GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze

Yanxia Zhang [*1], **Jörg Müller** [2], **Ming Ki Chong** [1], **Andreas Bulling** [3], **Hans Gellersen** [1]

[1] Lancaster University, Lancaster, UK
[2] Alexander von Humboldt Institute for Internet and Society, Berlin, Germany
[3] Max Planck Institute for Informatics, Saarbrücken, Germany

ABSTRACT

Public displays can be made interactive by adding gaze control. However, gaze interfaces do not offer any physical affordance, and require users to move into a tracking range. We present GazeHorizon, a system that provides interactive assistance to enable passers-by to walk up to a display and to navigate content using their eyes only. The system was developed through field studies culminating in a four-day deployment in a public environment. Our results show that novice users can be facilitated to successfully use gaze control by making them aware of the interface at first glance and guiding them interactively into the tracking range.

Author Keywords

Gaze interface; Public display; Eye tracking; Spontaneous interaction; Walk-up-and-use; Field study

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Gaze input has long been confined to controlled settings but there is increasing interest in gaze for public display contexts. Gaze has the advantage that it naturally indicates user interest, and that users can express gaze input fast [21], over a distance, and without the need of external devices. Gaze has been studied for target selection, content transfer, and navigation tasks on public displays [14, 15, 22]. A variety of works have focused on detection of eye contact to make displays attention-aware [13, 12, 16], on head pose to approximate gaze direction [7, 11], and software-only solutions requiring a single camera [23]. Techniques have been developed that leverage eye movement in novel ways for ad hoc interaction with displays, easing and overcoming the need for calibration to individual users [17, 22, 9]. In spite of these advances, deployment of gaze tracking with public displays has remained limited to passive observation of user attention [20].

A question that arises for the design of gaze interfaces for public display is how passers-by can discover and use gaze

* Corresponding author: yazhang@lancaster.ac.uk

Figure 1. GazeHorizon enables passers-by to interact with public displays by gaze. The system detects users walking up to a display (A), provides interactive guidance to bootstrap use of gaze (B) and lets users navigate displayed information using horizontal eye movement (C). When a user walks away, the system is immediately ready for the next user (D).

control "on the spot". Gaze interface design conventionally assumes that users know how the interface works and how to position themselves to use it, as a result of prior training or because they are guided by an expert (e.g., in a usability lab). However, passers-by attend spontaneously to public displays and usually without prior awareness of what interactions a display supports. Gaze interfaces offer no physical affordance and their availability is not directly visible to users. While we can often quickly glean how an interface works by observing the actions of others using it, this is not possible with gaze interfaces that are controlled by subtle eye movements. The challenge of making passers-by aware of the interactive affordances of a display has been highlighted previously, in studies of public multi-touch and gesture interfaces [3, 6, 19]. Other related work considered social factors in enticing users and groups to approach displays and begin interaction [1, 8, 4]. However, this work is first to focus on how gaze interaction can be bootstrapped in a public display context.

We present *GazeHorizon*, a vision system designed to enable passers-by to interact with a public display by gaze only. As illustrated in Figure 1, the system supports multiple interaction phases and adapts its behaviour when a user approaches the display (leveraging the proxemic interaction concept [2, 18]). It detects when users walk up to a display and provides interactive feedback to guide users into appropriate proximity for use of the gaze interface, and to provide cues to help users operate the interface. The gaze interface itself is purposely simple but robust, using horizontal eye movement of the user as relative input and an intuitive mapping to navigate display content (contrasting SideWays where eye movement
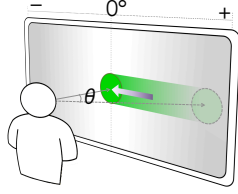
Figure 2. GazeHorizon maps horizontal eye movements for rate-controlled scrolling, with a self-centering effect.

Instruction Levels and Hints	Count
L1 Stand in front of the display.	10
L2 The system reacts to eyes.	11
L3 Keep head still, face towards the display, and move eyes only.	5
L4 Look specifically at each event object.	0
L5 Look at an object and follow it. The object stops in the center.	0
* Failed to use the system after all five levels were revealed	4

Table 1. The five levels of guidance and count of participants who needed the respective level of instruction to use GazeHorizon.

to left or right was mapped to screen regions [22]). When a user turns away from the display, it is immediately ready for the next user. GazeHorizon has been developed through iterative field studies, eliciting insight into the guidance needed for users to be able to use the gaze interface and testing the efficacy of visual cues and interactive feedback. The system was then deployed in-the-wild for four days to evaluate how passers-by would interact with it.

Our contribution from this work comprises insights from the design and deployment of GazeHorizon. By providing appropriate interactive guidance, novice users can be made aware of using only their gaze to control a public display without any expert assistance, and are able to adjust themselves to match GazeHorizon's vision tracking requirements.

DESIGN OF GAZEHORIZON
GazeHorizon is a scrolling system that reacts to users' eye movements. We designed the interface as a browser application with information (e.g., calendar events, latest movies) arranged horizontally. A passer-by first approaches and stands in front a GazeHorizon display. As the user looks at the screen, the item that the user focuses on will scroll towards the center. Figure 2 illustrates the interaction concept.

We use a single web camera for detecting users' eye input (at a resolution of 1280×720px). The system scans for the presence of users looking towards the display, and tracks only the person with a face positioned in the central region. When the tracked face region is larger than 200×200px, the system extracts eye image features using Zhang et al.'s image processing technique [22], and then uses the features to compute the Pupil-Canthi-Ratio (PCR) [23]. PCR represents the horizontal gaze angle θ relative from "looking straight ahead", and we use PCR as a relative input for rate controlled navigation. Looking at a displayed object on the right triggers content to scroll towards the left. As the user's gaze naturally follows the object-of-interest, the scrolling speed decreases and brings the object to a halt in the centre of the display.

Requirements for Interactive Guidance
We conducted a study to understand what information novice users need to use GazeHorizon. Over a two-day period, we invited passers-by in the reception area of a university building to use the system, introducing it as an interactive display that showed local events. We predefined five stages of instructions (listed in Table 1). Participants first stood in front of the display (hence, *L1*); we then prompted them for their perceived interaction model. The researcher gradually revealed the next level and the process continued iteratively until the

participants perceived the interaction correctly. Afterwards, we interviewed the participants for qualitative feedback.

We had 30 participants (8 female) and 26 were able to use the system with *L1* to *L3* guidance (full results shown in Table 1). The participants who only needed *L1* explained that they noticed content movements corresponded to their eye movements. Those who needed *L2* explained that eye-based input was uncommon, and initially assumed the system was "*motion-based*" (e.g. Kinect). After revealing *L2* instruction, the interaction became "*obvious*". *L2* instruction is crucial, as it helps users to eliminate attempts of other modalities for interactions. Those who needed up to *L3* instruction occasionally turned their head, but once told to keep facing forward, they realized the system reacted to eye movements. Four people failed because their eyes got distracted easily by moving objects, or the system could not detect their eyes properly.

Testing Interactive Guidance
We embedded the three levels – *Position* (*L1*), *Eye input* (*L2*), and *Head orientation* (*L3*) – as visual cues in the interface. We used textual labels with pulsing effects, as previous research shows they are effective for enticing interaction on public displays [3, 5, 19]. The system presents the cues in multiple stages for interactive assistance. First, it displays a message to invite users. Once the system detects the presence of a user's eyes, it displays 'look here' labels on the sides, which indicates where the user should look at. Corrective guidance labels appear based on different conditions. If the user is not facing forward, the system displays a "keep head facing forward" message. If the system detects the user's face but not the eyes, it advises the user to ensure that their eyes are not occluded, and if the detected face is too small it suggests that the user step closer.

We conducted a second two-day field study to test the interface. We invited 35 passers-by (5 female) to use the system but provided no assistance. Six failed to use the system due to: strabismus, myopia and eye occlusions. Interviews revealed that most of the participants found the visual cues informative – especially the '*Look Here*' label, but suggested that it was only needed for a short duration – and helped them to realise the interaction very quickly. Initially, the interface gave no precise information of how far the users should stand, and the participants often stood too far (over two meters) away from the screen. We placed a floor marker for positioning users, but it was ineffective as the users often ignored cues on the floor level. We later removed the marker and added an explicit label with distance information on the interface. This was more effective, but the users required time to adjust themselves to the correct position.
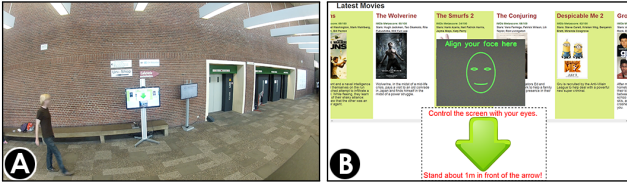
Figure 3. In-the-wild deployment. (A) GazeHorizon was placed in a lobby. (B) Interface showing a mirrored video feed and a label informing users that the system is eye-based and of the distance they should stand.

Figure 4. A histogram of overall users' interaction time over four days of deployment.

Figure 5. (A) & (B) Users adjust their height to align their face. (C) & (D) Incorrect positions. (E) Turned head to look at the side. (F) & (G) Users attempted to interact with the system by waving their hands. (H) A user attempted to make a selection by touching the screen. (I) A user attempted to scroll content by swiping.

Reflection

Most people were unfamiliar with the idea of gaze for display navigation, and required *L2* and *L3* instructions to become aware of the gaze control. Another crucial aspect is to guide users to position within a trackable region (*L1*). The guidance should be intuitive and informative, so the positioning can be done quickly. We added a mirrored video feed [6] overlaid with a face outline for our in-the-wild deployment. The eye scrolling interaction can be learned very quickly, thus we also make the '*Look Here*' label disappear after a period of time when scrolling events are triggered.

GAZEHORIZON IN THE WILD

To understand how people unaidedly use GazeHorizon, we deployed our system as a movie browser to determine its general effects on passers-by in an ecologically valid setting. We placed the system in the lobby of a university building (see Figure 3(A)), where we neither invited, interfered nor advised passers-by. The system logged usage data. Once interaction finished, a researcher approached the users for feedback. We adopted a two-phase deployment approach [10]. *Optimisation*: During day 1 and 2, we iteratively improved the interface according to our observations and users' feedback. After optimisation, the mirrored video feed is placed in the center of the interface (see Figure 3(B)). The video feed and the user positioning label disappears when users stand in the correct position. We added a 'Face position OK' label on the top to provide a constant feedback that the face is still correctly aligned. *Testing*: during day 3 and 4, the prototype was not modified, and we conducted detailed interviews with users.

Findings

Our log data revealed a total of 129 interaction instances, where each instance contains a full episode of uninterrupted use by one or more users [8]. Of the instances, 107 triggered continuous scrolling, with a mean interaction time of 67.1s (*SD*=54.2s). Figure 4 shows a histogram of the interaction time. The users spent an average of 27.0% (*SD*=15.1%) of the interaction time for scrolling. From the moment when users' presence is detected, on average, they required 4.8s (*SD*=8.5s) to align their face correctly and 7.2s (*SD*=11.0s) to perform a first scroll.

During the optimisation phase, we interviewed 46 users; 35 of them (76.0%) reported that they were able to use the system for scrolling information. This rate increased after optimisation. We interviewed 41 users during the testing phase, and 35 of them (85.4%) reported that they were able to scroll the content. Over the four day period, we observed 20 users
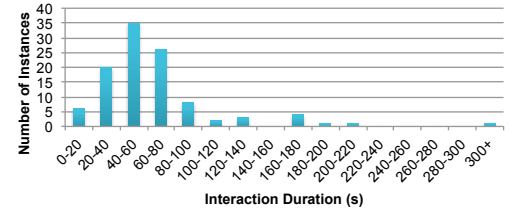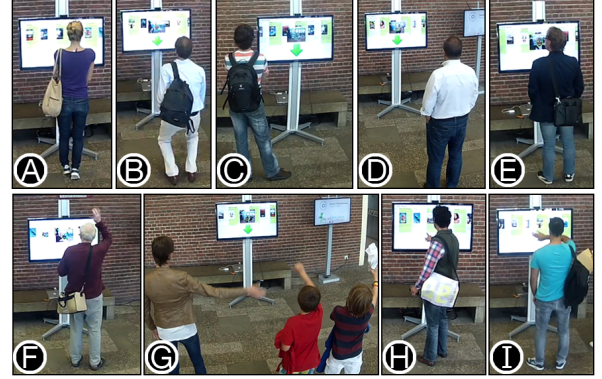
who wore glasses, and 9 of them were able to use our system without removing their glasses.

Interacting with GazeHorizon

When novice users approached our system, they estimated their position by aligning their face, and used the video feed for reference, as it provided realtime feedback. People stepped back and forth to adjust their distance, and then fine-tuned by leaning/tilting their body or adjusting their height (see Figure 5(A) & 5(B)). Sometimes after users noticed the display, they stepped away or stood on one side of the display to observe (see Figure 5(C) & 5(D)). We consider this behaviour as *mirror image avoidance*: although users were curious, they avoided being shown on the "mirror image".

Our results show that people are able to grasp the interaction of GazeHorizon from minimum interactive guidance. The participants were often attracted by the pictures of movie posters, and when they looked at the image the displayed content started to scroll. This made the users realised that the system reacted to what they focused on. Some users commented that the scrolling interaction was "logical" and "very fast to get used to [the system]". Seven users self-discovered difference in speed "*when [they] look more to the sides*".

Although our system was designed for scrolling content, twelve interviewees said they expected the content to be selectable. After scrolling, some users touched the screen in an attempt to trigger a selection (see Figure 5(H)) or used swipe gesture to control the content (see Figure 5(I)).

Figure 6. Honeypot effect. (A) Two passers-by observed a user. (B) The user explained the interaction. (C) The passers-by tried the system.

Some users approached our system without reading the displayed cues, so they did not realise that the system was eye-based. In several instances people attempted to wave their hands (see Figure 5(F) & 5(G)). After seeing no responses, they would then read and follow the interactive guidance. However, some impatient users abandoned further attempts. We also analysed instances where users failed to scroll. Many instances were caused by people standing at the wrong location (e.g. Figure 5(C) & 5(D)). Seven interviewees criticised that they had problems with moving only their eyes instead of turning their head (see Figure 5(E)). Other common causes of failure included: missed displayed labels, impatience due to system delays, impaired vision without glasses, and system failed to detect users' face/eyes.

In general, a large number of users had a positive experience with eye-based scrolling interaction and commented that it was "*interesting*" and "*fun*" (14/41). The majority (34/41) reported that they were comfortable with using GazeHorizon in public area and did not perceive any privacy risks. Some users commented that the technology is useful, e.g., "*good for when hands are dirty and busy*" and "*great for the disabled*". Eight interviewees criticised on system delay, e.g., "*a bit slow with glasses*","*takes too much time to make [the content] move*".

Group behaviour and sharing experience
Passers-by sometimes approached in groups, but usually one person interacted with GazeHorizon at a time. People in a group helped others by pointing to or reading the displayed guidance. We observed the *honeypot effect* [1]. Passers-by became curious after noticing someone using GazeHorizon (see Figure 6). Spectators first positioned themselves to observe from a distance. When the user noticed people were observing, the user explained the interaction and invited the spectators to try. We noticed instances where strangers engaged in short conversations to discuss about the operation.

Lessons Learned From GazeHorizon Deployment
Our deployment confirmed that by providing intuitive guidance novice users were able to control a GazeHorizon display without prior training, and we learned that:

- Gaze is currently an uncommon modality. Letting users know at first glance that the system reacts to eye movement helps them to eliminate attempts of other modalities.
- Impatient users lose interest very quickly and would abandon further attempts if they see no immediate system response from their first action. Similar findings were reported by Marshall *et al*. [4].
- The mirror image is an effective method for positioning users and it provides a real-time reference. Adding a face

outline helps users to figure out the system only support single-user operation, and users in groups took turns to interact with the display individually.

DISCUSSION
For users to learn the interaction of gaze-based system, they require guidance. Guidance could be provided either by an experience user or interface feedback. Requiring an experienced users is impractical for public displays. We adopted the interactive guidance approach, but avoided to add explicit instructions on the interface, and instead only provided guided assistance when the system detects an anomaly. We believe that this is more effective and potentially reduces the cognitive load of users, as they discover the interaction by exploring the interface at their own pace, and the guided assistance helps to prevent misconceptions and to correct user errors.

We learned that the 'look here' label naturally captured users' attention. Although the intention of the users was primarily to look at the label, the action activates scrolling as a after-effect with no extra cost. From a novice user's perspective the scrolling can be seen as an effect of his eye movement, which helps the user to conceptualise the activation of scrolling. We believe that the initial user experience was rather implicit; however, the interaction may become more explicit once the user understands the interaction. A few interviewees explained that once they learned the interaction, they explicitly moved their eyes to the sides for scrolling. Even though our interface did not provide any guidance for stopping the scrolling, somehow all of our participants self-discovered this operation.

Successful operations rely on users making their eyes available in the optimal eye tracking region by correctly positioning themselves and removing occlusion. From our field studies, we realised that there are many unpredictable factors that could hinder the tracking of users' eyes, such as unpredictable user behaviours. They could be corrected by giving appropriate interactive guidance based on specific aspects. If users are aware of a particular reason that causes the system to stop tracking their eyes, the users are generally cooperative and willing to adjust themselves, like removing glasses, stepping closer.

CONCLUSION
In this paper, we presented GazeHorizon, a vision-based system that enables passers-by to navigate information on public displays using their eyes. We conducted a succession of field studies and observed over 190 users (interviewed over 150 of them) to understand what guidance people require to discover the interaction of GazeHorizon. We evolved the system to provide visual cues and interactive assistance to bootstrap gaze interaction with unaware users. Finally, we deployed GazeHorizon "in the wild", where we neither invited nor assisted passers-by. The results showed that a large number of novice users successfully used GazeHorizon. Our work shows that it is possible to integrate spontaneous gaze-based interaction in public settings, and we believe that the work provides a foundation for the investigation of eye-based technology for public displays.

REFERENCES
1. Brignull, H., and Rogers, Y. Enticing people to interact with large public displays in public spaces. In *Proc. INTERACT 2003*, IOS Press (2003), 17–24.

2. Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. Proxemic interactions: The new ubicomp? *interactions 18*, 1 (Jan. 2011), 42–50.

3. Kukka, H., Oja, H., Kostakos, V., Gonçalves, J., and Ojala, T. What makes you click: exploring visual signals to entice interaction on public displays. In *Proc. CHI 2013*, ACM Press (2013), 1699–1708.

4. Marshall, P., Morris, R., Rogers, Y., Kreitmayer, S., and Davies, M. Rethinking'multi-user': an in-the-wild study of how groups approach a walk-up-and-use tabletop interface. In *Proc. CHI 2011*, ACM (2011), 3033–3042.

5. Müller, J., Alt, F., Michelis, D., and Schmidt, A. Requirements and design space for interactive public displays. In *Proc. MM 2010*, ACM Press (2010), 1285–1294.

6. Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F. Looking glass: a field study on noticing interactivity of a shop window. In *Proc. CHI 2012*, ACM Press (2012), 297–306.

7. Nakanishi, Y., Fujii, T., Kiatjima, K., Sato, Y., and Koike, H. Vision-based face tracking system for large displays. In *Proc. UbiComp 2002*, Springer-Verlag (2002), 152–159.

8. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., and Saarikko, P. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *Proc. CHI 2008*, ACM Press (2008), 1285–1294.

9. Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., and Gellersen, H. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proc. UIST 2013*, ACM Press (2013), 261–270.

10. Schmidt, C., Müller, J., and Bailly, G. Screenfinity: extending the perception area of content on very large public displays. In *Proc. CHI 2013*, ACM Press (2013), 1719–1728.

11. Sippl, A., Holzmann, C., Zachhuber, D., and Ferscha, A. Real-time gaze tracking for public displays. In *Proc. AmI 2010*, Springer-Verlag (2010), 167–176.

12. Smith, B. A., Yin, Q., Feiner, S. K., and Nayar, S. K. Gaze locking: Passive eye contact detection for human-object interaction. In *Proc. UIST 2013*, ACM Press (2013), 271–280.

13. Smith, J. D., Vertegaal, R., and Sohn, C. Viewpointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis. In *Proc. UIST 2005*, ACM Press (2005), 53–61.

14. Stellmach, S., and Dachselt, R. Look & touch: gaze-supported target acquisition. In *Proc. CHI 2012*, ACM Press (2012), 2981–2990.

15. Turner, J., Alexander, J., Bulling, A., Schmidt, D., and Gellersen, H. Eye pull, eye push: Moving objects between large screens and personal devices with gaze & touch. In *Proc. INTERACT 2013* (2013), 170–186.

16. Vertegaal, R., Mamuji, A., Sohn, C., and Cheng, D. Media eyepliances: using eye tracking for remote control focus selection of appliances. In *CHI EA 2005*, ACM Press (2005), 1861–1864.

17. Vidal, M., Bulling, A., and Gellersen, H. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proc. UbiComp 2013*, ACM Press (2013), 439–448.

18. Vogel, D., and Balakrishnan, R. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proc. UIST 2004*, ACM (2004), 137–146.

19. Walter, R., Bailly, G., and Müller, J. Strikeapose: revealing mid-air gestures on public displays. In *Proc. CHI 2013*, ACM Press (2013), 841–850.

20. Wooding, D., Mugglestone, M., Purdy, K., and Gale, A. Eye movements of large populations: I. implementation and performance of an autonomous public eye tracker. *Behavior Research Methods, Instruments, & Computers 34*, 4 (2002), 509–517.

21. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (magic) pointing. In *Proc. CHI 1999*, ACM Press (1999), 246–253.

22. Zhang, Y., Bulling, A., and Gellersen, H. Sideways: A gaze interface for spontaneous interaction with situated displays. In *Proc. CHI 2013*, ACM Press (2013), 851–860.

23. Zhang, Y., Bulling, A., and Gellersen, H. Pupil-canthi-ratio: A calibration-free method for tracking horizontal gaze direction. In *Proc. AVI 2014*, ACM Press (2014), 129–132.