Cross-Device Gaze-Supported Point-to-Point Content Transfer

Jayson Turner[*]
Lancaster University

Andreas Bulling[†]
Max Planck Institute for Informatics

Jason Alexander[‡]
Lancaster University

Hans Gellersen[§]
Lancaster University

Abstract

Within a pervasive computing environment, we see content on shared displays that we wish to acquire and use in a specific way i.e., with an application on a personal device, transferring from point-to-point. The eyes as input can indicate intention to interact with a service, providing implicit pointing as a result. In this paper we investigate the use of gaze and manual input for the positioning of gaze-acquired content on personal devices. We evaluate two main techniques, (1) Gaze Positioning, transfer of content using gaze with manual input to confirm actions, (2) Manual Positioning, content is selected with gaze but final positioning is performed by manual input, involving a switch of modalities from gaze to manual input. A first user study compares these techniques applied to direct and indirect manual input configurations, a tablet with touch input and a laptop with mouse input. A second study evaluated our techniques in an application scenario involving distractor targets. Our overall results showed general acceptance and understanding of all conditions, although there were clear individual user preferences dependent on familiarity and preference toward gaze, touch, or mouse input.

CR Categories: H.5.m [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous;

Keywords: eye-based interaction; gaze positioning; multimodal eye-tracking; content transfer; cross-device

1 Introduction

Digital content has become pervasive, with widespread adoption of both personal devices and shared displays. Despite their ubiquity, the transfer of content from large shared displays onto personal devices remains a significant challenge. In traditional single-device interactions, actions are applied to content selected within the bounds of the containing device. However, this immediate select-and-action process is not possible with remotely displayed content. Users regularly wish to select publicly advertised information, and apply actions to it on their personal device. Coping strategies typically require tedious and potentially inaccurate replication (copying down a phone number from a billboard) or inappropriate and potentially intrusive acquisition (photographing the text on a slide during a presentation).

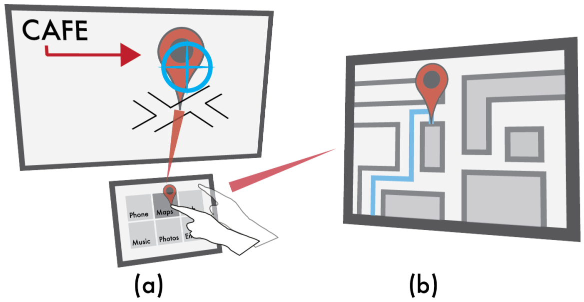


Figure 1: *Content acquisition with gaze: (a) a user obtains the address of a local cafe using gaze. The address is selected by holding down on the tablet. The address is then transferred to touch input where it is positioned and dropped onto a maps application. (b) a map displaying directions to the cafe is presented to the user.*

To illustrate, Figure 1 depicts a user selecting restaurant contact information from a public information display. In this example, the information is initially acquired using their eyes, the primary input modality is then switched as the data is transferred. The user may now position the contact data on one of two application icons: maps, to plot a route to the restaurant, or phone, to place a call and make a reservation. We aim to understand usability when modalities switch during interaction, from gaze to manual input, and how direct and indirect manual input configurations may pose further effects.

During interaction with the environment, our eyes naturally focus on content of interest. We utilise this eye-focus as a modality for pointing and combine it with a discrete trigger to confirm actions, making it possible to select remote targets [Stellmach and Dachselt 2012]. Pointing can now take place when other modalities (i.e., the user's hands) are unavailable. Gaze offers the additional flexibility of interaction with devices over varying distances from close-proximity to those that are out-of-reach, allowing the selection of any content within sight.

Previous work has demonstrated a variety of interaction paradigms where information can be transferred between situated and mobile devices [Hardy and Rukzio 2008; Schmidt et al. 2010; Bragdon et al. 2011]. Many methods are designed to support point-to-point transfer, allowing for the fine-grained positioning of acquired data [Baudisch et al. 2003; Hinckley et al. 2004; Collomb and Hascoët 2008]. However, there is currently no clear understanding as to the effect of switching from gaze to an alternate input modality on users' logical understanding and positioning success.

This work resides in a space where gaze is used to support the transfer of content between remote and close-proximity displays, where the spontaneity and speed of such interaction leverages the implicit nature of gaze-pointing. In this work we examine how gaze-acquired remote content can be transferred to- and positioned on personal devices. We develop two techniques, Gaze Positioning and Manual Positioning, and investigate the impact each complementary modality has on usability. We apply these techniques to direct and indirect manual input configurations, creating four conditions: (1) *Gaze Positioning + Direct*, selection by gaze, confirmed by touch, positioning by gaze, (2) *Gaze Positioning + Indirect*, selection by gaze, confirmed by mouse, positioning by gaze, (3) *Manual Positioning + Direct*, content is transferred from gaze to touch for final positioning (4) *Manual Positioning + Indirect*, gaze-

---

[*] e-mail: j.turner@lancaster.ac.uk
[†] e-mail: andreas.bulling@acm.org
[‡] e-mail: j.alexander@lancaster.ac.uk
[§] e-mail: hwg@comp.lancs.ac.uk

ETRA 2014, March 26 – 28, 2014, Safety Harbor, Florida, USA.

acquired content is transferred to the mouse for final positioning. We conducted two user studies to compare these conditions. The first aimed to evaluate users' performance, accuracy, and their logical understanding of each method when positioning onto a single target. The second study provides a more realistic scenario where the target device contains distractor icons. Our final results show that both gaze and manual positioning approaches show promise in terms of performance. Further insights revealed that transfer between gaze and touch felt natural to users as opposed to mouse.

2 Related Work

2.1 Cross-Device Information Transfer

Early work on cross-device information transfer demonstrated point-to-point interaction between close-by devices. *Pick and Drop* demonstrated a technique that allowed content transfer across devices using a pen device acting as a token to define selection locations, drop locations, and actions [Rekimoto 1997]. *HyperDrag*, allowed for content to be dragged seamlessly across devices [Rekimoto and Saitoh 1999]. Transfer was performed through a public medium (i.e., projected tabletop), a common approach in several works. An alternative approach involved the alignment and "stitching" of devices to create a combined display space for interaction [Hinckley et al. 2004]. Similarly, NFC-enabled mobile devices can be paired with a public display to enable direct device-to-device interaction and transfer [Hardy and Rukzio 2008]. *PhoneTouch* enabled the precise selection and manipulation of content using mobile sensing to issue actions, and computer vision to detect a mobile's location on a rear-projected multitouch surface [Schmidt et al. 2010]. In scenarios where close-proximity to a target is not achievable, drag-and-drop extensions have enabled out-of-reach targeting across multiple wall sized displays [Baudisch et al. 2003; Collomb and Hascoët 2008].

More recent work has focussed on transfer between distant and close-proximity displays. *Perspective Cursor* used head direction to transform multiple non-coplanar displays, enabling fast remote pointing [Nacenta et al. 2006]. Transfer between such displays involves inherent displayless space i.e., the gap between coplanar displays. Further work showed that accounting for this space in motor-controlled input does not improve performance [Nacenta et al. 2008]. In mobile settings, *Touch Projector* mapped touches from a smartphone, through its rear camera, to displays in the environment to enable remote interaction [Boring et al. 2010]. *Code Space* used smartphones as pointing devices for remote screens using touch to confirm actions and initiate transfer [Bragdon et al. 2011]. Unlike touch projector, this work did not rely on alignment through a camera feed but focussed on what could be considered a more natural method of pointing, catering to the spontaneity of such interaction. Motivated by this, our previous work, *Eye Pull, Eye Push*, combined gaze and touch for content transfer between public and personal displays [Turner et al. 2013]. The work compared three techniques and made no consideration for content positioning. Our findings validated the use of gaze and touch for interaction in this space. We adopt the same selection method in this work. In addition we investigate the switching of modalities from gaze to manual input to complete transfer.

2.2 Gaze-supported Interaction

Several works have examined the use of gaze for public display interaction. A common issue in gaze-supported interaction is one of accuracy. Methods using nearest target snapping and manual refinement can improve target selection in desktop environments [Monden et al. 2005]. Gaze has been used to control a fish eye lens interface combined with mobile touch-and-tilt to explore large image collections [Stellmach et al. 2011]. This work reiterates several known principles for gaze-based interaction, that interaction should not rely on accurate gaze positions, and that the "always on" nature of gaze should be countered with manual triggering [Jacob 1990]. This technique was compared against touch to control automotive dashboard widgets across multiple remote and close-proximity displays [Poitschke et al. 2011]. The outcome showed on average a higher cognitive load, induced by gaze interaction under driving conditions. Within a single display, mobile touch-and-tilt mechanisms have been used to improve gaze-based selection [Stellmach and Dachselt 2012] and positioning [Stellmach and Dachselt 2013] accuracy. Our techniques build upon these works, using gaze and manual input principles to enable transfer and positioning between displays. These works highlight further considerations in their design, in particular that actioning manual input during selection should not divert attention away from targeting with gaze, and that manual triggering should use simple actions.

One motivation for our work is users' requirement to interact with many displays throughout their environment. Head-worn eye-tracking has been used to map gaze to any planar digital display in a real-world environment [Mardanbegi and Hansen 2011]. Further work used nodding gestures combined with gaze to issue commands in remote applications [Mardanbegi et al. 2012]. Interaction with eye-gaze on portable devices poses additional challenges. The *MobiGaze* system used two external stereo cameras and infrared illumination to augment handheld mobile devices with gaze input [Nagamatsu et al. 2010]. *EyePhone* used full on-device hardware and processing [Miluzzo et al. 2010]. This system, unlike MobiGaze, did not distinguish between different gaze directions or recognise eye movement, but instead overlaid a grid on front-facing camera images. The location of the eye in this grid would issue commands. It has also been shown that eye-gestures can be used as input for mobile devices, but this approach can be tiring for users' eyes [Drewes et al. 2007]. Our studies use a head-worn eye-tracking system capable of mapping gaze to distant and close-proximity displays.

In traditional desktop interaction, *MAGIC* mouse cursor pointing relied on gaze to warp a cursor to a target area and switched to manual input for fine-grained target refinement [Zhai et al. 1999]. MAGIC was extended to use a touch-sensitive mouse that, when touched it repositioned the cursor to a user's current gaze position [Drewes and Schmidt 2009]. The approach considerably reduced the need for mouse movements in pointing tasks. We have applied a similar modality switching approach to one of our presented techniques. Our aim is to understand how this compares to transfer with gaze alone and how direct and indirect variants may affect performance and usability.

3 Techniques & Conditions

Throughout this work we aim to evaluate several factors that could impact usability when positioning gaze-acquired content. We consider, (1) The effect of using gaze alone versus switching modalities to manual input for final positioning, (2) We want to understand the effects of direct versus indirect input when used in the context of transfer with gaze.

In this section we outline two techniques, *Gaze Positioning* and *Manual Positioning*. These techniques were used in the evaluation of the above factors. We then apply these techniques in two configurations with direct and indirect manual input modalities, a tablet with touch (direct), and a laptop with mouse (indirect). These techniques aim to allow content, acquired by gaze from an out-of-reach context, to be positioned on a close-proximity device. The interac-

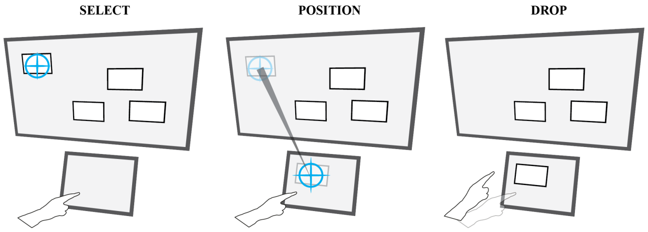


Figure 2: *Gaze Positioning + Direct (Touch): (Select) Look at content on shared display, confirm with touch hold. (Position) Content is attached to gaze and can be positioned on the tablet display. (Drop) Touch release drops the content.*

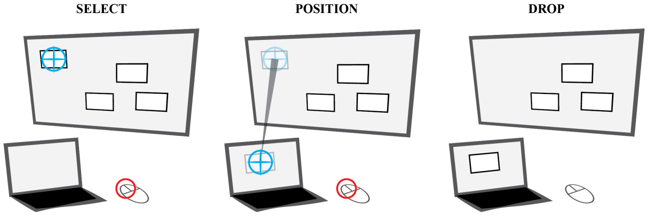


Figure 3: *Gaze Positioning + Indirect (Mouse): (Select) Look at content on shared display, confirm with mouse hold. (Position) Content is attached to gaze and can be positioned on the laptop display. (Drop) Mouse release drops the content.*

tion constitutes three parts, *Selection, Position, and Drop*.

3.1 Techniques

In the case of both techniques presented here, during selection, gaze is used to highlight content, and a manual trigger confirms the selection. As outlined in the related work, this approach has been utilised in several previous works [Jacob 1990; Stellmach and Dachselt 2013; Turner et al. 2013]. Each technique is distinguished in its method of content transfer and positioning.

Gaze Positioning. In this technique, gaze is used simultaneously to transfer selected content from one display to another and for positioning. The user selects content on a shared display using gaze combined with a manual trigger for confirmation. The content, now attached to the users gaze, can be moved on to the close-proximity device display and positioned by looking at the desired drop location. A second manual trigger drops the content.

Manual Positioning. Here transferring content additionally constitutes a change of modality from gaze to manual input, with final positioning also performed by manual input. The user selects content using gaze and a manual trigger. Content is then instantaneously transferred to the close-proximity device where it is held under the location defined by manual input, and positioned by the same means. A manual trigger then drops the content.

3.2 Manual Direct and Indirect Input Configurations

Here we show how we applied our Gaze Positioning and Manual Positioning techniques to include direct and indirect manual input configurations in our evaluation. This creates four conditions total, two for each technique, (1) Gaze Positioning + Direct (**GP+D**), (2) Gaze Positioning + Indirect (**GP+I**), (3) Manual Positioning + Direct (**MP+D**), (4) Manual Positioning + Indirect (**MP+I**). Direct conditions are applied to a touch enabled tablet device, and indirect conditions use a laptop with mouse input.

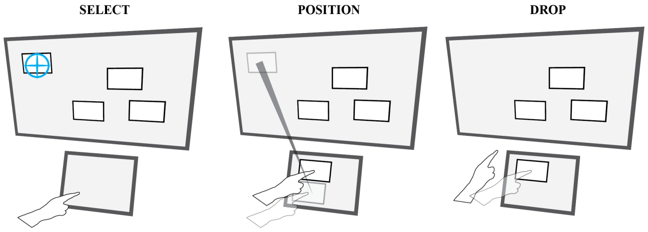


Figure 4: *Manual Positioning + Direct (Touch): (Select) Look at content on shared display, confirm with touch hold. (Position) Content is moved to touch location, touch can be moved to position the content. (Drop) Touch release drops the content.*

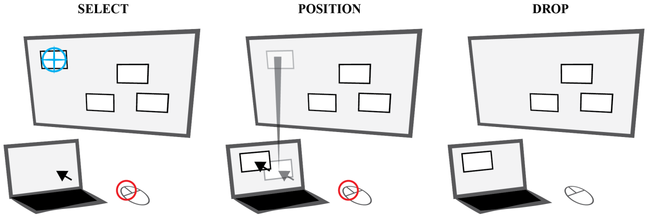


Figure 5: *Manual Positioning + Indirect (Mouse): (Select) Look at content on shared display, mouse hold confirms selection. (Position) Content is moved to mouse location, mouse can be moved to position the content. (Drop) Mouse release drops the content.*

Figure 2 shows the flow of interaction for **GP+D**, touch hold and release act as respective confirmation triggers for selection and drop, and positioning is performed by gaze. Figure 3 shows **GP+I**, a mouse hold and release confirm selection and drop, positioning is again performed by gaze. In manual positioning conditions, **MP+D** uses touch for positioning, selection, and drop confirmation (see Figure 4). **MP+I** uses mouse input to position content and mouse hold and release for selection and drop (see Figure 5).

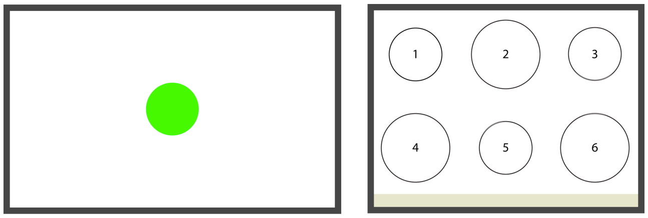
4 Study: Gaze vs. Manual, Direct vs. Indirect

We designed a user study to compare our four conditions (**GP+D**, **GP+I**, **MP+D**, **MP+I**). Our experiment aimed to answer the following research questions: (1) How does switching modalities during transfer affect performance and usability? (2) Does the type of manual input, direct or indirect, affect performance and usability?

4.1 Design & Setup

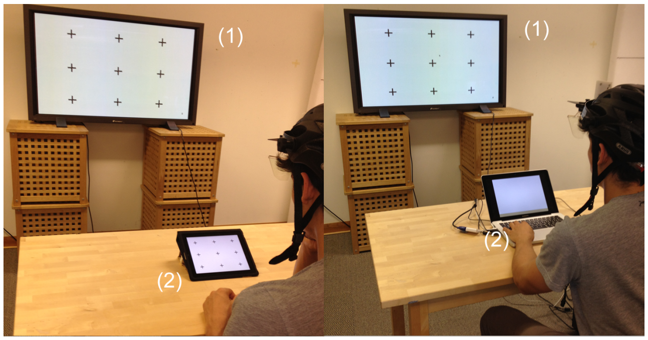
The study followed a repeated-measures within-subjects design with two independent variables, the first is *technique* with four levels: (1) **GP+D**, (2) **GP+I**, (3) **MP+D**, (4) **MP+I**. The second is *target size*, which varied between (1) Small, and (2) Large. The dependent variables were (1) Task completion time, (2) Accuracy, and (3) Number of errors.

The four top-level conditions were counter-balanced into eight unique orders to reduce learning effects. In our experimental setup (detailed below) 1° of visual angle equated to ~40 px on the shared display, ~44 px on the laptop, and ~54 px on the tablet. Objects were circular in shape sized at 200 px in diameter. Targets were also circular. Small targets (200 px) were chosen to be larger than our systems' eye-tracking accuracy (~1.5°) and at least 2.5 times greater (~3.7° − 5°) across all devices to reduce systematic accuracy effects. Large targets were sized +30% (260 px) and chosen to understand if more coarse-grained positioning would affect users performance and feedback.



(a) (b)

Figure 6: *Study 1 system interface: (a) the shared display interface showing the object origin location. (b) the personal device interface showing the possible locations of targets 1– 6 and small and large target sizes. All manual interaction was initiated within the grey bar below the targets.*



(a) (b)

Figure 7: *Study 1 setup: (a)(1) Shared Display, (a)(2) Tablet with touch input, (b)(1) Shared display, (b)(2) Laptop with mouse input*

Target positions were randomised between six locations (see Figure 6(b)) and ordered in two blocks which were further counterbalanced across all participants. All initial manual input began within a grey area at the bottom of each close-proximity devices' display, the area spanned the full width of the display (1024 px) and was 60 px in height (see Figure 6(b)).

Participants. We recruited 16 participants (6 F, 10 M) aged between 21 and 38 years ($M = 26.7, SD = .8$) with normal or corrected-to-normal vision using contact lenses. Participants were asked to answer 3 questions on a 5-point Likert scale (1: no experience at all, 5: expert user) on their previous experience with *eye-tracking* ($M = 2.5, SD = 1$), *eye-based interaction* ($M = 2.6, SD = 1.25$) and *touch screen usage* ($M = 3.5, SD = 1$).

Apparatus. Two experimental setups were required, a shared display with a tablet, and a shared display with a laptop. The experiment was conducted under dimmed lighting conditions to overcome infrared and computer vision limitations. The shared display was 50" (1280 x 768 px) in size and base-mounted at 1 m. For all conditions, participants were seated at a desk ~190 cm from the shared display. For direct input conditions, a tablet was located 60 cm from the participant and affixed to a wedge mount at an angle of ~35° from the surface of the desk (see Figure 7(a)). The tablet display was sized 9.7" (1024 x 768 px). For indirect conditions, a 13" laptop (1024 x 768 px cropped) and USB mouse were situated ~60 cm from the participant (see Figure 7(b)).

Eye-Tracking Setup. Both setups used a head-mounted eye-tracker that was calibrated to each participant at the beginning of each condition of the study. The eye-tracker is a customisation of SMI's iView X HED system with an additional scene camera to detect personal device displays at close-proximity [Turner et al. 2012]. Displays are detected using brightness thresholding with contour detection (see Figure 8). By minimising contours to four points, the rectangular surface of each display can be detected. Gaze is mapped from scene-camera to on-screen coordinates using a perspective transformation. To minimise error, both scene cameras were calibrated and undistorted to provide a rectilinear space. The system was accurate to within 1.5 degrees of visual angle. To compensate for parallax error and increase accuracy across varying distances, we used two 9-point calibrations, one for each display. The system transparently switched between these two calibrations in real time depending on which screen was detected. A moving average filter was applied to incoming eye movement data to reduce jitter and improve precision. This system was used in a controlled environment where users were seated directly in front of displays, the screen detection and gaze mapping were robust to up and down head movements under these conditions.

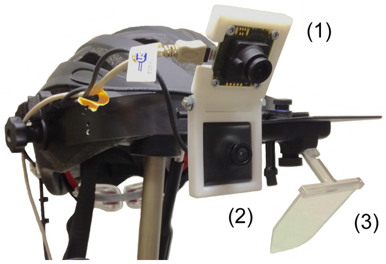


Figure 8: *Eye-tracking system: (1) and (2) show the scene cameras used to detect shared and close-proximity displays. (3) Eye camera*

Recorded Data. For each trial we recorded a timestamped log containing, object selection events, object drop events, touch events, mouse events, and errors. Errors were classified in to three categories: Failure to select on the first attempt (select failure), failure to drop within the bounds of the correct display (drop failure), and failure to drop within the target bounds (out-of-bounds failure).

4.2 Procedure

Participants first completed a demographics questionnaire on previous eye-tracking experience and eye health information. Each participant was seated at the desk with the laptop or tablet located in front of them. The head-mounted eye-tracker was then fitted and calibrated.

For each condition, participants were guided through 1 complete trial preceding 24 recorded trials. Trials were divided into two blocks, one using small targets and one using large targets. To begin each trial, participants looked at a green circle (small) centred on the shared display (see Figure 6(a)), a tap or click then started the trial. A red circle appeared in the centre of the display. Participants had to select and position the red circle within a target that would appear on the close-proximity device at one of the six defined target locations. As the circle entered a target area, its colour changed from red to grey signalling the participant to drop the object.

After each condition, participants were asked to complete a questionnaire with six questions on perceived speed, accuracy, ease of learning, the understandability of each technique, the intuitive sequence of each technique, control, and preference. For each condition, participants were also asked to complete a NASA Task Load Index exercise. The purpose of this was to gather standardised qualitative metrics that can be affected when users are subjected to unfamiliar interaction. At the end of the study, a final questionnaire was provided to gather comparative feedback and preferences for each of the four conditions.

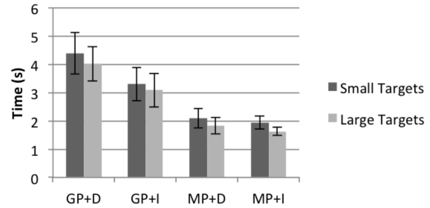


Figure 9: *Study 1 mean completion times in seconds with 95% confidence intervals (CI) for each condition with small and large targets.*

4.3 Results

4.3.1 Completion Time

Participants completed a total of 1536 trials. Figure 9 shows the mean completion times for small and large target trials. These times are calculated from the time an object was selected to when it was dropped within a target. Using a two-way repeated measures ANOVA we found no interaction effect between techniques and target size, however a significant main effect was found for techniques ($F_{3,41} = 36.130, p < .001$). We compared the times of each technique using a one-way repeated measures ANOVA. When positioning objects in small targets we found significant differences between our four conditions ($F_{3,45} = 25.708, p < .001$). Post-hoc analysis using paired samples t-tests (Bonferroni corrected ($p = 0.05/6 = .0083$)) revealed that manual techniques were significantly faster than their gaze based counterparts, **GP+D**-**MP+D** ($p < .001$), **GP+D**-**MP+I** ($p < .001$), **GP+I**-**MP+D** ($p < .012$), **GP+I**-**MP+I** ($p < .002$). No significance was found between **GP+D**-**GP+I** and **MP+D**-**MP+I**.

We also found significant differences when positioning objects in large targets ($F_{3,45} = 20.223, p < .001$). As with small targets, manual conditions significantly outperformed gaze-based conditions, **GP+D**-**MP+D**, ($p < .001$), **GP+D**-**MP+I** ($p < .001$), **GP+I**-**MP+D** ($p < .011$), **GP+I**-**MP+I** ($p < .001$). We also found that **GP+I** was significantly faster than **GP+D** ($p < .030$) when positioning in larger targets.

4.3.2 Error Rate and Accuracy

When analysing failure types we found that for small or large targets there were no significant differences in select failure rates between conditions. We did however find that there was significance ($F_{3,45} = 11.603, p < .001$) when dropping in small targets. There was a significantly higher drop failure rate for **GP+D** with small targets when compared to manual input conditions, **GP+D**-**MP+D** ($p < .006$), **GP+D**-**MP+I** ($p < .006$). The same is true for large targets ($F_{3,45} = 6.626, p < .001$), **GP+D**-**MP+D** ($p < .016$), **GP+D**-**MP+I** ($p < .016$). There was no significant difference between drop failure rates for large and small objects.

We found no significant difference in out-of-bounds failures between conditions but out-of-bounds failure rates were significantly higher between small and large targets for each condition, **GP+D**-Small - **GP+D**-Large ($t_{15} = 2.449, p < .027$) and **GP+I**-Small - **GP+I**-Large ($t_{15} = -2.324, p < .035$). No other pairs showed statistical significance.

For positioning accuracy, a one-way repeated measures ANOVA (Greenhouse Geisser corrected) showed statistical significance between conditions for small ($F_{2.386,35.789} = 72.498, p < .001$) and large ($F_{3,45} = 47.874, p < .001$) target sizes (Sphericity Assumed).

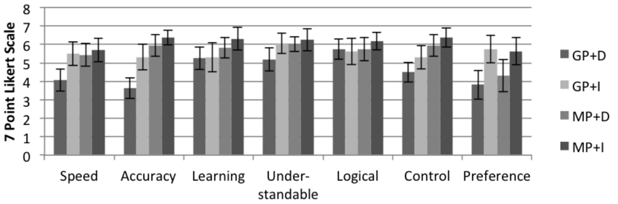


Figure 10: *Study 1 mean quantitative feedback responses*

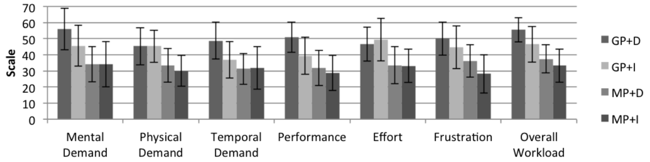


Figure 11: *Study 1 NASA TLX responses*

Post-hoc tests showed that **GP+D** was significantly less accurate for small targets than **MP+D** ($p < .001$) and **MP+I** ($p < .001$). This is also true of **GP+I** compared with **MP+D** ($p < .001$) and **MP+I** ($p < .001$). For large targets, **GP+D** was significantly more accurate than **GP+I** ($p < .027$) but less accurate than **MP+D** ($p < .001$) and **MP+I** ($p < .001$). **GP+I** was also found to be less accurate than manual input based conditions **MP+D** ($p < .001$) and **MP+I** ($p < .001$). In a paired t-test comparison between small and large target tasks, three of the four conditions were found to be more accurate with larger targets, **GP+D**-Small - **GP+D**-Large ($t_{15} = -5.337, p < .001$), **MP+D**-Small - **MP+D**-Large ($t_{15} = -3.357p < .004$), **MP+I**-Small - **MP+I**-Large ($t_{15} = -5.633, p < .001$).

4.3.3 Questionnaire

Questionnaire responses were analysed using a Friedman test and post-hoc Wilcoxon-Signed Rank tests (Bonferroni corrected ($p = 0.05/6 = .0083$)).

Figure 10 summarises questionnaire responses for each condition. Significance was found for perceived speed ($\chi^2_3 = 13.356, p < .004$) with **GP+I** being perceived as faster than **GP+D** ($Z = -2.954, p < .003$) and **MP+I** being faster than **GP+D** ($Z = -2.994, p < .003$). In terms of perceived accuracy ($\chi^2_3 = 28.061, p < .001$), **GP+D** was perceived to be the least accurate when compared against all conditions **GP+D**-**GP+I** ($Z = -3.248, p < .001$), **GP+D**-**MP+D** ($Z = -3.338, p < .001$), **GP+D**-**MP+I** ($Z = -3.533, p < .001$). Participants perceived to be most in control ($\chi^2_3 = 22.130, p < .001$) with **MP+D** ($Z = -2.891, p < .004$) and **MP+I** ($Z = -3.574, p < .001$) when compared with **GP+D**. No significant difference was found for perceived control in comparisons with **GP+I**. In addition, no significant differences were found as to which condition seemed most intuitive to participants. The same result was also found for ease of learning, understandability, and preference.

4.3.4 NASA TLX

Mean responses from NASA TLX worksheets are shown in Figure 11. Significant differences were found for the following factors, Mental Demand ($\chi^2_3 = 14.110, P < .003$), Physical Demand ($\chi^2_3 = 10.762, p < 0.13$), Temporal Demand ($\chi^2_3 = 10.838, p < .013$), Frustration ($\chi^2_3 = 12.691, p < .005$), and Overall Weighted Workload ($\chi^2_3 = 14.325, p < .002$). Post-hoc analysis revealed that **MP+I** required significantly less mental demand than **GP+I**

($Z = -2.630, p < .001$). **MP+D** required less physical demand than **GP+I** ($Z = -2.701, p < .007$). **MP+I** had less temporal demand than **GP+I** ($Z = -2.701, p < .007$) and **MP+D** was less frustrating than **GP+I** ($Z = -3.223, p < .001$). No other significant results were found between techniques and the mean reported scales. We found that **GP+I** required significantly more workload than **MP+D** ($Z = -3.464, p < .001$) and **MP+I** ($Z = -3.103, p < .002$). No significance in workload was found between other conditions.

4.3.5 Subjective Feedback

Five questions were asked at the end of each session to compare the following factors: (1) Gaze Input vs. Touch Input, (2) Gaze Input vs. Mouse Input. (3) Touch Input vs. Mouse Input, (4) Transferring content using Gaze, (5) Transferring content to Manual Input.

Subjectively, participants responded primarily in favour of touch over gaze stating several factors including, greater accuracy, control, and that *human eyes are not used to moving objects*. One participant pointed out that gaze was a useful complimentary modality for reaching while keeping manual input as the primary modality, *after gazing, the mouse would still be used*. In favour of gaze positioning, one participant envisaged dragging content with the eyes across multiple displays.

The manual positioning condition allowed for content to be handed from gaze to manual input for final positioning. Participants were asked if this mechanism seemed logical when used. The direct variant of this technique was again subjectively preferred with one participant stating that *[using the] mouse seemed like a weird "hand over"* and *touch seemed to flow naturally after eye movement*. Overall participants found this mechanism logical, although visualisation of transfer was found to be confusing for one participant, *[it was] confusing, [the] content vanishing from [the shared display] and appearing on the [tablet] under your finger*.

4.4 Summary

The results of our study show that conditions with manual input i.e., **MP+D** and **MP+I**, were significantly faster than their respective gaze-based counterparts i.e., Gaze Positioning + Direct and Gaze Positioning + Indirect. NASA TLX results also align with this finding, showing higher workload for gaze variants. Our results show that **GP+D**, **MP+D**, and **MP+I** were all significantly more accurate with large targets. One would expect large targets to result in more accurate positioning. In light of this result, in our next study we use only large targets. Participants reported that if eye-tracking accuracy had been higher, it would have been preferred. In contrast, it was also reported that the eyes are not accustomed to human-computer interaction which could account for lower performance.

Subjective responses provided several interesting insights. Firstly, a preference toward **MP+D** was reported due to a greater sense of control, and that handing over content from gaze to touch felt more natural than to mouse. This is interesting as both gaze and touch are direct input modalities, which could account for the perceived naturalness of this technique. Secondly, the visualisation of transfer in manual positioning techniques lead to confusion, this was due to content being transferred immediately upon triggering selection. This confusion was minor but considerations for future development are made in our discussion section. Thirdly, one participant noted how gaze is useful as a secondary modality, with manual remaining primary. This is applicable to our manual positioning techniques where gaze is only used for distant selection and not local close-proximity tasks.

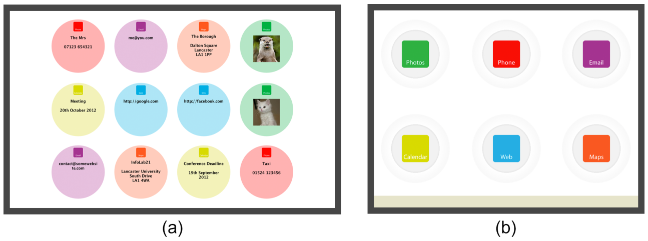


Figure 12: *Study 2 interface: (a) Shared display content layout. (b) Personal device target layout*

In a second study we evaluate how our techniques perform in an application context involving distractor targets and decision making. To further reduce the influence of accuracy issues with gaze positioning techniques, we incorporate targeting assistance.

5 Study: Positioning with Distractors

In this study we aim to understand how our techniques perform in a real-world application, in a task that involves distraction and decision making. As a consequence of our previous findings, we improved the accuracy of gaze-based input by introducing a simple targeting assistance algorithm. Our hypothesis is that this will improve performance and bring gaze input closer to manual input. The study application requires participants to select information from a shared display and, depending on the context of that information, drop it onto one of six application icons located on a tablet device.

5.1 Design & Setup

Participants and Apparatus. We recruited 8 participants (2 F, 6 M) aged between 24 and 34 years ($M = 28.3, SD = 3.8$) with normal or corrected-to-normal vision using contact lenses. One participant was medically colour blind. As in our first study, participants were asked to rate their previous experience on a 5-point Likert scale (1: no experience, 5: expert user) for *eye-tracking* ($M = 2.5, SD = 0.75$), *eye-based interaction* ($M = 2.5, SD = 0.75$) and *touch screen usage* ($M = 4.1, SD = 0.83$). We used an identical setup to our first study.

Design and Procedure. As shown in Figure 12, participants were shown 12 coloured objects on the shared display. Each contained information that they had to match up with six coloured icons displayed on either the tablet device or laptop. Objects contained information that contextually matched one of the six target icons, these were as follows: *Phone*, *Maps*, *Photos*, *Email*, *Web*, *Calendar*. Objects and icons were coloured in red, orange, green, purple, blue, and yellow. The 12 objects were split into six pairs, one pair per icon category. Objects and icons were ordered randomly for each condition. Conditions were counter-balanced as described in our first study. Only large target sizes were used as we found participants were more likely to succeed across all conditions.

Participants were verbally guided through one trial of the study and then asked to complete a further 11 recorded trials, matching up objects to target icons. The following measures were recorded: completion time from selection to drop, drop failures (object dropped on incorrect display), incorrect drop failures (object dropped on incorrect target icon). We did not use erroneous trials in our completion time analysis. Qualitative feedback was again obtained though questionnaires between conditions and at the end of the study.

Gaze Targeting Assistance. Results from our first study indicated that gaze-based input could benefit from higher accuracy. To help participants to position icons on the tablet and laptop we imple-

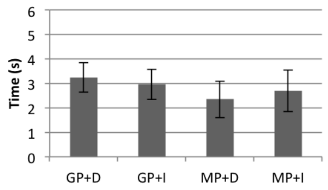


Figure 13: *Study 2 mean completion times with 95% CI*

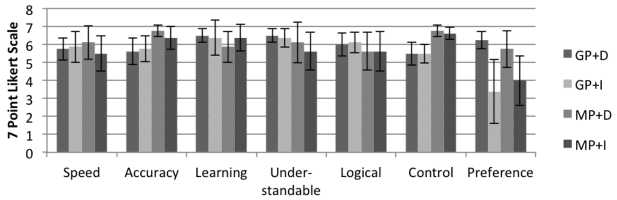


Figure 14: *Study 2 mean quantitative feedback responses*

mented a simple targeting assistance algorithm that smoothly snaps objects to target centres. For each incoming gaze sample, the algorithm first checks whether the gaze point is within the bounds of a target. If so, the centre of this target is added as a point in the moving average used to smooth raw gaze points from the eye-tracking system. The effect of this smoothly moves the object to the centre of the target. If the gaze point is not within the bounds of a target, the object is positioned at the true gaze location.

5.2 Results

5.2.1 Completion Time, Errors & Quantitative Feedback

Figure 13 shows the task completion times for this experiment but we did not find any significant results using a one-way repeated measures ANOVA. We did not measure drop accuracy as targeting assistance was used in the gaze based conditions. We recorded errors where participants dropped objects on incorrect icons. Paired samples t-tests found no significant results.

Participants were asked to provide responses to the same six scales used in our first study, these are summarised in Figure 14. A Friedman test showed significance in responses to the control scale ($X^2_3 = 14.318, p < .003$). Post-hoc analysis with a Wilcoxon-Signed Rank Test (Bonferroni corrected ($p = (0.05/6) = .0083$)) showed that participants perceived significantly less control with **GP+I** over **MP+I** ($Z = -2.714, p < .007$).

Participants were also asked to complete a NASA Task Load Index worksheet (see Figure 15). A Friedman test showed no significance when comparing across techniques.

5.2.2 Subjective Feedback

Participants were asked to leave free-text feedback to compare modalities and content transfer methods. These questions were the same as in our first experiment.

When comparing gaze to direct and indirect manual input variants, participants wrote primarily in favour of gaze positioning, stating *it was easier and less effort*, although manual positioning also received positive remarks. With regard to transfer, participants responded positively to gaze in comparison to switching modalities, *dependency on peripheral devices seem like a bit of an ordeal*, although generally both gaze and manual approaches were reported as logical. One participant reported that transfer to mouse was preferred due to its firm grounding in human computer interaction.

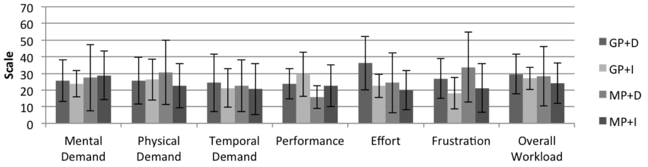


Figure 15: *Study 2 NASA TLX responses*

6 Discussion

Our initial study aimed to answer the following questions, (1) How does switching modalities during transfer affect performance and usability? (2) Does the type of manual input, direct or indirect, affect performance and usability?

In answer to our first question, results from study 1 showed that switching to manual input for final positioning was significantly faster than positioning with gaze. After applying targeting assistance in a second study to reduce system accuracy effects, we found all conditions to be similar in performance. This also lead to a reduction of workload in gaze conditions, revealed by NASA TLX feedback. It should be noted that in a real-world implementation, targeting assistance is but one solution to increase selection accuracy, another approach may size targets dependent on proximity to a display. Study 2 resulted in longer completion times, which we account to the distraction and decision making involved in the task, this did not lead to increased workload, thus validating our techniques' usability in such a task.

Subjective feedback from both studies revealed that the sequence of interaction for all conditions seemed logical to participants. In line with findings from our previous work [Turner et al. 2013], study 2 participants responded favourably towards gaze positioning conditions, describing the automatic nature of content following the eyes as easier. This is converse to the results of study 1 where manual touch was stated as more natural. This difference is most likely due to the improved accuracy of gaze in study 2 and demonstrates that participants were generally accepting of either approach. It was also noted that gaze was useful as a complementary modality to manual input to enable distant reaching, this aligns with the findings of [Zhai et al. 1999]. This also suggests that some users may prefer to use gaze only for distant acquisition, and a primary manual modality for close-proximity tasks.

Our second question focussed on direct and indirect input comparison and how transfer to each from gaze was regarded by participants. In study 1 it was commented that transferring from gaze to touch felt more natural than from gaze to mouse. Results from study 2 revealed a stronger preference to direct variants and a reduction in indirect preferences. It is not clear if this can be attributed to preference toward touch in general but also raises further questions i.e., is there a deeper relationship between gaze and touch? Does the direct-ness of each modality make interaction in this space more natural to users? Whether this is the case or that this result is dependent on individual user preference warrants further investigation.

Participants reported some confusion regarding the visualisation of transfer with manual input positioning. Upon selection, content would be removed from the shared display and transferred immediately to the personal device for positioning. We believe further consideration should be made around timing when removing content from view. For instance, in a scenario with multiple shared displays, a user may not wish to transfer to their personal device but instead between shared displays. In this case, it would make more sense to keep content visible until the direction of transfer can be inferred. Only when the change of context is clear, should content be moved. Furthermore, the type of personal device could also affect interaction, for instance if a smartphone with a small display is used, pragmatically only manual input would perform final positioning due to the affordance of higher accuracy and precision. We consider these areas for further investigation.

Previous work has shown how gaze can be combined with touch to support fine-grained positioning within a single display [Stellmach and Dachselt 2013]. Currently, only our manual positioning techniques allow very high granularity in this respect. Our gaze conditions use targeting assistance, however other methods would need to be devised or adopted to provide more control when positioning with gaze. One solution would be to use target expansion, zooming mechanisms, or manual refinement as demonstrated in the related work [Stellmach and Dachselt 2012; Stellmach and Dachselt 2013].

7 Conclusion

In this paper we investigated how gaze and manual input can be used for point-to-point content transfer between distant and close-proximity displays. Our first study showed that manual positioning techniques outperformed gaze positioning, we attribute this result to system accuracy differences between conditions. A second study incorporated targeting assistance and we found that all conditions performed similarly under a real application scenario. Our studies revealed that transfer between gaze and direct input felt most natural. Additionally, this work poses further questions of how best to visualise transfer between gaze and manual input modalities. Overall, no one condition performed best. Both Gaze Positioning and Manual Positioning are promising solutions for the transfer and positioning of gaze-acquired content. Users showed individual preference to techniques and modalities.

References

BAUDISCH, P., CUTRELL, E., ROBBINS, D., CZERWINSKI, M., TANDLER, P., BEDERSON, B., AND ZIERLINGER, A. 2003. Drag-and-pop and drag-and-pick: Techniques for accessing remote screen content on touch and pen-operated systems. In *Proc. INTERACT '03*.

BORING, S., BAUR, D., BUTZ, A., GUSTAFSON, S., AND BAUDISCH, P. 2010. Touch projector: mobile interaction through video. In *Proc. CHI '10*.

BRAGDON, A., DELINE, R., HINCKLEY, K., AND MORRIS, M. R. 2011. Code space: touch + air gesture hybrid interactions for supporting developer meetings. In *Proc. ITS '11*.

COLLOMB, M., AND HASCOËT, M. 2008. Extending drag-and-drop to new interactive environments: A multi-display, multi-instrument and multi-user approach. *Interacting with Computers 20*.

DREWES, H., AND SCHMIDT, A. 2009. The magic touch: Combining magic-pointing with a touch-sensitive mouse. In *Proc. INTERACT '09*.

DREWES, H., DE LUCA, A., AND SCHMIDT, A. 2007. Eye-gaze interaction for mobile phones. In *Proc. Mobility '07*.

HARDY, R., AND RUKZIO, E. 2008. Touch & Interact: Touch-based interaction of mobile phones with displays. In *Proc. MobileHCI '08*.

HINCKLEY, K., RAMOS, G., GUIMBRETIERE, F., BAUDISCH, P., AND SMITH, M. 2004. Stitching: Pen Gestures That Span Multiple Displays. In *Proc. AVI '04*.

JACOB, R. J. K. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proc. CHI '90*.

MARDANBEGI, D., AND HANSEN, D. W. 2011. Mobile gaze-based screen interaction in 3d environments. In *Proc. NGCA '11*.

MARDANBEGI, D., HANSEN, D. W., AND PEDERSON, T. 2012. Eye-based head gestures. In *Proc. ETRA '12*.

MILUZZO, E., WANG, T., AND CAMPBELL, A. T. 2010. Eyephone: activating mobile phones with your eyes. In *Proc. MobiHeld '10*.

MONDEN, A., ICHI MATSUMOTO, K., AND YAMATO, M. 2005. Evaluation of gaze-added target selection methods suitable for general guis. *IJCAT 24*, 1.

NACENTA, M. A., SALLAM, S., CHAMPOUX, B., SUBRAMANIAN, S., AND GUTWIN, C. 2006. Perspective cursor: Perspective-based interaction for multi-display environments. In *Proc. CHI '06*.

NACENTA, M. A., MANDRYK, R. L., AND GUTWIN, C. 2008. Targeting across displayless space. In *Proc. CHI '08*.

NAGAMATSU, T., YAMAMOTO, M., AND SATO, H. 2010. Mobigaze: development of a gaze interface for handheld mobile devices. In *Proc. CHI EA '10*.

POITSCHKE, T., LAQUAI, F., STAMBOLIEV, S., AND RIGOLL, G. 2011. Gaze-based interaction on multiple displays in an automotive environment. In *Proc. SMC '11*.

REKIMOTO, J., AND SAITOH, M. 1999. Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments. In *Proc. CHI '99*.

REKIMOTO, J. 1997. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proc. UIST '97*.

SCHMIDT, D., CHEHIMI, F., RUKZIO, E., AND GELLERSEN, H. 2010. PhoneTouch: A Technique for Direct Phone Interaction on Surfaces. In *Proc. UIST '10*.

STELLMACH, S., AND DACHSELT, R. 2012. Look & touch: Gaze-supported target acquisition. In *Proc. CHI '12*.

STELLMACH, S., AND DACHSELT, R. 2013. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proc. CHI '13*.

STELLMACH, S., STOBER, S., NÜRNBERGER, A., AND DACHSELT, R. 2011. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. NGCA '11*.

TURNER, J., BULLING, A., AND GELLERSEN, H. 2012. Extending the visual field of a head-mounted eye tracker for pervasive eye-based interaction. In *Proc. ETRA '12*.

TURNER, J., ALEXANDER, J., BULLING, A., SCHMIDT, D., AND GELLERSEN, H. 2013. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Proc. INTERACT '13*.

ZHAI, S., MORIMOTO, C., AND IHDE, S. 1999. Manual and gaze input cascaded (magic) pointing. In *Proc. CHI '99*.