
Copyright © 2023 The Authors. Reproduced under a Creative Commons Attribution 4.0 International License.

For the purpose of open access, the author(s) has applied a Creative Commons Attribution license to any Accepted Manuscript version arising.

https://eprints.gla.ac.uk/298323/

Deposited on: 12 May 2023

Enlighten – Research publications by members of the University of Glasgow
https://eprints.gla.ac.uk
GazeCast: Using Mobile Devices to Allow Gaze-based Interaction on Public Displays

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

Penpicha Sinrattanavong
chingpenpicha@gmail.com
University of Glasgow
Glasgow, United Kingdom

Andreas Bulling
florian.alt@uni-bw.de
University of Stuttgart
Stuttgart, Germany

Florian Alt
florian.alt@uni-bw.de
University of the Bundeswehr Munich
Munich, Germany

Yasmeen Abdrabou
yasmeen.essam@uni-bw.de
University of the Bundeswehr Munich
Munich, Germany

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

Figure 1: We propose GazeCast, a novel system that leverages users’ personal hand-held mobile devices to enable gaze-based interaction with surrounding displays using Pursuits (A). In a user study, we evaluate GazeCast by experimenting with three tracking setups where participants’ gaze was tracked: using a web camera (B), GazeCast with a mobile device attached to a stand (C), and GazeCast with a mobile device held by the participants (D). The illustration is adapted from GTmoPass by Khamis et al. [Khamis et al. 2017a].

ABSTRACT
Gaze is promising for natural and spontaneous interaction with public displays, but current gaze-enabled displays require movement-hindering stationary eye trackers or cumbersome head-mounted eye trackers. We propose and evaluate GazeCast – a novel system that leverages users’ handheld mobile devices to allow gaze-based interaction with surrounding displays. In a user study (𝑁 = 20), we compared GazeCast to a standard webcam for gaze-based interaction using Pursuits. We found that while selection using GazeCast requires more time and physical demand, participants value GazeCast’s high accuracy and flexible positioning. We conclude by discussing how mobile computing can facilitate the adoption of gaze interaction with pervasive displays.

CCS CONCEPTS
- Human-centered computing → Interaction design; Interaction techniques.

KEYWORDS
Public Displays, Eye Tracking, Pursuits, Gaze Interaction, Mobile Devices

1 INTRODUCTION
Gaze-based interaction is deemed particularly promising for public displays. Interacting using gaze is fast [Sibert and Jacob 2000] possible from a distance, hence offering users an attractive alternative to touch interaction that is also suitable for unreachable displays (e.g., displays behind glass windows [Davies et al. 2014]). So far,
gaze interaction has been enabled for public displays by either augmenting the display using remote eye trackers and webcams or requiring the user to wear a head-mounted eye tracker [Khamis et al. 2016; Lander et al. 2015a]. While public displays require users to interact from different positions and distances to the display [Müller et al. 2012], stationary eye trackers have a limited range. This means that for users to interact with public displays using gaze, they need to position themselves in the display’s “sweet spot” [Alt et al. 2015]. While head-mounted eye trackers allow for freedom of movement and interaction from a distance [Lander et al. 2015a], they require person-specific calibration and gaze mapping to each display. Although head-mounted trackers have recently become affordable [Kassner et al. 2014] and are envisioned to be integrated into daily Eyewear [Bulling and Kunze 2016], they are still not in wide-spread use and require augmentation of each user [Lander et al. 2015a]. At the same time, off-the-shelf mobile phones allow for accurate gaze estimation, which resulted in many mobile apps in which gaze outperforms touch [Khamis et al. 2022; Liu et al. 2015; Zhang et al. 2017a]. In this paper, we introduce GazeCast—a novel system that leverages users’ handheld devices for gaze-based interaction on public displays. GazeCast uses the smartphone’s front-facing camera to cast users’ gaze onto the display. As such, GazeCast does not require special-purpose eye tracking equipment, does not restrict the user’s positioning, can support calibration-free interaction as demonstrated in our study, and can be easily extended to support multiple users. Furthermore, collecting gaze data through public displays has privacy implications due to the sheer amount of sensitive information inferred from eye movements [Bozkir et al. 2021, 2020; David-John et al. 2022]. GazeCast gives more control to the user on which gaze data is transferred to the display. In our implementation, we transfer the gaze estimates to the display for processing, but the concept allows gaze data to be processed locally on the user’s phone.

We present the results of a user study in which 20 participants provided gaze input on a situated display using Pursuits [Vidal et al. 2013]. To evaluate GazeCast, we experimented with three tracking setups: participants’ gaze was tracked using 1) a webcam (baseline), 2) GazeCast with the smartphone attached to a stand, GazeCast (on a stand), and 3) GazeCast with the smartphone held by the participant, GazeCast (handheld). Results show that while selection time and perceived workload increase with GazeCast, error counts are lower and subjective feedback indicates users value how GazeCast makes their task easier and more accurate and has the potential to make the experience more privacy-preserving. We discuss how mobile computing facilitates the adoption of gaze interaction on public displays.

2 RELATED WORK

We build on prior work on gaze-based interaction on public displays and gaze-enabled handheld mobile devices.

2.1 Gaze-based Interaction on Public Displays

Researchers have investigated the use of gaze input for public display applications, such as voting [Khamis et al. 2016], consuming news [Lander et al. 2015b], interacting with medical images data [Hatscher et al. 2017], gaming [Vidal et al. 2013; Zhang et al. 2013], measuring attention [Alt et al. 2016], and authentication [Khamis et al. 2017a, 2018c]. A recurring challenge in gaze interaction with public displays is that public displays expect users to interact from different positions and distances to the display [Müller et al. 2012]. However, current remote eye trackers limit users’ mobility [San Agustin et al. 2010] and require users to keep their heads in a confined tracking box about 70 cm away from the screen [Khamis et al. 2017b]. Approaches to address this included the use of head-mounted eye trackers [Lander et al. 2015a], which require gaze mapping to each display and are cumbersome to wear. Another approach was to use active eye tracking, where the eye tracker physically moves depending on the user’s position, as done in EyeScout [Khamis et al. 2017b]. In EyeScout, the eye tracker was mounted on a conveyor-belt-like rail but only covered a limited area. A second challenge is that most remote eye trackers are optimized for viewing angles that correspond to screens up to 24 inches, which means that they may not track gaze accurately on larger displays [Rajanna and Hammond 2018]. A third challenge is that public displays expect multiple users to interact simultaneously [Memarovic et al. 2014], but most remote eye-tracking setups support one user at a time. GazeCast works around these challenges as it does not pose positioning and display size requirements, and multiple users can connect to the same display using their handheld mobile devices.

2.2 Gaze-enabled Handheld Mobile Devices

Mobile devices are now equipped with high-resolution front-facing depth cameras and high-performance processors. These advances, alongside the recent developments in computational gaze estimation, have made eye tracking increasingly available on mobile devices [Khamis et al. 2018a]. This resulted in increased eye-tracking applications that run directly on smartphones and tablets. A survey on gaze-enabled handheld mobile devices classifies eye tracking applications on mobile devices to: 1) gaze behaviour analysis, where users’ eye movements are silently tracked for later analysis, 2) implicit gaze-based interaction where the system reacts to the users’ natural eye movements, and 3) explicit gaze-based interaction where the user deliberately moves their eyes to provide input [Khamis et al. 2018a]. We focus on the last as GazeCast is an application for explicit gaze-based interaction. Examples of mobile apps that feature real-time gaze-based interaction include GazeSpeak [Zhang et al. 2017a], which supports communication for people with motor disabilities, gaze-based authentication using GazeTouchPass [Khamis et al. 2022], and many other generic gaze interaction applications from research [Kong et al. 2021; Lee et al. 2017; Li et al. 2017; Miluzzo et al. 2010] and industry [Access 2019; SeeSo 2022].

Most relevant to our work are WorldGaze [Mayer et al. 2020] and GTmopass [Khamis et al. 2017a]. WorldGaze estimates the head-pose using the front-facing camera and detects the gazed-at objects using the rear camera. This was then used to improve voice assistants e.g., looking at a store and asking “What time does this close?”. GazeCast is different in that it detects eye gaze rather than head-pose, and relies on stimuli that are displayed on nearby displays. GTmopass detects multimodal gaze and touch input on mobile devices to authenticate users on public displays. While GTmopass only detects gaze gestures and only communicates
3 GAZECAST CONCEPT AND IMPLEMENTATION

3.1 Concept

The idea behind GazeCast is to use one’s smartphone’s camera as an eye tracker to interact with public displays. There are different ways to connect the smartphone to the public display. In our work, a QR code is shown on the display, which users can scan, directing them to the web application with instructions allowing them to interact with the display. We implemented two web applications (server and client) that can be accessed using different internet browsers e.g., Google Chrome. The concept of GazeCast can be extended to allow more than one user to interact with the same display concurrently by connecting their devices to the public display and establishing the connection.

Building on previous work [Khamis et al. 2016; Vidal et al. 2013], our system uses Pursuits to enable calibration-free gaze-based interaction with public displays. Being a calibration-free interaction technique, Pursuits is particularly suitable for public displays which need to be “immediately usable” [Davies et al. 2014]. By adopting Pursuits, GazeCast allows spontaneous interaction with public displays eliminating the need for calibration. Pursuits checks for motion correlation between the user’s eye movements and trajectories of on-screen moving targets [Esteves et al. 2015; Vidal et al. 2013]. The method’s strength lies in its ability to determine which object the user is gazing at by studying their eye behaviour without the need for accurate gaze estimates.

3.2 Implementation

GazeCast has two main components: the web application user interface (Client) and the Server, implemented using Next.js. We used “socket.io-client” for two-way communication between the client and the server. The client can access the application via a browser, such as Google Chrome, and can connect to the application by sending a message to join the connection. The server receives the gaze data from the client and calculates the correlation every second (30 frames). After detecting the target, it sends the message back to the client to trigger the following view. The message can be sent directly to the client by the server or broadcast to all connected clients. Throughout the application, we continuously log the gaze data and the moving target’s locations in x-, y- coordinates. Gaze data was collected using the SeeSo.io SDK [SeeSo 2022]. We used Pearson’s product-moment correlation coefficient to compute the correlation between the user’s gaze and the targets’ movement as done in prior work [Drewes et al. 2018, 2019; Esteves et al. 2015; Velloso et al. 2017; Vidal et al. 2013]. A target is selected if it has the highest correlation to the user’s eye movements as long as the correlation of the smallest of x and y is greater than the threshold value of 0.8 as suggested by the literature [Esteves et al. 2015]. The choice of the window size of 1 second (30 samples at a rate of 30 frames per second) is based on prior work [Esteves et al. 2020, 2015] and pilot testing with two participants.

While in principle, the calculation of the correlation can be done on the client’s side (i.e., the personal mobile device), we decided to process the gaze data on the server. This was done to ensure a fair comparison with the baseline – this way, the data is processed by the same machine for all experimental conditions. Extending GazeCast to allow the processing of gaze data on the phone would have privacy benefits, as we discuss further in Section 6.

4 GAZECAST EVALUATION

4.1 Study Design

To evaluate the performance of the tracking setups for interaction with public displays, we designed a repeated measures lab study. The experiment has one independent variable: The tracking setups, which has three levels: (1) A webcam mounted on a stand (baseline) (Figure 1B), (2) GazeCast with the mobile device attached to a stand,
4.2 Participants and Apparatus

We recruited 20 participants (10 females), with an average age of 27.7 (SD = 3.96) and an average height of 168.58 cm. Nine participants wore glasses, and three wore contact lenses during the study. Two participants had previous experience with eye tracking, but none had any experience with Pursuits.

For our setup, we used a 32-inch display (1920 × 1080 pixels) in a controlled light environment. We used a standard USB web camera (1080P) for the webcam condition. For GazeCast, we used an iPhone X running on iOS version 15.3.1 with a front-facing camera with a 7 MP sensor. Our web application was opened on a Microsoft Edge browser on an Apple MacBook Air (13”, M1, 2020), which was connected to the display directly. To enable eye tracking on the mentioned devices, we used SeeSo SDK [SeeSo 2022], an eye-tracking library that provides real-time gaze data recorded as an (x,y) position in screen coordinates at 30 frames per second and a reported accuracy of 1.6°. We added a marker on the floor (60 cm away from the screen) indicating where the users should stand for a consistent distance between the participants and tracking setups.

4.3 Task and Procedure

We designed a game, “Follow the pink potato” inspired by EyeVote [Khamis et al. 2016]. In each task, the application displays three potatoes floating in different trajectories. Participants are asked to select the target highlighted in pink. Selecting the wrong target would result in an error. Each displayed target moves at a different speed in one of the three trajectories: circular, zigzag, or diagonal [Khamis et al. 2016] (see Figure 2). Each target moves along its path back and forth until the system detects a selection or the task exceeds the time limit of 30 seconds [Dwivedi et al. 2017]. Once the system detects a selection or a timeout, a message is shown on the screen for seconds telling the user what their selection was, followed by a new round. Figure 3 shows the different screens for the study.

For the procedure, when the participants arrived at the lab, we explained the I am of the study and asked them to fill in the consent and demographics forms. Participants were asked to perform 6 selections per input technique using Pursuits. The order of the conditions was counterbalanced across participants by a Latin square. For each technique, participants started with one training session (1 selection) to familiarise themselves with the task and the eye-tracking device. After each technique, participants rated their experience through 5-point Likert-scale questions, responded to the NASA-TLX questionnaire, and open-ended questions assessing the ease of use, speed, and effort. At the end of the study, participants were required to answer a final questionnaire to rank the eye-tracking devices based on their preferences and performance. The experiment took an average of 60 minutes.

5 RESULTS

Our data showed normal distribution using Shapiro-Wilk test. hence we used one-way repeated measures ANOVA unless stated otherwise. We reported Greenhouse-Geisser-corrected degrees of freedom in cases where Mauchly’s test showed a violation of sphericity. Bonferroni-correction to account for multiple comparisons was used when applicable.

5.1 Selection Time

Figure 4 shows the selection time. Our statistical test showed a significant effect of each tracking setup on selection time $F_{2,38} = 18.340$, $p < .001$. Post-hoc analyses showed significant differences between webcam ($M = 4024.11, SD = 2909.01$) and GazeCast (handheld) ($M = 6627.13, SD = 4284.71$) and between webcam and GazeCast (on a stand) ($M = 7137.47, SD = 4679.40$). As a result, using the webcam yielded the fastest selection time. However, no significant differences were found between both GazeCast conditions.

5.2 Error Counts and Timeout

The results revealed a significant effect of each tracking setup on errors $F_{2,38} = 6.961$, $p = .002$. Post-hoc analyses showed significant differences between webcam (35 errors out of 120 selections) and GazeCast (handheld) (18 errors out of 120) at $p = .039$, and between
We investigated the potential of using a mobile device’s front-facing webcam setup to be the fastest (webcam and GazeCast (on a stand) (15 errors out of 120) at \( p = .006 \). Reflecting on the trajectories, the zigzag trajectory was the most error prone, then the circular and the diagonal trajectories. For the setups, we found that the webcam setup is the most error-pruned, then GazeCast (on a stand) and finally, GazeCast (handheld) (see Figure 5). For timeout, out of all trials by all participants (360 trials), six timeouts were found.

### 5.3 Perceived Workload and Subjective Rating

Figure 6 shows the average NASA-TLX scores out of 100. The statistical tests revealed that the tracking setup has a significant effect on physical demand, \( F_{2,38} = 5.347, p < .05 \). Post-hoc analyses showed significant differences between physical demand (\( p < .05 \)) induced by GazeCast (handheld) (\( M = 29.75, SD = 25.93 \)) and GazeCast (on a stand) (\( M = 16.75, SD = 19.21 \)). Hence, the participants perceived GazeCast (handheld) as more strenuous than the webcam and GazeCast (on a stand), which is expected. No significant effects of the tracking setups on other workload dimensions were found.

Figure 7 shows participants’ subjective ratings. No statistically significant difference was found by Freidman test for the tracking setup on any of the self-reported aspects: ease, speed, accuracy, eye tiredness, and willingness to use daily. However, participants reported that they could perform tasks most easily (\( M = 4.4, SD = 0.99 \)) and accurately (\( M = 3.95, SD = 1.1 \)) with GazeCast (on a stand). This aligns with the results of Error counts. Nonetheless, compared to both GazeCast conditions, participants perceived the webcam setup to be the fastest (\( M = 4.25, SD = 1.02 \)), least tiring (\( M = 2.2, SD = 1.15 \)), easiest to use (\( M = 4.3, SD = 1.13 \)), and most likely to be used in daily life (\( M = 4, SD = 1.12 \)).

### 5.4 Feedback and Ranking

For Usability and user experience, out of the 20 participants, 10 participants reported that they liked to use the mobile as an eye tracker because it was easy to use, “using GazeCast (handheld) is easier than using a webcam” (P17), “using GazeCast (handheld) was the easiest, as I could hold the device at my comfort height” (P19). For participants’ preferences, ten participants reported that they would prefer a mobile phone as an eye tracker over a webcam for privacy and security reasons, as reported by four participants, “Although it is more comfortable to go with a built-in eye-tracker, I prefer to use my phone as an eye-tracker for security reasons” (P14). Finally, for the ranking of the tracking setups, after finishing all tasks, We asked the participants to rank the tracking setups based on their preference, we found that GazeCast (handheld), and webcam were preferred equally (8 out of 20), then GazeCast (on a stand) with four votes.

### 6 DISCUSSION

We investigated the potential of using a mobile device’s front-facing camera as an eye tracker to interact with public displays. In the following, we discuss the implications of using GazeCast for gaze-based interaction with public displays.

#### 6.1 Accuracy and Tracking Setups

Comparing our three tracking setups, the webcam one had the highest error and false detection rate (35 out of 360 selections), compared to 18 for GazeCast (handheld) and 15 for GazeCast (on a stand). Demonstrating that in our implementation, using a smartphone decreases inaccuracy by over 50% and outperforms using a webcam. This might be due to the distance between the user and the webcam (60 cm), which makes gaze estimation difficult, especially when the webcam has a lower resolution than the mobile device. Finally, due to the smaller distance when participants hold the phone with their hand (around 30 cm) the error rate was almost negligible. Another possible reason is that, because users are accustomed to apps that utilize the front-facing camera, they know how to hold their phone in a way that ensures their face is captured by the front-facing camera. On the downside, holding the mobile device for a long time results in a shaky setup, which in turn slows down selection compared to the webcam setup. However, the interaction with public displays is normally short, around 8 seconds [Gentile et al. 2020; Müller et al. 2010] which won’t require huge physical demand. Our findings show that utilizing a mobile device for gaze interaction while holding it works better for a shorter period of time. This aligns with our anticipated use cases for GazeCast, where users interact shortly with public displays to obtain information.

#### 6.2 Ubiquitous and Privacy-preserving Gaze Interaction

The novel idea of GazeCast enables the ubiquitous use of eye-gaze interaction on different interfaces. Even if it is anticipated that eye trackers will soon be incorporated into most devices, this would
require hardware integration which is costly. GazeCast allows gaze interaction with public displays with minimal costs for the display owner, and leveraging personal mobile devices, which are already ubiquitous. GazeCast would have promising applications in shopping centers and in airports where public info displays are common, and on ATMs.

However, with gaze interaction becoming ubiquitous, users’ privacy is at risk. Privacy is particularly important in this context given the rich information content available in human eye movement [Kröger et al. 2020]. several user characteristics, including age, gender, mental illnesses, personality traits, and more, can be inferred from users’ collected gaze data. On the other hand, the majority of users do not recognize the risks of sharing their gaze data, therefore they are willing to share their data with multiple parties [Steil et al. 2019]. In our implementation, the processing was done on the server to ensure a fair comparison with the webcam condition, but the GazeCast concept can enable privacy-preserving gaze-based interaction as processing could be done on the user’s phone. This would prevent gaze data from being shared with a server or a third party. This shows that GazeCast is an effective and low-cost solution to enable gaze-based interaction on public displays while maintaining users’ privacy.

### 6.3 Supporting Multiple Users

The GazeCast concept allows multiple users to interact with the same screen at the same time, which is a common requirement for pervasive public displays [Memarovic et al. 2014]. Although in our experiment one user interacted at a time, below we discuss how our implementation can be extended to support multiple users in terms of 1) connection methods and 2) feedback and screen indicators. For the connection methods, there are various strategies for connecting with the public display. The simplest method we used in our implementation was QR codes. However, since the majority of smartphones have NFC readers, it is also possible to utilize other types of technology, such as NFC tags or BLE beacons [Mäkelä et al. 2017]. To allow multiple users to connect, multiple sockets can be generated randomly and sent to various devices to establish connections. For the feedback and screen indicators, when multiple users interact with the same display, users need a way to distinguish their controls. Previous work visualized the users’ gaze and employed different colours to differentiate users’ gaze cursors [Zhang et al. 2017b]. Another approach could be to split the screen into multiple sections, one per user. Similarly, each section can be mapped to a colour that is mapped to a particular user. On the downside, this may impact the accuracy of the interface as the targets will have to be smaller and potentially closer to each other to fit into one display. Public displays researchers studied ways to support users in identifying their on-screen representation [Khamis et al. 2018b], but not for gaze interfaces which open future work directions.

### 7 LIMITATIONS

As the work aims to investigate GazeCast as a concept, we did not explore the effect of the holding posture, face visibility, and lighting conditions which might affect users' experiences. However, these are ongoing research topics in mobile eye tracking that are forecasted to be addressed in the near future, and thus we leave for future work. Also, we attempted to maintain consistency in the distance between participants and the camera by using the same distance for both the Webcam and GazeCast (on stand). In contrast, for GazeCast (handheld), we told the participants to hold the phone they would do naturally. However, it is important to acknowledge that the different camera resolutions may have introduced a potential confounding factor that could impact the accuracy of our results.

### 8 CONCLUSION AND FUTURE WORK

We presented GazeCast, a novel system that leverages users’ personal handheld devices to enable gaze-based interaction with public displays. We used qualitative and quantitative measures to evaluate GazeCast. While the results show that the selection time and perceived workload increase with GazeCast, users’ feedback suggested that they appreciate how GazeCast allows for flexible positioning and potential privacy benefits when interacting with public displays. Future work can explore how many concurrent users can use the display without affecting the detection accuracy. Another interesting direction is to investigate the different user distances and sweet spots in relation to the detection accuracy.

### ACKNOWLEDGMENTS

This work has been funded by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK EPSRC under grant number EP/S003562/1, and an EPSRC New Investigator award (EP/V008870/1), the Islamic university of Madinah, the Royal Society of Edinburgh (RSE) grant number 1931. A. Bulling was funded by the European Research Council (ERC; grant agreement 801708).

### REFERENCES


