. We focus on pairing two gaze input techniques in the same interface, where we employ one technique for selection and another one for navigation. The topic is under-explored except for a few studies that split gaze-based selection into pointing and confirmation steps on a desktop computer, utilising different gaze input techniques for each (Schenk, Dreiser, Rigoll, & Dorr, 2017; Schenk, Tiefenbacher, Rigoll, & Dorr, 2016). Our work is the first to explore pairing gaze-input techniques for selection and navigation in general, and the first to examine the concept of pairing gaze input techniques on handheld mobile devices and in mobile settings, including while walking. We hypothesise that pairing gaze-based inputs, in this study, these are Dwell time, Pursuits and Gestures, allows for effective, efficient and well-perceived

interaction with mobile devices.

To evaluate our concept, we conducted a user study ($N = 24$) to assess the usability of six pairs of gaze-based techniques for selection and navigation tasks, in two motor activities: while sitting and while walking. We compared the pairs in terms of task completion time, task completion rates, error rates, and perceived workload using NASA-TLX. We also gathered participants' perceptions of pairing different eye movements through semi-structured interviews. In a task involving navigating a mobile home-screen-like interface to select a music player and play a given track, results show certain pairings improve the performance of gaze input, especially while walking. Using Gestures for navigation paired with Dwell time (19.1s) or Pursuits (21.6s) for selection was significantly faster than using Pursuits alone for both tasks (27.4s). Pairing also lowered error rates from (57.1%) (Pursuits only) to 19.6% (*DwellGestures*) and 32.3% (*PursuitsGestures*), and significantly improved completion rates. Participants generally appreciated pairing techniques rather than using one eye movement during the interaction.

Our findings suggest that pairing gaze-based input techniques can improve the effectiveness of gaze interactions on mobile devices for navigation and selection tasks. Based on our findings, we discuss and present the implications of pairing gaze-based input techniques when designing user interfaces for enhanced gaze-based interaction on mobile devices. For example, we discuss the underlying mechanism that may impact the performance of such pairing. We also discuss how, depending on the context, such as whether the user is walking or sitting, the studied techniques can be combined to maximise their strengths while minimising their limitations.

Contribution Statement

The contribution of our work is threefold. Firstly, we introduce and investigate the novel concept of pairing gaze-based input techniques within the same interface. Secondly, we provide an in-depth analysis of this approach by (a) evaluating user performance and perception and (b) presenting implications for enhancing gaze interactions through pairing input techniques. Finally, we explore the effect of motor activity (sitting vs. walking) on user preferences and performance, offering valuable insights into the contextual adaptability of gaze-based interactions.

2. Background and Related Work

Gaze has long been studied for interaction. We build on insights from work that studied individual eye movements in isolation for explicit interaction or together with other eye movement types, with a focus on handheld mobile devices.

2.1. Eye Movements and Gaze-based Interaction on Mobile Devices

Previous research in the HCI domain has categorised the implementation of gaze-based interaction methods according to various eye movements, broadly, fixations, smooth pursuits, and saccades (Møllenbach et al., 2013; Rantala et al., 2020).

Fixation refers to a state where the eyes appear motionless, though true fixations do not occur as microsaccades, drift, and tremors accompany them (Carpenter, 1988;

Møllenbach et al., 2013). Dwell time (Jacob, 1990), a fixation-based gaze input technique, is widely used for selecting static targets (Dybdal, Agustin, & Hansen, 2012; Li et al., 2021; Majaranta, Räihä, Hyrskykari, & Špakov, 2019; Namnakani et al., 2023). It involves a prolonged fixation to address the Midas touch problem (Jacob, 1990), with durations ranging from under 200 ms to over 1000 ms (Fernandez et al., 2020; Majaranta & Bulling, 2014; Nagamatsu, Yamamoto, & Sato, 2010; Rantala et al., 2020). Dwell interfaces limit target numbers due to tracking inaccuracy and micro-movements (Møllenbach et al., 2013). On mobile devices, Dwell time has been compared with other gaze techniques (Dybdal et al., 2012; Lei, Wang, et al., 2023; Namnakani et al., 2023), used for gaze-only interaction (Drewes et al., 2007), and combined with other modalities (Kong et al., 2021a; Nagamatsu et al., 2010).

Smooth pursuits, a fixation-in-motion, occur when tracking a moving object (Esteves, Bouquet, Pfeuffer, & Alt, 2022; Møllenbach et al., 2013; Vidal, Pfeuffer, et al., 2013). Pursuits input technique, based on this eye movements, compares the user's eye trajectory with that of the moving targets for selection (Esteves et al., 2015; Khamis et al., 2016; Vidal, Pfeuffer, et al., 2013). With Pursuits, accuracy drops and demands increase with more targets (Esteves et al., 2015; Isomoto, Yamanaka, & Shizuki, 2022). On mobile, Pursuits was the fastest compared with Dwell time and Gestures (Namnakani et al., 2023) but was less preferred when walking due to higher demands (Lei, Wang, et al., 2023).

Saccades, rapid ballistic eye movements between fixations (30–120 ms) (Carpenter, 1988; Møllenbach et al., 2013; Rantala et al., 2020), form the basis of gaze gestures (Hyrskykari et al., 2012; Namnakani et al., 2023; Schenk et al., 2016). Gestures range from single strokes to multi-stroke patterns (Majaranta et al., 2019; Mollenbach et al., 2009; Rantala et al., 2020), can be fast, require no visual stimuli, and scale with more commands (Mollenbach et al., 2009; Møllenbach et al., 2010; Møllenbach et al., 2013). However, complexity reduces memorability, may cause fatigue, and simple gestures risk accidental activation (Møllenbach et al., 2013; Namnakani et al., 2023). Gestures have been explored for mobile interaction (Dybdal et al., 2012; Kangas et al., 2014; Lei, Wang, et al., 2023; Namnakani et al., 2023; Rantala et al., 2020; Rozado, Moreno, San Agustin, Rodriguez, & Varona, 2015). Since both Gestures and Pursuits techniques rely on relative eye movements, they do not require appropriate calibration as the absolute gaze positions are less important (Khamis, Alt, & Bulling, 2018a; Møllenbach et al., 2013; Schenk et al., 2017).

Overall, Prior work shows that mobile contexts present unique challenges (Khamis, Alt, & Bulling, 2018a). Thus, enabling gaze input on mobile devices has led recent research to explore gaze-input techniques in mobile settings (Lei, Wang, et al., 2023; Namnakani et al., 2023), as findings from other settings, such as HMD (Khamis, Oechsner, Alt, & Bulling, 2018), public displays (Khamis, Klimczak, Reiss, Alt, & Bulling, 2017), and desktop (Rajanna, Malla, Bhagat, & Hammond, 2018) are not transferable to the context of mobile device.

2.2. Combination of Gaze-based Input Techniques

While there is work on combining gaze with other input modalities, such as mouse and/or keyboards (Fono & Vertegaal, 2005; Salvucci & Anderson, 2000), motion gestures (Kong, Ahuja, Goel, & Harrison, 2021b), voice (M. Zhao et al., 2022) or touch input (Pfeuffer, Alexander, Chong, & Gellersen, 2014; Pfeuffer, Alexander, Chong, Zhang, & Gellersen, 2015; Pfeuffer & Gellersen, 2016; Rivu, Abdrabou, Mayer, Pfeuf-

fer, & Alt, 2019; Stellmach & Dachselt, 2012), in this work, we focus on pairing gaze-based input techniques in the same interface. Prior work suggested combining gaze techniques to improve target selection, such as SPOCK and MSDT (Schenk et al., 2017, 2016), which used Pursuits or Dwell time depending on whether stimuli were dynamic or static, with stimuli appearing after a 500 ms dwell on a particular position. Ismoto et al. used a two-level stroke gesture for command activation and a short dwell for target selection (Isomoto, Yamanaka, & Shizuki, 2020), while Morimoto and Amir combined a short dwell to select a letter with a gesture across the opposite keyboard to activate it (Morimoto & Amir, 2010). Since Gestures can be performed without visual cues (Møllenbach et al., 2013), it has been suggested, though not explored, that combining techniques such as Dwell time and Gestures could strengthen gaze-based interaction (Heikkilä & Räihä, 2012; Møllenbach et al., 2010).

As demonstrated, prior work mainly split a single interaction into steps using different techniques, rather than combining multiple gaze-based techniques to facilitate various interactions. While this addresses the Midas-touch problem, it can be limiting. We hypothesise that integrating multiple gaze techniques on mobile devices can improve performance and usability by enabling more natural interactions. To test this, we explore our research questions in mobile contexts, both seated and walking.

2.3. Summary and Research Questions

While prior work has used different eye movements for pointing and selection, no studies have examined pairing gaze-based techniques within the same interface for different functions. Existing combinations have focused only on improving target acquisition by dividing selection into pointing and confirmation. Namnakani et al. (Namnakani et al., 2023) compared various gaze-based input techniques when used independently for selection tasks within user interfaces and studied the impact of the number of targets on the techniques. We build on this research by investigating how pairing gaze input techniques for different functionalities on the same interface impacts the usability and user perceptions of the techniques. We study that in the context of mobile devices, where users interact with the device while sitting or walking. Towards this end, we aim to answer the following questions: **RQ1:** How does pairing gaze-based techniques (Dwell time, Pursuits, Gestures) on a mobile interface affect performance, interaction efficiency and user perception compared to a single technique? And **RQ2:** How do motor activities (sitting vs. walking) influence performance and preferences with paired gaze-based techniques?

3. Pairing Gaze-based Input Techniques Concept and Implementation

To evaluate the usability of pairing different gaze-based techniques, we focus on two canonical user interface functions: *selection*: commonly performed on mobile touch interfaces via tapping for discrete selection, and *navigation*: typically executed using a single gesture by hand, i.e. swiping (Møllenbach et al., 2010) or tapping navigation buttons (Gonzalez-Calleros, Guerrero-Garcia, Vanderdonckt, & Munoz-Arteaga, 2009). We employ one gaze technique for selection and another for navigation. We developed an iPhone application using Xcode and Swift. Below, we list our implementation of the gaze-based inputs and how we pair them.

Dwell Time

In our implementation, selection requires an 800 ms fixation on the target. The dwell activation time was inspired by prior work (Chen, Acharya, Oulasvirta, & Howes, 2021; Fernandez et al., 2020; Namnakani et al., 2023; Niu, Gao, Zhang, Xue, & Yang, 2019; Zhang, Ren, & Zha, 2010). The timer is fired when one gaze point crosses a dwell target, and a selection is performed when the user maintains their gaze within the boundaries of the target within the time window. Rather than treating each gaze point as individually landing inside the target, we calculate the mean fixation point within each time window and determine if it lies inside the target, to compensate for tracking inaccuracy in the mobile context (Namnakani et al., 2023).

Pursuits

Our Pursuits implementation follows Namnakani et al. and Esteves et al. (Esteves et al., 2015; Namnakani et al., 2023). As Pursuits requires dynamic interfaces (Esteves et al., 2015; Vidal, Bulling, & Gellersen, 2013), users select targets by following moving objects. On the mobile's limited display, targets are shown as static circles, each with a small orbiting circle. Selection occurs by following the orbiting circle of the desired target (Esteves et al., 2015; Namnakani et al., 2023). We use the Pearson correlation coefficient (Esteves et al., 2015; Khamis, Oechsner, et al., 2018; Namnakani et al., 2023; Velloso et al., 2017; Vidal, Bulling, & Gellersen, 2013) to compare each target's horizontal and vertical trajectory with the x and y coordinates of the gaze data within a time window of one second (30-sample) (Esteves, Shin, & Oakley, 2020; Esteves et al., 2015; Namnakani et al., 2023). The lower correlation value between x and y is stored, and a target is selected if the stored correlation exceeds a threshold of 0.8, following practices used in prior work (Esteves et al., 2015; Namnakani et al., 2023). If not, correlation is recalculated with each new sample in a rolling window of 30 samples (Esteves et al., 2015; Namnakani et al., 2023). All orbiting circles move at 120°/sec, with half moving in opposite directions and all starting at different positions, following (Esteves et al., 2015; Namnakani et al., 2023).

Gaze Gestures

We used single-stroke right and left gestures, as they are efficient for basic navigation (Mollenbach et al., 2009; Møllenbach et al., 2013). A gesture had to be completed within 1000 ms (30 samples), with the Pearson correlation coefficient (Namnakani et al., 2023) used to compare the user's horizontal gaze path with gesture templates. Consistent with prior work (Namnakani et al., 2023) and based on pilot tests, the template with the highest correlation > 0.8 was taken as the intended gesture. To prevent unintentional gaze gesture activation by distinguishing natural horizontal eye movements from navigation gestures, we employed the screen bounds gazing to complete them, similar to prior work (Heikkilä & Räihä, 2012; Namnakani et al., 2023). By incorporating a 160 pixels-wide horizontal buffer at both the left and right screen edges, spanning the entire length of the device (Namnakani et al., 2023), the gesture is activated once the final horizontal gaze point crosses the buffer area, signaling completion of a gesture (Heikkilä & Räihä, 2012; Mollenbach et al., 2009; Namnakani et al., 2023).

Figure 2. A step-by-step showing how participants perform the task for the *DwellPursuits* condition. In the *DwellPursuits* condition, Dwell time handled selection and Pursuits handled navigation: (1) navigate from page 1 (left) to 2 (middle) via Pursuits in the home screen, (2) select the music player with Dwell time (middle), (3) navigate four screens to the target soundtrack via Pursuits (right), and (4) play it with Dwell time (a selection step). Navigation buttons remain on-screen for consistency, though unused with Gestures.

Pairing Gaze-based Techniques Prototype

We implemented a prototype application with different pairings of gaze-based techniques (see Figure 2). The interfaces mimic typical Android and iOS layouts. The home screen shows six selectable apps and two navigation targets (left and right). The music player provides three controls: play/pause, next, and previous. Although gestures can be used without visual targets (Møllenbach et al., 2013), all targets, including navigation buttons, were made visible for consistency. The number of home screen targets was limited to eight, as prior work found higher error rates with Dwell time and Pursuits when more targets are displayed (Esteves et al., 2015; Namnakani et al., 2023). We used a circular target with a diameter of 65 pt (1.08 cm or 2.09° at 29.6 cm viewing distance, measured with Eyedid during the experiment), corresponding to 195 pixels on the iPhone X used in the study. Our decision follows Apple design guidelines (Apple Inc., 2024a) and is inspired by prior work on gaze interaction on mobile devices (Namnakani et al., 2023). Targets were positioned closer to the centre of the display as previous work reported reduced accuracy toward screen edges (Hohlfeld, Pomp, Link, & Guse, 2015; Namnakani et al., 2023; Park, Park, & Cha, 2021). When integrating Pursuits and Gestures, we used iOS Swift's concurrency features, such as DispatchQueues and DispatchGroup, to process gaze data for each technique concurrently within a 1000 ms window. DispatchQueue handles the parallel execution, calculating correlations and returning the target with the highest correlation that exceeds a 0.8 threshold. DispatchGroup ensures that all processes are complete before selecting and executing the target. If both techniques exceed the threshold, we choose the target with the highest correlation. To integrate Dwell time with Pursuits or Gestures, we run each technique's processes concurrently on separate queues, while sharing a state that is updated upon input detection. Using DispatchQueue, we ensured exclusive access to this state. Upon detecting input, all processes are reset, and data collection starts again.

Table 1. The six conditions of gaze input techniques used for selection and navigation.

		Navigation		
		Dwell time	Pursuits	Gestures
Selection	Dwell time	*DwellDwell*	*DwellPursuits*	*DwellGestures*
	Pursuits	*PursuitsDwell*	*PursuitsPursuits*	*PursuitsGestures*
	Gestures	-	-	-

4. Evaluation

We conducted our study in a low-traffic corridor of a university building to examine the usability of pairing gaze-based selection and navigation tasks (see Figure 1). The study was approved by the University of Glasgow, College of Science and Engineering Ethics Committee (Approval No. 300220210).

4.1. Study Design

To study the impact of gaze-based input pairing in the same interface, we designed a $6x2$ within-subjects factorial design with the following two independent variables. **[IV1] Input techniques:** We tested six technique pairs: 1) Dwell time and Dwell time, 2) Dwell time and Pursuits, 3) Dwell time and Gestures, 4) Pursuits and Pursuits, 5) Pursuits and Dwell time, and 6) Pursuits and Gestures (see table 1). In each pair, the first technique was used for selection (e.g., activating the music player or playing a soundtrack), and the second for navigation (e.g., browsing apps or changing soundtracks). **[IV2] Motor activity:** We covered sitting and walking conditions.

4.2. Recruitment and Participants

Through social media, word of mouth, and posters across campus, we recruited 24 participants for the study (13 male, 11 female; ages 18–52, $M = 28.87$, $SD = 8.0$, 22 right-handed). Although none of the participants wore eyeglasses during the experiment, six reported nearsightedness, while three indicated they are farsighted. The participants indicated moderate use of phones while walking ($M = 3.08, SD = 1.10$) on a 5-point scale (0: never to 5: always). They also reported little to no experience with eye-tracking ($M = 0.96, SD = 1.63$) on a 5-point scale (0: no experience to 5: very experienced). The participants received a compensation of £10.

4.3. Apparatus and Task

We used an iPhone X running iOS 16.6 to run the prototype described in Section 3. The device features a front-facing camera with a 7-megapixel sensor. Eyedid SDK (Eyedid, 2023) was used to enable eye tracking on the mobile device. With a gaze tracking rate of 30 Hz (30 frames per second), the SDK provides real-time gaze estimation data in the form of (x,y) screen coordinates and reports an accuracy of 1.6°. The (x,y) values are provided even when the estimated gaze position falls outside the screen boundaries.

The Music player prototype consisted of 30 soundtracks downloaded from Pixabay.com and ordered randomly in each task. We chose random soundtracks without

showing the artists' names, rather than songs, to minimise the participants' familiarity with the songs. To avoid the possibility that participants might forget the soundtracks' titles during the tasks, we referred to them only by single-word titles.

The main task is to play a specific soundtrack given to the participant. This is conducted through four main steps:

Step 1: (Navigation), Swipe through a Home screen-like application to find the music player prototype (see Figure 2).
Step 2: (Selection), Open the music player prototype.
Step 3: (Navigation), Swipe through soundtracks to find the given soundtrack.
Step 4: (Selection), Play the given soundtrack once found.

In the sitting condition, participants sat in a relaxed posture. In the walking condition, participants were instructed to walk as they normally would when using a mobile device, holding it in whichever hand felt natural. During walking, they were guided to walk the length of the corridor from one end to the other, travelling back and forth without stopping. The experimenter monitored participants to ensure continuous movement during task performance.

4.4. Measures

To evaluate our input technique pairings, we measured task completion time, completion rate, error rate, and perceived workload (NASA-TLX (Hart, 2006)). Completion time is calculated as the duration from the Home screen to playing the correct soundtrack. Completion rate is calculated as the percentage of tasks completed within 60 seconds. Error rate is calculated as the proportion of incorrect actions to total actions (e.g., 8 errors in 15 actions = 53.3%). We also collected qualitative feedback through semi-structured interviews and Likert-scale questions on usability.

4.5. Procedure

Upon arrival, participants were briefed, signed the consent, and provided demographics, including eye-tracking experience. They received a demonstration of the interface and task, then completed the study tasks (Section 4.3). A calibration phase preceded all techniques to ensure consistent tracking, even for Pursuits and Gestures where it was not required (Dybdal et al., 2012; Khamis, Alt, & Bulling, 2018a, 2018b; Namnakani et al., 2023; Vidal, Bulling, & Gellersen, 2013). Prior to their participation in the study, participants were shown the designated location and the specific path they were to follow during the walking tasks to acquaint them with the environment. **Before each condition,** participants viewed a screen containing detailed information about the task to be performed, including the gaze input techniques to use, the name of the application to open, the soundtrack's title to play, and whether to perform the task sitting or walking. **In each condition**, participants completed one training trial, which was similar to the main task, with the only difference being the soundtrack titles, to familiarise themselves with the task and techniques. To focus on selection performance rather than search time (Namnakani et al., 2023), the music player was placed in the same location for all runs (steps 1 and 2). Training helped participants locate the app so search time would not affect analysis. Soundtrack titles for steps 3 and 4 were randomly chosen but consistently positioned. Participants completed three trials per condition, each requiring two selections (activating the music player

and playing the soundtrack) and five navigation steps (one step to move to the second page in the Home screen, and four steps to find the soundtrack), see Figure 2. Incorrect selections of apps triggered a one-second alert and were logged, while playing incorrect soundtracks were logged silently. The trial began when the participant was presented with the Home screen in the prototype and ended when the correct soundtrack was played, prompting a completion message. Trials were unsuccessful and also ended if the correct soundtrack was not played within 60 seconds or if the next correct action was not taken within 20 seconds at any stage (Namnakani et al., 2023). This approach was adopted to limit the experiment duration and minimise participant fatigue. The starting and ending times of each trial were automatically logged using timestamps from the phone. Feedback is provided to the participant whenever a button is activated by changing the button's colour to yellow for one second. **After each condition,** participants completed the NASA-TLX questionnaire and a 5-point Likert scale to reflect on various usability aspects. **At the end of the experiments,** participants filled out a questionnaire to rank the paired techniques based on their preference, and we conducted a semi-structured interview. Participants completed a total of 36 trials. All conditions were counterbalanced using Latin square to counteract order effects.

4.6. Limitations

In our study, participants located the music player app during the training step (see Section 4.5), thereby eliminating the need for a search step during the main tasks. While this design decision may not reflect real-world scenarios, it was essential to minimise search time, which could vary regardless of the input technique, as we did not want longer searches to be mistaken for slower gaze input. We did not study the impact of the search phase on gaze techniques as this was already explored in prior work by Namnakani et al. (Namnakani et al., 2023). The researchers explored Dwell time, Pursuits, and Gestures for target selection using various numbers of targets in two motor activities: sitting and walking. In a task in which participants had to locate and select discrete targets, gaze gestures was reported as the most accurate input technique. Dwell time was more accurate than Pursuits as the number of targets increased, especially when walking.

Although participants in our experiment did not face obstacles typical of real-world settings, conducting the study in a corridor may have introduced partially controllable, yet unpredictable, factors. For example, distractions such as people passing through or interacting with the experimenter could momentarily disrupt concentration; however, no such incidents occurred, and we were prepared to repeat any trials that were notably disrupted if necessary. We intentionally chose an indoor environment with minimal distractions to evaluate the performance of pairing the techniques in isolation, yet this may not fully reflect the wide range of real-world contexts where gaze interaction can be used, such as outdoor environments, crowded spaces, or varying lighting conditions. While these contextual factors can be investigated, in our experiment, we controlled them to mitigate the effects of confounding variables on results, while still ensuring that the study was conducted under natural usage conditions to maximise the relevance and applicability of our findings.

We conducted the study on an iPhone X, released in 2017. Since pairing gaze techniques required running multiple threads concurrently to detect inputs via Dispatch Queues, this may have introduced additional computational load due to the hardware's age. While we did not explicitly monitor CPU usage or latency, we did not observe

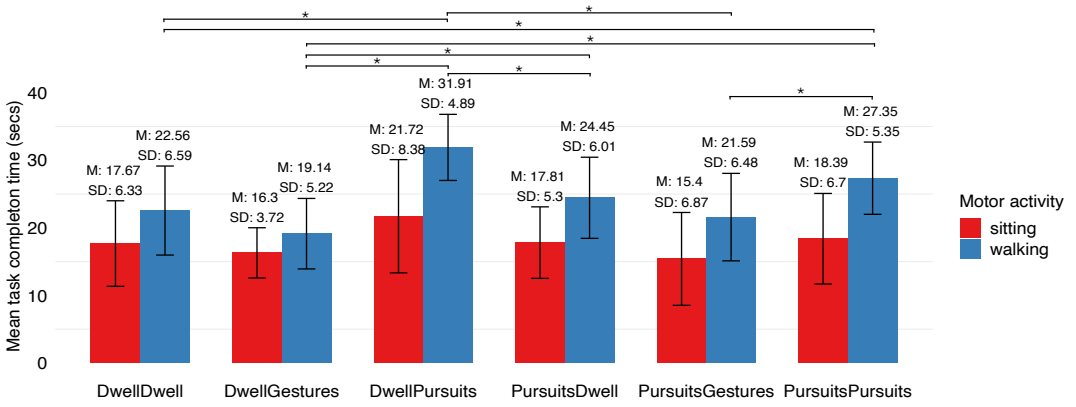

Figure 3. This figure represents the mean task completion time in seconds. No significant differences were observed between the techniques during seated motor activity. While walking revealed significant differences between several pairs, *DwellGestures* and *PursuitsGestures* showed faster completion time than *PursuitsPursuits*. The error bars represent the standard deviation.

any latency or delays during the pilot testing and the experiment, nor was it reported as such by participants. While performance may vary across newer mobile devices, in our study, all paired techniques were evaluated under identical hardware and software conditions in the same controlled environment. Thus, any performance limitations are expected to have affected all techniques equally.

Finally, our results are based on tracking data from the Eyedid library on mobile devices. While accuracy may vary with other software, our focus was on the conceptual understanding of user interfaces, specifically, how participants perceive paired gaze input techniques. Prior work suggests that tracking quality is influenced more by individual user differences than by the tracker or lighting (Feit et al., 2017).

5. Results

For the statistical tests, unless specified otherwise, we analysed the collected measures using repeated measures ANOVA with post-hoc pairwise comparisons and Bonferroni correction. For sphericity violations, we report Greenhouse-Geisser-corrected degrees of freedom and p-values. Where data transformation was required, we applied the Aligned Rank Transform (ART) (Wobbrock, Findlater, Gergle, & Higgins, 2011). If participants failed to complete all the trials within 60 seconds for a condition, task time was replaced with the mean of the remaining values for that condition (Namnakani et al., 2023). Due to a device error, one trial for one participant in one condition was lost, leaving two trials for that condition and a total of $24 \times 12 \times 3 - 1 = 863$ measurements.

5.1. Task Completion Time

Figure 3 summarises the descriptive statistics and shows the significant differences between pairs. The statistical test revealed a two-way interaction between Input techniques and Motor activity on task completion time, $F_{3.99,91.74}$=2.79, $p < 0.05$. While we found no significant differences between the input techniques in the sitting motor activity on the task completion time, when walking, the post hoc analysis showed that *DwellGestures* ($M = 19.1s, SD = 5.22$, $p < .001$) and *PursuitsGes-*

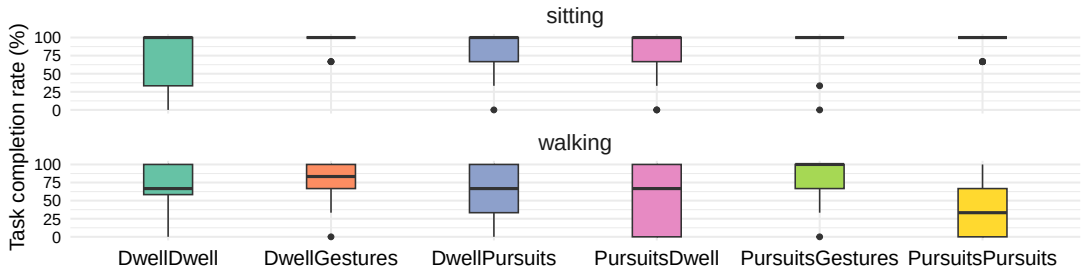

Figure 4. The figure illustrates the task completion rate by input techniques and motor activities. *DwellGestures* and *PursuitsGestures* resulted in a significantly higher completion rate compared to *PursuitsPursuits* when walking.

tures ($M = 21.6s, SD = 6.48$, $p < .05$) are significantly faster than *PursuitsPursuits* ($M = 27.4s, SD = 5.35$). *DwellGestures* also had significantly faster completion time than *DwellPursuits* ($M = 31.91s, SD = 4.9$, $p < .0001$) and *PursuitsDwell* ($M = 24.5s, SD = 6.01$, $p < .05$). Using t-test, we also looked at the impact of the motor activity on the techniques' completion time. We found that *DwellDwell*, *DwellPursuits*, *PursuitsDwell*, *PursuitsGestures*, and *PursuitsPursuits* are significantly slower when performing the task while walking (all $t(23) <= -2.867, p < .01$). This implies that there is not enough evidence to suggest that changing the motor activity results in additional time when using *DwellGestures*, $t(23) <= -2.008, p > .05$.

Overall, *DwellGestures* reduces task completion time without clear evidence of whether the motor activity impacts its performance.

5.2. Task Completion Rate

Figure 4 summarises the descriptive statistics. The statistical test revealed a two-way interaction between input techniques and motor activity on task completion rate, $F_{5,253} = 5.13$, $p < .001$. While the Friedman test revealed a statistically significant variation in the task completion rate between the input techniques while sitting, $\chi^2(2) = 18.05$, $p = .003$, post hoc analyses using Wilcoxon signed-rank tests, adjusted with a Bonferroni correction, did not show notable differences between the pairs. When walking, there were significant differences between the techniques, $\chi^2(2) = 18.093$, $p = .003$. Post hoc analysis revealed significant differences between *PursuitsPursuits* ($Med = 33.3\%, IQR = 66.7\%$) and both *PursuitsGestures* ($Med = 100\%, IQR = 33.3\%$) ($p < .01$) and *DwellGestures* ($Med = 83.3\%, IQR = 33.3\%$) ($p < .05$). This suggests that pairing Gestures with either Dwell time or Pursuits significantly enhances the task completion rate compared to using Pursuits alone.

5.3. Task Error Rate

Figure 5 summarises the descriptive statistics and details significant differences between multiple pairs. We found a statistically significant two-way interaction between Input techniques and Motor activity on task error rate, $F_{3.37,\ 77.52} = 3.83$, $p < 0.05$. When seated, post hoc analysis revealed that *DwellGestures* is significantly less error-prone ($M = 9.98\%, SD = 12.1$) compared to *DwellPursuits* ($p < .01$), *PursuitsDwell* ($p < .01$), *PursuitsGestures* ($p < .05$), and *PursuitsPursuits* ($p < .0005$). When

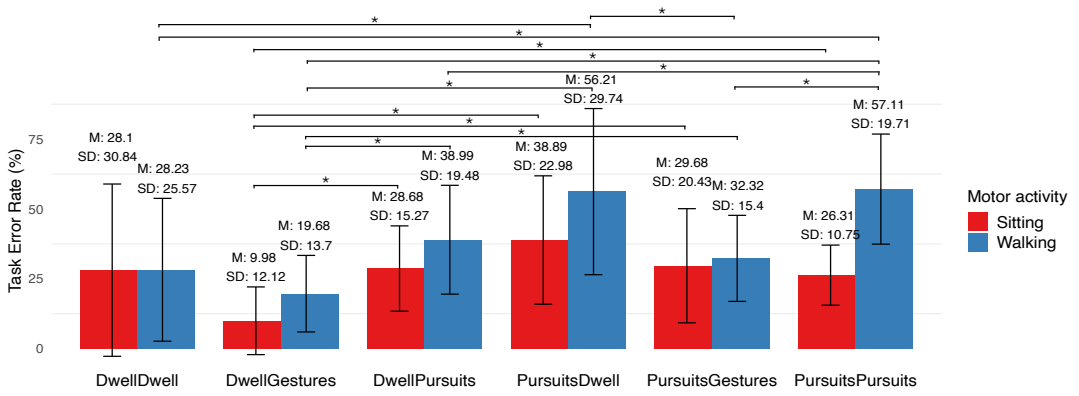

Figure 5. Sitting and walking task error rates. For tasks performed while seated, *DwellGestures* demonstrated significantly higher accuracy than *DwellPursuits*, *PursuitsDwell*, *PursuitsGestures*, and *PursuitsPursuits*. In contrast, during walking, *DwellDwell*, *DwellGestures*, and *PursuitsGestures* significantly outperformed *PursuitsPursuits* in terms of accuracy. The error bars represent the standard deviation.

walking, *PursuitsGestures* is significantly more accurate ($M = 32.3\%, SD = 15.4$) compared to *PursuitsPursuits* ($M = 57.1\%, SD = 19.7$). The analysis shows that *DwellGestures*, *DwellPursuits*, *PursuitsDwell* (all $p < .05$) and *PursuitsPursuits* ($p < .0001$) are substantially affected by motor activities.

Overall, *DwellGestures* led to a significantly lower error rate than any paired techniques involving Pursuits when sitting. When walking, the pairing of Gestures with either Dwell time or Pursuits proved to be significantly more accurate than *PursuitsPursuits*.

5.4. Task-load Index

We calculated mean TLX scores across all six dimensions for the twelve conditions using the raw NASA-TLX method (Hart, 2006). The scores are based on a scale of 100 (Jansen, Schjerlund, & Hornbæk, 2019; Mathis, Vaniea, & Khamis, 2021; Namnakani et al., 2023; Speicher, Feit, Ziegler, & Krüger, 2018). The lower the scores, the lower the workload (See Figure 6). There was a statistically significant interaction between Input techniques and motor activity on the task load, $F_{5,253}= 2.87$, $p < .05$. Post hoc analysis showed that *DwellGestures* resulted in significantly lower workload than *PursuitsDwell* while sitting ($M = 16.5, SD = 13.8$ compared to $M = 28.5, SD = 17.5$, $p < .05$) and walking ($M = 26.7, SD = 18.2$ compared to $M = 39.8, SD = 19.5$, $p < .05$). *PursuitsPursuits* showed a significantly higher cognitive workload ($M = 44.9, SD = 20.9$) compared to both *DwellDwell* ($M = 28.0, SD = 20.3$) and *DwellGestures* when walking ($p < .001$).

5.5. Likert Scales & Preference Rankings

Participants used a 5-point Likert scale (1= Strongly Disagree; 5= Strongly Agree), to answer questions that evaluated various aspects of navigation and selection using different eye movements. We structured questions as statements such as "I found navigating with { Input Technique } and selecting with { Input Technique } to be easy, accurate, easy to learn, enjoyable, not requiring much attention, tiring, and something

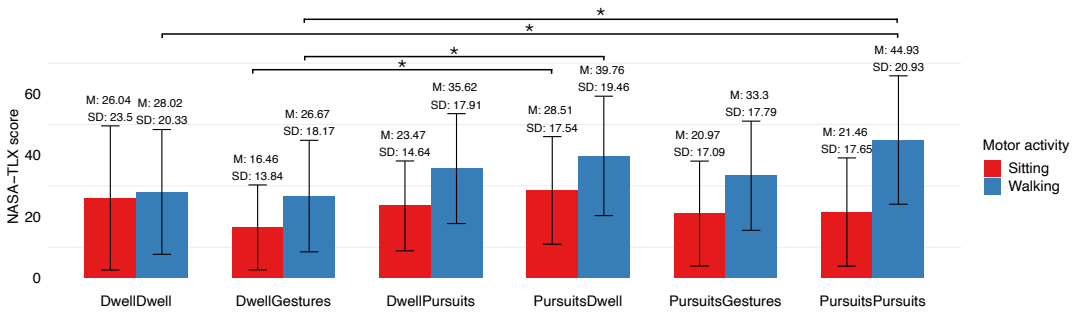

Figure 6. Mean overall NASA-TLX workload per input techniques in each motor activity. The error bars represent the standard deviation.

I would use in the presence of others. Participants perceived *DwellGestures* ($Med = 4, IQR = 1$) as significantly more accurate than *PursuitsDwell* ($Med = 3, IQR = 1.5$) while seated. When walking, they found *DwellGesture* to be significantly more accurate ($Med = 4, IQR = 1$) and enjoyable ($Med = 4, IQR = 0$) compared to *DwellPursuits* ($Med = 2.5, IQR = 2$, $Med = 3, IQR = 2$), *PursuitsDwell* ($Med = 3, IQR = 1$, $Med = 3, IQR = 2$), and *PursuitsPursuits* ($Med = 2.5, IQR = 1$, $Med = 3, IQR = 2$). *DwellDwell* ($Med = 4, IQR = 2.25$), *DwellGestures* ($Med = 4, IQR = 0$), and *PursuitsGestures* ($Med = 4, IQR = 1$) were perceived as significantly easier to use than *PursuitsPursuits*. Table 2 presents the medians, Interquartile Ranges (IQR) and significance.

We also asked the participants to express their preference for the proposed input techniques based on their experience (1= most preferred; 6= least preferred). In seated motor activities, nine participants identified *DwellGestures* as their most favoured technique, whereas ten participants ranked *DwellGestures* as their top preference in walking scenarios (see Figure 7). Friedman test revealed a significant effect of input techniques on rankings in both activities: sitting, $\chi^2(5) = 31.5$, $p < .0001$ and walking, $\chi^2(5) = 21.52$, $p < .001$, motor activities. Pairwise comparison using Wilcoxon signed-rank tests, adjusted with a Bonferroni correction, showed that *DwellDwell* ($Med = 2, IQR = 3, p < .01$), *DwellGestures* ($Med = 2, IQR = 2, p < .001$), and *DwellPursuits* ($Med = 3, IQR = 1, p < .05$) were ranked significantly better compared to *PursuitsDwell* ($Med = 5, IQR = 1.50$) and *PursuitsPursuits* ($Med = 5, IQR = 1.50$) in sitting condition. While walking, *PursuitsPursuits* ($Med = 5, IQR = 2$) was significantly less preferred compared to *DwellDwell* ($Med = 3.50, IQR = 2.25, p < .05$), *DwellGestures* ($Med = 2, IQR = 2.50, p < .005$), and *DwellPursuits* ($Med = 3, IQR = 0.25, p < .01$). *DwellGestures* was also significantly ranked better compared to *PursuitsDwell* ($Med = 4, IQR = 2, p < .05$).

5.6. Qualitative Feedback

We collected feedback through semi-structured interviews that were recorded and transcribed. We employed thematic analysis with an inductive approach to allow us to determine frequently discussed topics by participants. Through this process, the codes were aggregated into subcategories, resulting in three primary themes.

Table 2. While participants perceived *DwellGestures* as significantly more accurate compared to *PursuitsDwell* during seated motor activities, Significant differences were found between multiple pairs on various usability aspects while walking. The numbers represent medians and Interquartile Ranges (IQR), in brackets, along with statistical significance between pairs.

	DwellDwell (1)	*DwellGestures (2)*	*DwellPursuits (3)*	*PursuitsDwell (4)*	*PursuitsGestures (5)*	*PursuitsPursuits (6)*	*Friedman test*	*p <.05*
Sitting								
Ease of use	4.00 (1.50)	4.00 (1.00)	4.00 (.25)	4.00 (1.00)	4.00 (.50)	4.00 (1.25)	$\chi^2(5) = 10.22, p > .05$	
Accuracy	3.00 (2.00)	4.00 (1.00)	4.00 (1.00)	3.00 (1.25)	4.00 (1.00)	4.00 (1.25)	$\chi^2(5) = 16.85, p < .01$	(2,4)
Learnability	5.00 (1.00)	5.00 (1.00)	4.00 (1.00)	4.00 (1.00)	4.00 (1.00)	5.00 (1.00)	$\chi^2(5) = 2.18, p > .05$	
Enjoyability	4.00 (1.00)	4.50 (1.00)	4.00 (2.00)	4.00 (1.25)	4.00 (2.00)	5.00 (2.00)	$\chi^2(5) = 14.21, p < .05$	
Effortlessness	2.00 (2.00)	3.00 (2.00)	2.50 (2.00)	2.00 (1.00)	2.00 (1.25)	3.00 (1.00)	$\chi^2(5) = 3.85, p > .05$	
Tiredness	2.00 (2.25)	3.00 (2.25)	3.00 (2.00)	3.00 (1.25)	3.00 (2.00)	2.00 (2.00)	$\chi^2(5) = 3.37, p > .05$	
Social Acceptability	4.00 (2.00)	4.00 (1.00)	4.00 (1.00)	4.00 (2.00)	4.00 (2.00)	4.00 (1.25)	$\chi^2(5) = 2.22, p > .05$	
Walking								
Ease of use	4.00 (2.25)	4.00 (0.00)	3.00 (2.00)	3.00 (2.00)	4.00 (1.00)	2.00 (1.25)	$\chi^2(5) = 29.15, p < .0005$	(1,6), (2,6), 5,6)
Accuracy	4.00 (2.00)	4.00 (1.00)	2.5 (2.0)	3.00 (1.00)	3.00 (1.00)	2.5 (1.00)	$\chi^2(5) = 23.40, p < .001$	(2,3), (2,4), (2,6)
Learnability	5.00 (1.00)	5.00 (1.00)	4.00 (.25)	4.00 (1.00)	4.00 (1.00)	4.00 (1.25)	$\chi^2(5) = 22.63, p < .001$	(1,6)
Enjoyability	4.0 (1.00)	4.00 (0.00)	3.00 (2.00)	3.00 (2.00)	4.00 (1.00)	3.00 (2.00)	$\chi^2(5) = 25.62, p < .001$	(2,3), (2,4), (2,6)
Effortlessness	2.50 (1.00)	2.00 (1.25)	2.00 (.25)	2.00 (1.00)	2.00 (2.00)	2.00 (1.00)	$\chi^2(5) = 13.67, p < .05$	(1,6)
Tiredness	3.00 (2.00)	2.50 (1.25)	4.00 (1.00)	3.50 (1.25)	3.00 (2.00)	4.00 (1.00)	$\chi^2(5) = 16.26, p < .01$	
Social Acceptability	4.00 (2.25)	4.50 (2.00)	3.50 (2.25)	4.00 (1.50)	4.00 (2.00)	3.50 (2.25)	$\chi^2(5) = 13.35, p < .05$	

5.6.1. Theme: Participants' Preference Evaluation:

DwellGestures was associated with seven positive attributes (n=8), such as intuitive (2), easy to understand (1), natural (1) and fast when Dwell time is paired with Gestures, as reported by P16. *DwellDwell* was associated with five positive attributes (n=8), such as easy to perform (4) and fast (1). Some participants preferred *DwellGestures* and *PursuitsGestures*, attributing this to the gesture part for navigation, as P4 stated that any pair of techniques that include Gestures was ranked at the top. *PursuitsGestures* was associated with five positive attributes (n=5), such as being accurate (1), intuitive (1), and less tiring to the eyes (1). P6 reported experiencing eye fatigue when relying on one eye movement (Pursuits). However, they noted that the pairing of Pursuits and Gestures appeared to mitigate this issue, potentially reducing eye strain. Using Gestures for navigation was associated with 18 positive attributes, such as being easy for navigation (3), intuitive (3), and accurate (2). It was also explicitly reported as the favoured technique when paired with either Dwell time or Pursuits (5), although it took longer to learn (4). P18 appreciated utilising Gestures for navigation, deeming it a notably natural action to swipe with their eyes. Dwell time, when paired with either Pursuits or Gestures, was associated with five positive attributes (n=5), such as being easier (2) and less effort (1) compared to Pursuits, and natural (1). However, Dwell time was less preferred where there were many targets (1). P4 expressed their preference for Pursuits when paired with other techniques, reporting that any pair of techniques that included Pursuits was ranked higher compared to Dwell time. P18 reported that Pursuits, when used, show a clear interaction. However, Pursuits was also perceived as less accurate (3), tiring to the eyes (2), uncomfortable to look at (1), and not suitable for navigation (1).

5.6.2. Theme: Participants' Experiences While Mobile:

Two participants expressed the relevance and comfort of using gaze input while walking. P12 noted that although using Dwell time for selection was easy when walking, integrating it with Gestures for navigation further enhanced the interaction by allowing them to take a break from shifting their gaze between various targets on the screen, as gaze gestures could be performed without the need to focus on a specific target. While P21 felt that gaze gestures were hard to perform while walking, other participants perceived it more enjoyable (1) and easy (1). Although *PursuitsGestures* was the favoured technique for P13 during walking, due to the perception of increased accuracy with Pursuits as compared to Dwell time, Pursuits was also viewed as confusing (1),

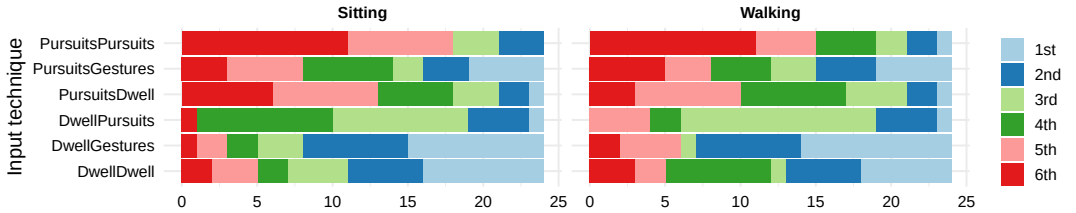

Figure 7. Participants ranked their preference for the input techniques (1= Most preferred; 6= Least preferred). In both seated and walking motor activities, *DwellDwell*, *DwellPursuits*, and *DwellGestures* were consistently preferred, with a significant preference over *PursuitsDwell* and *PursuitsPursuits* while seated. While walking, *PursuitsPursuits* was significantly less favoured than the three.

hard (1), and requiring concentration (1). In contrast, Dwell time was perceived as better than Pursuits for use while walking (1). While P19 ranked *DwellDwell* as the most preferred technique for sitting, they ranked *DwellGestures* highest when walking, finding it easier to perform the task with this pair, as the head moves naturally while walking.

5.6.3. Theme: Contexts and Applications:

Five participants mentioned using gaze when the temperature is low and wearing impeding clothing, such as gloves; as P10 noted, "*I still have to use my nose sometimes, or take my hand out of the gloves, to interact with the phone, which is not perfect*". They also suggested gaze use when both hands are busy (7), dirty or wet (2), carrying bags (1), or when touch is impossible, such as while driving (1) or cycling (1). Four participants expressed a utilitarian view on gaze input, one stating, "*I would have no problems using that anywhere; if it is useful, I would use it*". Participants reported home use such as reading in bed and swiping pages with gaze (1), listening to music on the couch or bed (1), interacting with hands under a blanket (1), and in stationary situations on public transport (1) or when the device was out of reach (1). Four participants also highlighted accessibility benefits. They also suggested applications to use with gaze, including music players (7), taking photos (2), predefined text selection for typing (1), page navigation (2), album browsing (2), initiating calls (2), and emergency calls (1). While participants were enthusiastic about applications, seven raised concerns about using gaze while walking outdoors.

6. Discussion & Future work

6.1. Gestures can Enhance Navigation when Paired with Other Techniques

Our experiment revealed that pairing Pursuits with Gestures (*PursuitsGestures*) led to significantly faster task completion times, reduced error rate, and improved completion rates compared to using Pursuits (*PursuitsPursuits*) solely during walking activities. The collected feedback showed that participants generally appreciated pairing input techniques rather than using a single eye movement, as one participant explicitly reported that combining Gestures with Pursuits reduced eye fatigue caused by using Pursuits only, and another ranked any input technique paired with Pur-

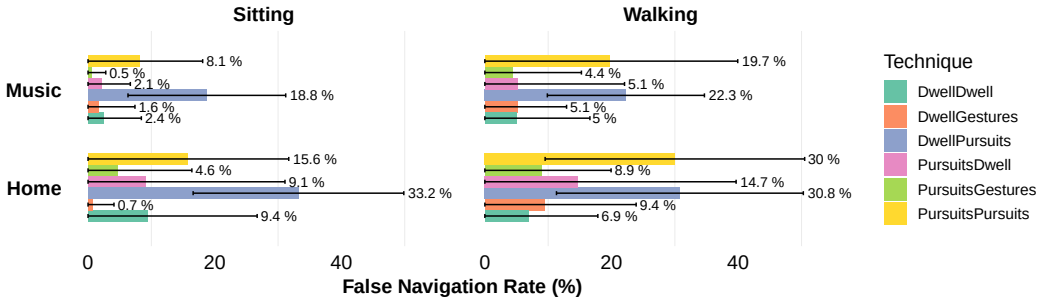

Figure 8. Mean false navigation rate for each navigation technique when paired with other techniques. It is calculated as the proportion of incorrect navigation relative to the total number of navigations performed. Gestures, when paired with Dwell time or Pursuits, resulted in fewer errors compared to Dwell time or Pursuits when used for navigation in sitting motor activity. Pursuits, in *DwellPursuits*, caused the highest navigation error rate. The error bars represent the standard deviation.

suits as most preferred. Gestures, in particular, when paired with other techniques, was associated with 18 positive attributes. Likert scale responses also showed that participants perceived Gestures as significantly easier to use when paired with Dwell time or Pursuits whilst walking than when using Pursuits only. Unlike other pairs, we also found no evidence that the motor activity, whether sitting or walking, has an impact on the completion time when using *DwellGestures*. This pair led participants to rank *DwellGesturees* as the most preferred in both motor activities. While pairing Gestures with Dwell time and Pursuits positively impacted both techniques, the preference for *DwellGestures* over *PursuitsGestures* could be due to the minimal deliberate eye movements Dwell time requires (Penkar, Lutteroth, & Weber, 2012), its convenience (Jacob, 1991), and being the gaze counterpart to touch tap (Fernandez et al., 2020). Our findings align with prior work that suggested the effectiveness of single stroke gaze gestures (Møllenbach et al., 2013), the recommendation of gestures with few targets (Namnakani et al., 2023), and its proven robustness while walking (Lei, Wang, et al., 2023). While complex gaze gestures can cause physiological and cognitive difficulties (Møllenbach et al., 2013; Porta & Turina, 2008) and was least favoured for selection with many targets (Namnakani et al., 2023), our results show they can be preferred when used to complement rather than replace other gaze input techniques. In our study, Gestures was employed for navigation, a process that P18 perceived as a natural eye-swiping action. This extends prior work that suggested combining multiple techniques, such as Dwell time and Gestures, for issuing different commands (Møllenbach et al., 2013) and the suggested guidelines for enabling users to utilise different techniques based on context and the number of targets (Namnakani et al., 2023).

Based on the results, we emphasise the value of integrating various gaze input techniques in a single user interface. This opens avenues for future research to explore the impact of task complexity and investigate the effectiveness of integrating gaze input techniques for more complex tasks, such as zooming and scrolling. Future work could explore which specific functionalities or commonly used basic gestures (Apple Inc., 2024b), such as zooming, scrolling, or double-tapping, are most effectively executed using particular gaze input techniques. These techniques can be integrated into a single interface, enabling users to familiarise themselves with the functionality of each technique and facilitating the design of an effective mobile gaze-enabled interface that supports fundamental gestures. Gaze interfaces may also adapt to individual

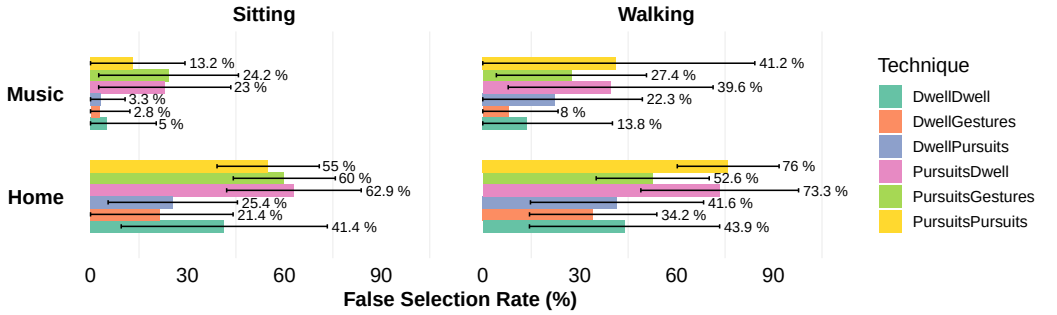

Figure 9. Mean false selection rate for each technique used for selection when paired with other techniques. It is calculated as the proportion of incorrect selections relative to the total number of selections performed. Dwell time, when used for selection and paired with Gestures for navigation (*DwellGestures*, resulted in fewer errors compared to when the technique is used for selection and navigation (*DwellDwell*). The error bars represent the standard deviation.

users, with personalised gaze input techniques paired based on users' performance and preferences.

6.2. Underlying Mechanisms Influencing Performance

Given that Pursuits and Gestures both rely on relative eye movements (Khamis, Alt, & Bulling, 2018a; Vidal, Pfeuffer, et al., 2013), and when paired (*PursuitsGestures*), led to a noticeable decline in gesture performance, particularly in terms of task completion time, on both the home and music player interfaces. Regardless of the motor activities, we noticed that Gestures were harder to initiate and took longer to perform (Home screen: $M = 4.0s, SD = 4.07$; Music Player: $M = 13.1s, SD = 6.69$), compared to when Gestures was paired with Dwell time (Home screen: $M = 3.3s, SD = 2.90$; Music Player: $M = 12.3s, SD = 4.65$) for selection. This is likely due to the visual clutter that Pursuits adds to the interface caused by the moving trajectories, which can be visualised in the rightmost screenshots in Figure 10, where red colour crosses in Home Screen 1 show how Pursuits targets were activated while participants were trying to perform the gesture to navigate from Home Screen 1 to 2. However, despite these challenges, Gestures have been shown to be more robust than Pursuits during walking (Lei, Wang, et al., 2023). Our experiment supports this by demonstrating that pairing Pursuits with Gestures significantly reduced error rates during walking tasks (Figure 5). This suggests that although the pairing of Pursuits and Gestures may increase task time, it enhances accuracy and reliability during dynamic activities, such as walking.

When Pursuits is used for navigation and paired with Dwell time for selection (*DwellPursuits*), we noticed an improvement in the navigation speed ($M = 4.59s, SD = 3.66$), but not the error rate in the home screen compared to using Pursuits for navigation in *PursuitsPursuits* ($M = 5.34s, SD = 4.51$). This could be attributed to the inherent speed of the technique (Namnakani et al., 2023) while mitigating the negative effects associated with a large number of targets to select from. To reduce errors and avoid false correlation detection with Pursuits during navigation, which showed the highest error rate compared to Dwell time and Gestures on both home and music player screens in both motor activities (Figures 8 and 9), correlation could be computed only when the gaze is within proximity to the Pursuit targets.

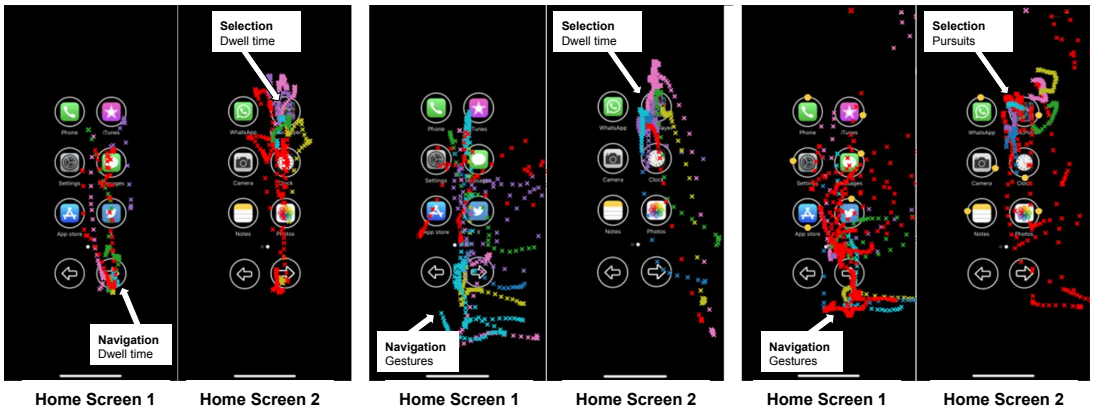

Figure 10. Gaze points from six participants are plotted on the phone's home screen. The red colour represents the gaze points that have led to incorrect actions. Other colours represent the gaze points that ended with each participant's correct actions. **Left:** participants navigated the home screen using Dwell time and then selected the music player on Home Screen 2 using the same technique. **Middle:** participants navigated with Gestures and selected the music player using Dwell time. **Right:** participants navigated with Gestures and selected the music player using Pursuits.

Figure 10 shows that participants' gaze often stayed within the same region between screens. For instance, when performing right or left gestures on the home screen (Middle and Right screenshots in Figure 10), the eyes moved to the screen edge and remained there on the next screen. This behaviour suggests UI elements, such as navigation buttons, could remain in consistent areas to support continuity and faster navigation. Placing content where users' gaze naturally rests can also help maintain focus and reduce search effort.

6.2.1. Gestures for Navigation Reduces Dwell Unintentional Activation

Pairing Dwell time with Gestures for navigation reduced the occurrence of unintentional selections, or the Midas touch effect. This is likely due to the gaze behaviour between screens. This improvement is evident in Figure 9 and the middle screenshot of Figure 10, where fewer red crosses appear on Home Screen 2, indicating fewer unintentional activations compared to when Dwell time was used alone. In contrast, the left-hand screenshots of Figure 10 show a higher number of accidental dwell target activations when participants shifted their gaze from the navigation button to the music player. The mean false selection rate in the *DwellDwell* condition was 41.39% ($SD = 31.9$) on the Home screen, and 5.03% ($SD = 15.3$) in the music player, while this error reduced to 21.37% ($SD = 22.7$) on the home screen and to 2.78% ($M = 9.4$) in the music player when using Dwell time for selection in the *DwellGesture* (see Figure 9).

6.3. Number of Targets, Size, and Placement Affect Gaze-Based Technique Performance

While pairing different techniques proved promising, some pairings were not as promising. While this could be due to the nature of the task, we suspect three other factors may be involved, which we discuss in the following sections.

6.3.1. The Effect of Number of Targets

The effect of the number of targets may impact the accuracy and completion time of paired techniques in various degrees, as suggested by prior work (Namnakani et al., 2023). For example, when walking, the experiment data showed that *PursuitsGestures* is less accurate when used in the home screen, with an average error rate 50.26% (See Figure 2), than when used in the music player application ($M = 15.35\%$). *PursuitsDwell* also resulted in more errors compared to *DwellPursuits*, as the former displayed more targets to be selected using Pursuits, while the latter pair used Pursuits for navigation with only two targets (see Figures 8 and 9). One approach to mitigate the impact of the number of targets is to display a limited number of functions or targets on each screen. This would allow users to concentrate on a more concise set of tasks, potentially enhancing their overall performance.

6.3.2. The Effect of Target Size

In our implementation, we ensured that the target size remained consistent across all the input techniques employed, following Apple design guidelines and inspired by prior work on gaze interaction on mobile devices (Apple Inc., 2024a; Namnakani et al., 2023). While this approach guarantees a balanced comparison among the techniques, it impacts the performance of each technique to varying extents. This issue could be mitigated by designing the interface with specific consideration for the input techniques, where targets with Dwell time, for example, can be sized to better compensate for tracking inaccuracy (Feit et al., 2017; Namnakani, Abdrabou, Grizou, & Khamis, 2025).

6.3.3. The Effect of Target Placement

The placement of the targets on the screen may also impact the performance of the input techniques, as this was suggested by prior work (Feit et al., 2017; Namnakani et al., 2023, 2025). In our implementation, we endeavoured to position all targets near the centre of the screen to minimise such an effect. However, some targets were placed closer to the edges due to limitations on screen real estate. Although randomising the target positions might mitigate such effects in the study, it could introduce a time cost due to prolonged search times, potentially affecting the analysis. In designing a gaze-enabled user interface, careful attention must be given to the placement of targets, ensuring they align with the variations in accuracy and precision across different screen regions (Feit et al., 2017; Namnakani et al., 2023).

6.4. Performance in Real-world Mobile Usage

By conducting our experiment in a corridor instead of a lab-controlled setting, we introduced uncontrollable variables such as the passersby. While this approach may contribute towards understanding the real-world applicability of our findings, outdoor or in-the-wild studies provide more realistic conditions due to the inherent environmental and contextual factors such settings introduce (Chittaro, 2010; Kane et al., 2008). Issues include changes in the weather, the various lighting conditions, the shaky environments due to the dynamic use, disturbance by pedestrians or obstacles, the limited or split attention, and the added distractibility by public events (Chittaro, 2010; Kane et al., 2008; Khamis, Alt, & Bulling, 2018b). As mentioned in Section 4.6, we intentionally chose an indoor environment, with minimal distraction, and controlled

lighting conditions to explore the performance of pairing gaze-based techniques in isolation. Our findings provide an initial step toward future investigations of pairing gaze techniques in the wild.

To address issues in a shaky environment, which can often break calibration data, eye-tracking systems could leverage IMU sensor data to detect when recalibration is needed, depending on the user's context, as done in MAC-GAZE (Lei et al., 2025). As explicit calibration is tedious and time-consuming (Khamis, Alt, & Bulling, 2018b), implicitly calibrating users is promising (Cai, Hong, Wang, & Lu, 2025; Yang, Jin, & He, 2023), especially for the mobile context, where users are likely to calibrate more often due to the dynamic nature of mobile use (Khamis, Alt, & Bulling, 2018b).

While Eyedid SDK, the eye-tracking library used in our experiment, utilises RGB images for gaze estimation, RGB cameras are sensitive to lighting conditions (Khamis, Alt, & Bulling, 2018b; Majaranta & Bulling, 2014). For example, tracking eyes using RGB cameras in the dark is challenging. Similar to Eyedid, ASGaze proposed using RGB cameras to enable eye tracking based on geometric eye models (Cao, Lin, Liu, & Li, 2023). While avoiding the high cost of commercial solutions, ASGaze enables tracking on various external surfaces, such as non-electronic whiteboards and computer screens, and is not limited to phone screens. When tested under various settings, ASGaze achieved good tracking performance, but its performance degrades when illumination is very low, as it becomes harder to extract the iris boundaries reliably. The tracking error with ASGaze also increases with the user's walking speed. This could be due to increased motion blur and larger variations between consecutive frames caused by hand-induced phone shaking, affecting iris boundary extraction, a key component of eye tracking in ASGaze. On the other hand, the use of the Infrared-based Pupil-Corneal Reflection (IR-PCR) mitigates tracking issues in the dark but comes at the cost of not being reliable under direct sunlight, as sunlight may interfere with the IR sensors (Khamis, Alt, & Bulling, 2018b; Majaranta & Bulling, 2014).

Moving from frame-based cameras to event-based cameras for tracking may be a plausible approach to take, given that conventional frame-based cameras are limited by the fixed rate sampling, often bounded by a few hundred hertz, depending on the tracking hardware (G. Zhao et al., 2023). Recent system EV-Eye (G. Zhao et al., 2023) utilised a bio-inspired event camera to capture asynchronous, pixel-level intensity changes with sub-microsecond latency, enabling adaptive and high-frequency eye tracking. In EV-Eye, each pixel asynchronously reports brightness changes with precise timestamps, enabling continuous, low-latency representation of motion without using frames. As event cameras produce events only when intensity changes are detected, they save sensing energy and bandwidth (G. Zhao et al., 2023). Given their high dynamic range (140 dB compared to 60 dB for traditional RGB cameras), such cameras operate effectively under challenging lighting conditions, making them suitable for eye tracking on handheld mobile devices, where lighting conditions are likely to vary.

6.5. Gaze Has Potential for Addressing Situational Impairments for Mobile Users

Participants suggested using gaze for interaction in contexts or situations that may negatively affect their experience when interacting with mobile phones. While walking is one of these factors, they suggested using gaze when the device is out of reach, when hands are encumbered with various objects such as shopping bags (Sarsenbayeva et al., 2017), when hands are busy, or when wearing gloves and the weather is cold.

These factors were previously identified by Wobbrock et al. (Wobbrock, 2019) as contextual factors that can negatively affect user interaction with mobile devices, leading to situationally induced impairments. Given that situational impairments can impact users of various abilities (Sarsenbayeva et al., 2017; Wobbrock, 2006, 2019), as users experiencing situational impairments might interact with their touch devices in ways similar to individuals experiencing hand tremors, it is crucial to explore design solutions to accommodate such impairments and mitigate their impact. Although we did not optimise the interface to minimise the effect of walking in our experiment, as it was out of scope, our results offer insight into how each technique pairing performs and is perceived by participants while walking. Future work could investigate different ways the interface can be designed, including what gaze input technique to use for specific situations and functionalities to account for the complexity present during mobile interaction. Solutions can be borrowed from prior work that evaluated the use of touch for interaction when walking (Kane et al., 2008) and the presented preliminary design guidelines to account for situational impairments (Saulynas, Burgee, Bendigeri, & Kuber, 2022).

7. Conclusion

In this work, we conducted a study to investigate the impact of pairing gaze input techniques, namely Dwell time, Pursuits, and Gestures, within the same interface on the usability of navigation and selection tasks while interacting with mobile devices during two motor activities: sitting and walking. Our results demonstrate that pairing techniques can generally enhance the effectiveness of gaze interaction on mobile devices, particularly for navigation and selection tasks. While Gestures generally improved performance when combined with either Dwell time or Pursuits, the improvement was more pronounced with Pursuits than with Dwell time. *PursuitsGestures* resulted in significantly faster completion times, reduced error rates, and improved completion rates when walking compared to using Pursuits alone. Participants reported better performance and ease of use with *PursuitsGestures* compared to Pursuits alone. In both motor activities, *DwellGestures* was the most favoured technique by the participants. Our results show several future directions to explore the impact of combining gaze techniques for more complex tasks. We also discussed the implications of combining gaze-based interactions in the same interface and investigated why certain combinations work better than others. Our work opens new directions for enhancing gaze-based interactions on mobile devices and better leveraging gaze input for interacting with mobile devices.

Acknowledgements and Funding

This work was supported by the Islamic University of Madinah. We acknowledge the use of ChatGPT-5 to enhance the readability and shorten certain parts of the manuscript through paraphrasing. It was not used for the entire manuscript.

Author Contributions

CRediT: **Omar Namnakani:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Yasmeen Abdrabou:** Conceptualization, Methodology, and Writing – review & editing; **Jonathan Grizou:** Supervision; **Mohamed Khamis:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing.

Disclosure Statement

The authors report there are no competing interests to declare.

About the Authors

Omar Namnakani is a PhD student at the University of Glasgow. While his Phd research focuses on designing gaze interaction techniques for target acquisition on handheld mobile devices, he is interested in exploring eye tracking and gaze-based interaction across ubiquitous platforms, such as XR.

Yasmeen Abdrabou is a postdoctoral researcher and team leader at TU Munich. Her work focuses on user behavior, eye tracking, and usable security, investigating how behavioral signals such as gaze and interaction dynamics reveal attention, workload, and vulnerability in interactive systems.

Jonathan Grizou is an Honorary Lecturer at the University of Glasgow specialising in interactive machine learning and self-calibrating interfaces. His interdisciplinary research includes human–AI interaction, brain–computer interfaces, and AI for scientific discovery, with a focus on curiosity-driven, closed-loop systems that adapt to users and environments without prior calibration.

Mohamed Khamis is a Professor at the University of Glasgow and leads the SIRIUS Lab. His research interests include eye-tracking and gaze-based interaction for ubiquitous platforms, including handheld mobile devices, XR, and situated displays. He has 150+ publications and leads international projects impacting industry, government, and UK and EU policy.

References

Apple Inc. (2024a). *App icons.* Retrieved from https://developer.apple.com/design/human-interface-guidelines/foundations/app-icons/ (accessed 10 September 2024)

Apple Inc. (2024b). *Gestures - human interface guidelines.* Retrieved from https://developer.apple.com/design/human-interface-guidelines/gestures (accessed 10 September 2024)

Cai, Z., Hong, J., Wang, Z., & Lu, F. (2025). Gazeswipe: Enhancing mobile touchscreen reachability through seamless gaze and finger-swipe integration. In *Proceedings of the 2025 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3706598.3713739

Cao, J., Lin, C., Liu, Y., & Li, Z. (2023). Gaze tracking on any surface with your phone. In *Proceedings of the 20th acm conference on embedded networked sensor systems* (p. 320–333). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3560905.3568544

Carpenter, R. H. (1988). *Movements of the eyes, 2nd rev.* Pion Limited.

Chen, X., Acharya, A., Oulasvirta, A., & Howes, A. (2021). An adaptive model of gaze-based selection. In *Proceedings of the 2021 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3411764.3445177

Chittaro, L. (2010). Distinctive aspects of mobile interaction and their implications for the design of multimodal interfaces. *Journal on Multimodal User Interfaces*, *3*(3), 157–165. Retrieved from https://doi.org/10.1007/s12193-010-0036-2

Drewes, H., De Luca, A., & Schmidt, A. (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* (p. 364–371). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1378063.1378122

Drewes, H., & Schmidt, A. (2007). Interacting with the computer using gaze gestures. In *Proceedings of the 11th ifip tc 13 international conference on human-computer interaction - volume part ii* (p. 475–488). Berlin, Heidelberg: Springer-Verlag. Retrieved from https://doi.org/10.1007/978-3-540-74800-7_43

Dybdal, M. L., Agustin, J. S., & Hansen, J. P. (2012). Gaze input for mobile devices by dwell and gestures. In *Proceedings of the symposium on eye tracking research and applications* (p. 225–228). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2168556.2168601

Esteves, A., Bouquet, E., Pfeuffer, K., & Alt, F. (2022, jul). One-handed input for mobile devices via motion matching and orbits controls. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *6*(2). Retrieved from https://doi.org/10.1145/3534624

Esteves, A., Shin, Y., & Oakley, I. (2020). Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies*, *139*, 102414. Retrieved from https://www.sciencedirect.com/science/article/pii/S1071581920300185

Esteves, A., Velloso, E., Bulling, A., & Gellersen, H. (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* (p. 457–466). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2807442.2807499

Eyedid. (2023). *Eyedid sdk, eye-tracking software for your mobile devices.* Webpage. Retrieved from https://sdk.eyedid.ai

Feit, A. M., Williams, S., Toledo, A., Paradiso, A., Kulkarni, H., Kane, S., & Morris, M. R. (2017). Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 1118–1130). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025599

Fernandez, M., Mathis, F., & Khamis, M. (2020). Gazewheels: Comparing dwell-time feedback and methods for gaze input. In *Proceedings of the 11th nordic conference on human-computer interaction: Shaping experiences, shaping society.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3419249.3420122

Fono, D., & Vertegaal, R. (2005). Eyewindows: evaluation of eye-controlled zooming windows for focus selection. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 151–160). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1054972.1054994

Goncalves, J., Sarsenbayeva, Z., van Berkel, N., Luo, C., Hosio, S., Risanen, S., ... Kostakos,

V. (2017). Tapping task performance on smartphones in cold temperature. *Interacting with Computers*, *29*(3), 355–367.

Gonzalez-Calleros, J. M., Guerrero-Garcia, J., Vanderdonckt, J., & Munoz-Arteaga, J. (2009). Towards canonical task types for user interface design. In *2009 latin american web congress* (p. 63-70).

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904-908. Retrieved from https://doi.org/10.1177/154193120605000909

Heikkilä, H., & Räihä, K.-J. (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* (p. 147–154). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2168556.2168579

Hohlfeld, O., Pomp, A., Link, J. A. B., & Guse, D. (2015). On the applicability of computer vision based gaze tracking in mobile scenarios. In *Proceedings of the 17th international conference on human-computer interaction with mobile devices and services* (p. 427–434). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2785830.2785869

Huang, Q., Veeraraghavan, A., & Sabharwal, A. (2017). Tabletgaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications*, *28*(5), 445–461.

Huynh, S., Balan, R. K., & Ko, J. (2022, dec). imon: Appearance-based gaze tracking system on mobile devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *5*(4). Retrieved from https://doi.org/10.1145/3494999

Hyrskykari, A., Istance, H., & Vickers, S. (2012). Gaze gestures or dwell-based interaction? In *Proceedings of the symposium on eye tracking research and applications* (p. 229–232). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2168556.2168602

Isomoto, T., Yamanaka, S., & Shizuki, B. (2020). Gaze-based command activation technique robust against unintentional activation using dwell-then-gesture. In *Graphics interface 2020*.

Isomoto, T., Yamanaka, S., & Shizuki, B. (2022, sep). Dwell selection with ml-based intent prediction using only gaze data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *6*(3). Retrieved from https://doi.org/10.1145/3550301

Jacob, R. J. K. (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* (p. 11–18). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/97243.97246

Jacob, R. J. K. (1991, apr). The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Inf. Syst.*, *9*(2), 152–169. Retrieved from https://doi.org/10.1145/123078.128728

Jansen, Y., Schjerlund, J., & Hornbæk, K. (2019). Effects of locomotion and visual overview on spatial memory when interacting with wall displays. In *Proceedings of the 2019 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3290605.3300521

Kane, S. K., Wobbrock, J. O., & Smith, I. E. (2008). Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces. In *Proceedings of the 10th international conference on human computer interaction with mobile devices and services* (p. 109–118). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1409240.1409253

Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P., & Raisamo, R. (2014). Gaze gestures and haptic feedback in mobile devices. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 435–438). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2556288.2557040

Khamis, M., Alt, F., & Bulling, A. (2018a). 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* (pp. 1–17).

Khamis, M., Alt, F., & Bulling, A. (2018b). 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.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3229434.3229452

Khamis, M., Buschek, D., Thieron, T., Alt, F., & Bulling, A. (2018, jan). Eyepact: Eye-based parallax correction on touch-enabled interactive displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *1*(4). Retrieved from https://doi.org/10.1145/3161168

Khamis, M., Klimczak, A., Reiss, M., Alt, F., & Bulling, A. (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.* New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3126594.3126630

Khamis, M., Oechsner, C., Alt, F., & Bulling, A. (2018). Vrpursuits: Interaction in virtual reality using smooth pursuit eye movements. In *Proceedings of the 2018 international conference on advanced visual interfaces.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3206505.3206522

Khamis, M., Saltuk, O., Hang, A., Stolz, K., Bulling, A., & Alt, F. (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* (p. 274–285). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2971648.2971679

Khamis, M., Trotter, L., Mäkelä, V., Zezschwitz, E. v., Le, J., Bulling, A., & Alt, F. (2018, dec). Cueauth: Comparing touch, mid-air gestures, and gaze for cue-based authentication on situated displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(4). Retrieved from https://doi.org/10.1145/3287052

Kong, A., Ahuja, K., Goel, M., & Harrison, C. (2021a). Eyemu interactions: Gaze+ imu gestures on mobile devices. In *Proceedings of the 2021 international conference on multimodal interaction* (pp. 577–585).

Kong, A., Ahuja, K., Goel, M., & Harrison, C. (2021b). Eyemu interactions: Gaze + imu gestures on mobile devices. In *Proceedings of the 2021 international conference on multimodal interaction* (p. 577–585). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3462244.3479938

Lei, Y., He, S., Khamis, M., & Ye, J. (2023, sep). An end-to-end review of gaze estimation and its interactive applications on handheld mobile devices. *ACM Comput. Surv.*, *56*(2). Retrieved from https://doi.org/10.1145/3606947

Lei, Y., Wang, Y., Caslin, T., Wisowaty, A., Zhu, X., Khamis, M., & Ye, J. (2023). Dynamicread: Exploring robust gaze interaction methods for reading on handheld mobile devices under dynamic conditions. *Proceedings of the ACM on Human-Computer Interaction*, *7*(ETRA), 1–17.

Lei, Y., Zhao, M., Wang, Y., He, S., Sugano, Y., Khamis, M., & Ye, J. (2025). *Mac-gaze: Motion-aware continual calibration for mobile gaze tracking.* Retrieved from https://arxiv.org/abs/2505.22769

Li, Z., Zhao, M., Wang, Y., Rashidian, S., Baig, F., Liu, R., ... others (2021). Bayesgaze: A bayesian approach to eye-gaze based target selection. In *Proceedings. graphics interface (conference)* (Vol. 2021, p. 231).

Majaranta, P., & Bulling, A. (2014). Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing* (pp. 39–65). Springer.

Majaranta, P., Räihä, K.-J., Hyrskykari, A., & Špakov, O. (2019). Eye movements and human-computer interaction. *Eye movement research: An introduction to its scientific foundations and applications*, 971–1015.

Mathis, F., Vaniea, K., & Khamis, M. (2021). Replicueauth: Validating the use of a lab-based

virtual reality setup for evaluating authentication systems. In *Proceedings of the 2021 chi conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3411764.3445478

Mollenbach, E., Hansen, J. P., Lillholm, M., & Gale, A. G. (2009). Single stroke gaze gestures. In *Chi '09 extended abstracts on human factors in computing systems* (p. 4555–4560). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1520340.1520699

Møllenbach, E., Lillholm, M., Gail, A., & Hansen, J. P. (2010). Single gaze gestures. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (p. 177–180). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1743666.1743710

Morimoto, C. H., & Amir, A. (2010). Context switching for fast key selection in text entry applications. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 271–274).

Møllenbach, E., Hansen, J. P., & Lillholm, M. (2013, May). Eye movements in gaze interaction. *Journal of Eye Movement Research*, *6*(2). Retrieved from https://bop.unibe.ch/JEMR/article/view/2354

Nagamatsu, T., Yamamoto, M., & Sato, H. (2010). Mobigaze: Development of a gaze interface for handheld mobile devices. In *Chi '10 extended abstracts on human factors in computing systems* (p. 3349–3354). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1753846.1753983

Namnakani, O., Abdrabou, Y., Grizou, J., Esteves, A., & Khamis, M. (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*. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3544548.3580871

Namnakani, O., Abdrabou, Y., Grizou, J., & Khamis, M. (2025). Stretch gaze targets out: Experimenting with target sizes for gaze-enabled interfaces on mobile devices. In *Proceedings of the 2025 chi conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3706598.3713092

Niu, Y.-f., Gao, Y., Zhang, Y.-t., Xue, C.-q., & Yang, L.-x. (2019). Improving eye–computer interaction interface design: Ergonomic investigations of the optimum target size and gaze-triggering dwell time. *Journal of Eye Movement Research*, *12*(3).

Park, J., Park, S., & Cha, H. (2021). Gazel: Runtime gaze tracking for smartphones. In *2021 ieee international conference on pervasive computing and communications (percom)* (pp. 1–10).

Penkar, A. M., Lutteroth, C., & Weber, G. (2012). Designing for the eye: Design parameters for dwell in gaze interaction. In *Proceedings of the 24th australian computer-human interaction conference* (p. 479–488). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2414536.2414609

Pfeuffer, K., Alexander, J., Chong, M. K., & Gellersen, H. (2014). Gaze-touch: Combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th annual acm symposium on user interface software and technology* (p. 509–518). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2642918.2647397

Pfeuffer, K., Alexander, J., Chong, M. K., Zhang, Y., & Gellersen, H. (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* (p. 373–383). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2807442.2807460

Pfeuffer, K., & Gellersen, H. (2016). Gaze and touch interaction on tablets. In *Proceedings of the 29th annual symposium on user interface software and technology* (p. 301–311). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/

10.1145/2984511.2984514
Porta, M., & Turina, M. (2008). Eye-s: a full-screen input modality for pure eye-based communication. In *Proceedings of the 2008 symposium on eye tracking research & applications* (p. 27–34). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1344471.1344477
Rajanna, V., Malla, A. H., Bhagat, R. A., & Hammond, T. (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)* (p. 1-8).
Rantala, J., Majaranta, P., Kangas, J., Isokoski, P., Akkil, D., Špakov, O., & Raisamo, R. (2020). Gaze interaction with vibrotactile feedback: Review and design guidelines. *Human–Computer Interaction*, *35*(1), 1-39. Retrieved from https://doi.org/10.1080/07370024.2017.1306444
Rivu, S., Abdrabou, Y., Mayer, T., Pfeuffer, K., & Alt, F. (2019). Gazebutton: Enhancing buttons with eye gaze interactions. In *Proceedings of the 11th acm symposium on eye tracking research & applications*. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3317956.3318154
Rozado, D., Moreno, T., San Agustin, J., Rodriguez, F., & Varona, P. (2015). Controlling a smartphone using gaze gestures as the input mechanism. *Human–Computer Interaction*, *30*(1), 34–63.
Salvucci, D. D., & Anderson, J. R. (2000). Intelligent gaze-added interfaces. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 273–280). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/332040.332444
Sarsenbayeva, Z., van Berkel, N., Luo, C., Kostakos, V., & Goncalves, J. (2017). Challenges of situational impairments during interaction with mobile devices. In *Proceedings of the 29th australian conference on computer-human interaction* (p. 477–481). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3152771.3156161
Saulynas, S., Burgee, L., Bendigeri, A., & Kuber, R. (2022). Putting situational impairments in context: developing guidance for situational impairments and severely constraining situational impairments by examining parallel domains. *Universal Access in the Information Society*, 1–26.
Schenk, S., Dreiser, M., Rigoll, G., & Dorr, M. (2017). Gazeeverywhere: Enabling gaze-only user interaction on an unmodified desktop pc in everyday scenarios. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 3034–3044). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025455
Schenk, S., Tiefenbacher, P., Rigoll, G., & Dorr, M. (2016). Spock: A smooth pursuit oculomotor control kit. In *Proceedings of the 2016 chi conference extended abstracts on human factors in computing systems* (p. 2681–2687). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2851581.2892291
Speicher, M., Feit, A. M., Ziegler, P., & Krüger, A. (2018). Selection-based text entry in virtual reality. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3173574.3174221
Stellmach, S., & Dachselt, R. (2012). Look & touch: gaze-supported target acquisition. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 2981–2990). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2207676.2208709
Velloso, E., Carter, M., Newn, J., Esteves, A., Clarke, C., & Gellersen, H. (2017, apr). Motion correlation: Selecting objects by matching their movement. *ACM Trans. Comput.-Hum. Interact.*, *24*(3). Retrieved from https://doi.org/10.1145/3064937
Vidal, M., Bulling, A., & Gellersen, H. (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* (p. 439–448). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2493432.2493477

Vidal, M., Pfeuffer, K., Bulling, A., & Gellersen, H. W. (2013). Pursuits: Eye-based interaction with moving targets. In *Chi '13 extended abstracts on human factors in computing systems* (p. 3147–3150). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2468356.2479632

Ware, C., & Mikaelian, H. H. (1986). An evaluation of an eye tracker as a device for computer input2. In *Proceedings of the sigchi/gi conference on human factors in computing systems and graphics interface* (p. 183–188). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/29933.275627

Wobbrock, J. O. (2006). The future of mobile device research in hci. In *Chi 2006 workshop proceedings: what is the next generation of human-computer interaction* (pp. 131–134).

Wobbrock, J. O. (2019). Situationally-induced impairments and disabilities. *Web Accessibility: A Foundation for Research*, 59–92.

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 143–146). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1978942.1978963

Yang, S., Jin, M., & He, Y. (2023). Continuous gaze tracking with implicit saliency-aware calibration on mobile devices. *IEEE Transactions on Mobile Computing*, *22*(10), 5816-5828.

Zhang, X., Ren, X., & Zha, H. (2010). Modeling dwell-based eye pointing target acquisition. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 2083–2092). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1753326.1753645

Zhao, G., Yang, Y., Liu, J., Chen, N., Shen, Y., Wen, H., & Lan, G. (2023). Ev-eye: Rethinking high-frequency eye tracking through the lenses of event cameras. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in neural information processing systems* (Vol. 36, pp. 62169–62182). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2023/file/c41b5d8c1ba15b2aa83e4fa1541f02c8-Paper-Datasets_and_Benchmarks.pdf

Zhao, M., Huang, H., Li, Z., Liu, R., Cui, W., Toshniwal, K., ... Bi, X. (2022). Eyesaycorrect: Eye gaze and voice based hands-free text correction for mobile devices. In *Proceedings of the 27th international conference on intelligent user interfaces* (p. 470–482). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3490099.3511103