Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch

Jayson Turner[1], Jason Alexander[1], Andreas Bulling[2], Dominik Schmidt[3], and Hans Gellersen[1]

[1] School of Computing and Communications, InfoLab21, Lancaster University, Lancaster, LA1 4WA, United Kingdom
[2] Max Planck Institute for Informatics, Perceptual User Interfaces Group, Campus E1 4, 66123 Saabrücken, Germany
[3] The Human Computer Interaction Lab, Hasso Plattner Institute, Prof-Dr-Helmert Str. 2-3, 14482 Potsdam, Germany
{j.turner,j.alexander}@lancaster.ac.uk,
andreas.bulling@acm.org, dominik.schmidt@hpi.uni-potsdam.de,
hwg@comp.lancs.ac.uk

Abstract. Previous work has validated the eyes and mobile input as a viable approach for pointing at, and selecting out of reach objects. This work presents Eye Pull, Eye Push, a novel interaction concept for content transfer between public and personal devices using gaze and touch. We present three techniques that enable this interaction: Eye Cut & Paste, Eye Drag & Drop, and Eye Summon & Cast. We outline and discuss several scenarios in which these techniques can be used. In a user study we found that participants responded well to the visual feedback provided by Eye Drag & Drop during object movement. In contrast, we found that although Eye Summon & Cast significantly improved performance, participants had difficulty coordinating their hands and eyes during interaction.

Keywords: Eye-Based Interaction, Mobile, Cross-Device, Content Transfer, Interaction Techniques.

1   Introduction

We are surrounded by out-of-reach digital information. Our private TVs and public shared displays often present URLs, physical addresses, phone numbers, route descriptions, and other information that we wish to 'pull' to our personal devices. Equally, we often wish to add personal content to notices, discussions, presentations and collections on shared screens. Yet we lack fluid mechanisms for moving content between public and personal displays.

We present Eye Pull, Eye Push, a novel interaction concept that allows for the acquisition (pulling) and publication (pushing) of content between personal and remote devices. Using a combination of gaze and touch it is possible to define techniques that enable this interaction style. Gaze is a natural modality choice for selecting objects that catch our visual attention, while touch actions can be performed on personal
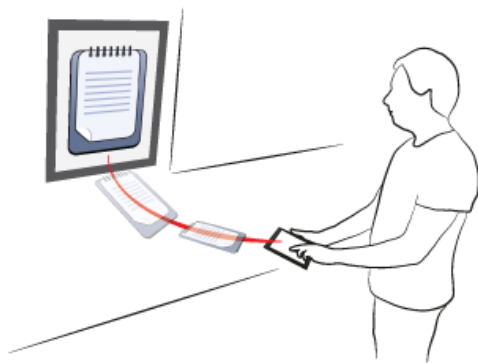
Fig. 1. Eye Pull, Eye Push: users pull and push objects between remote screens and their personal devices with a combination of gaze and touch. In this scenario, the user selects a form on a public service terminal simply by looking it, retrieves it to their touch device with a swipe, fills it in, and returns it with a swipe while looking up at the terminal.

devices without visual attention. Related work has employed handheld device input, combined with gaze interaction, to assist panning and zooming [22] and target acquisition [20] on desktop displays. Our work is distinct in demonstrating the gaze-supported transfer of objects across devices.

Figure 1 illustrates our vision: a user selects an object on a public display and, while still visually fixating on it, swipes on their handheld personal device to pull the object down for editing. Once editing is complete, the user re-fixates on the remote target, and returns the object with a further touch gesture. This style of interaction would benefit many contexts of use: group collaboration, classrooms, and public community displays [11, 17, 7]; our homes for lazy interaction between the TV screen and mobile devices; public terminals that we may find too exposed or too grimy for direct data entry; and anywhere that digital objects exist, that users would like to edit but cannot reach.

This paper makes a two-fold contribution. First, we introduce Eye Pull, Eye Push, a concept for multimodal cross-device content transfer. We define the required input attributes for such interaction and explore application scenarios where it makes a compelling impact.

Second, we define three novel techniques for the transfer of objects between remote screens and personal touch devices, each combines gaze and touch: *Eye Cut & Paste* (ECP): Objects are cut and pasted using gaze and touch tap events. *Eye Drag & Drop* (EDD): Objects are moved using gaze and touch hold/release events. *Eye Summon & Cast* (ESC): Objects are pulled using gaze and a swipe down action, and pushed using gaze and a swipe up action. All three techniques were implemented using a portable eye tracker extended for wider field of view [24]. We evaluated these techniques in a user study to understand their strengths and weaknesses in performance and usability. The results demonstrate that users are able to transfer content efficiently using our techniques, thus validating our approach. ECP and EDD performed similarly, with EDD being preferred due to the continuous visual feedback provided by drag-and-drop. ESC was the fastest of the techniques but was rejected by users due to the more complex hand-eye coordination required.

2   Related Work

2.1   Cross-Device Information Transfer

The case for moving objects easily between handheld devices and larger screens has been made widely, for group work settings [11, 17] as well as serendipitous encounters with public displays [7, 1].

Several works have focused on pushing and pulling content using touch-surfaces as proxies to public displays. Touch Projector [5] demonstrated improvements to work by Tani et al. that enabled the control of remote machinery through live video feeds while maintaining spatial context [23]. Boring's work made use of a phone camera feed to project touches on to public displays to manipulate and move objects. Similarly, Bubble Radar showed how users could interact at a distance using the representation of a public display on a tablet PC [1].

Bragdon et al. developed Code-space, a set of techniques focused on interactions between mobile and situated devices in developer meetings [6]. Their work utilised situated depth cameras and inertial sensors embedded in mobile devices to enable intuitive pointing and information transfer for collaboration. Earlier work by Rekimoto et al. entitled Pick-and-Drop [18] has shown how physical objects can be used to transfer content from one display to another. In this case a pen was used to represent a faux storage device that could pick and drop content.

Several techniques have explored obtaining content at a distance within a single large display. Baudisch et al. investigated different techniques for dragging and dropping objects [3]. Their Drag-and-Pop and Drag-and-Pick techniques used proxies of distant icons to effectively bring them closer to a user. Drop-and-Drag by Doeweling et al. [9] was a technique similar to traditional drag and drop technique that allowed for interaction to be suspended mid-transfer, thus allowing the user to perform fine-grained navigation before dropping an object. The above techniques were all found to be faster than traditional drag and drop for sufficiently distant targets. Finally, Schmidt et al. [19] described a range of interactions made available by combining a mobile phone with a multi-touch surface. Their techniques allow for fluid content transfer, personalisation of the surface and access-control over publicly visible elements.

2.2   Gaze Pointing

Early work on eye-based interaction showed that the eyes could be used as input in desktop environments. However an issue coined by Jacob et al. known as the *Midas Touch Problem* causes unwanted interactions when trying to explicitly issue commands [12]. Dwell-time overcomes this problem by allowing a user to fixate on a control for a set delay before activation occurs. Studies by Jacob et al showed that the delay incurred by dwell-time could be overridden, by using manual input to activate controls interaction can be sped up.

Prior to Jacob et al., Ware et al. [25] examined three picking techniques that used gaze combined with dwell, a virtual button and a hardware button for selection. Their experiments found that confirmation via a hardware button was fastest. It was also found that users would attempt to synchronise their eye movement with hardware button presses, causing occasional selection errors as the eyes move away before

selection is confirmed. A fully developed alternative to mouse input using gaze and keyboard commands was demonstrated by Kumar et al. [13].

Further studies have evaluated gaze as an assistive modality for manual input. Zhai et al. [26] developed MAGIC pointing. In their paper they designed two techniques that combined gaze with mouse input: *liberal*, the mouse cursor is warped to objects being looked at, the final selection is performed by the mouse, *conservative*, the mouse cursor is only warped after the user moves the mouse. Their experiment found that users subjectively felt they could interact faster with MAGIC techniques. Their liberal technique was faster than manual input and their conservative technique was slower. Drewes et al. [10] followed up on this experiment by combining gaze with a touch enabled mouse. They found that warping the cursor based on when the mouse was touched, as opposed to moved, reduced the need for mouse repositioning, thus improving the overall speed. Bieg et al. [4] showed however that MAGIC pointing offered no performance boost over mouse only input when used on large displays.

2.3   Multi-modal Gaze Interaction with Public Displays

Gaze-based and gaze-supported interactions with public displays are concepts already explored in the literature. Mardenbegi et al. demonstrated the use of head gestures in combination with gaze to interact with applications on a public display [15]. This work followed the same principles as in the previously described work on pointing. Gaze is used to point, and an additional modality is used to issue commands.

Stellmach et al. evaluated techniques in several works that combine gaze and mobile input, i.e., inertial sensing and touch [22, 20]. Their work developed techniques to navigate large image collections on public displays. Techniques used gaze for pointing while touch and accelerometer values were used to pan and zoom through images [22]. Users perceived increased effort and complexity when panning and zooming. This was considered as acceptable however, as it allowed for simultaneous interactions not usually possible with gaze alone. In a later work they combined gaze with touch commands. This work defined five techniques for the remote selection of varying sized targets in a desktop setting [20]. Their findings gave rise to one technique in particular, MAGIC Tab, which allowed users to tab through a series of objects within close proximity to a users gaze, thus overcoming eye tracking accuracy issues. In further work Stellmach et al. evaluated techniques that utilise a combination of eye and head directed pointing with touch interaction for selection and manipulation of distant objects [21]. Their results highlighted that further improvements are required to allow for more precise distant cursor control with large displays.

2.4   Summary

The literature demonstrates success both using multimodal eye-based interactions for remote target acquisition and using touch-based proxies for distant content interaction. Our work joins these areas by using gaze and touch to pull and push objects between public and close proximity devices.

3   Eye Pull, Eye Push

Here we describe the concept of Eye Pull, Eye Push. 'Pulling' refers to moving content from a public context to a personal one. 'Pushing' refers to the opposite of this, moving from personal to public. The overall concept presents an interaction style whereby these tasks can be completed using a combination of gaze and touch.

Below we outline three techniques designed to pull and push objects. We define the stages of interaction required to transfer content between personal and public displays, and explain how each of our techniques provides the required input attributes for each stage.

3.1   Input and Interaction Flow

The transfer of an object between a public display and a personal device can be broken down in to four main steps: object location, confirmation of selection, destination location, and confirmation of drop. Each of these requires two attributes to be fulfilled: *Locate* (the location of the object or target) and *Confirm* (an action to confirm the location). The three techniques we propose combine gaze and touch actions in different ways. They are able to execute the outlined main steps and fulfil their attributes.

Each technique uses one of three touch commands: *Tap*, *Hold/Release* and *Swipe* and each is performed with a single finger. Tap combines two touch events, touch down and touch up, performed in quick succession. Hold/Release also combines touch down and touch up but they are used in considerably slower succession to confirm actions. Swipe combines touch down, touch moved and touch up, each must be performed in quick succession for the gesture to be recognised. The mappings of touch and gaze for each technique are shown in Table 1.

Table 1. Mapping of gaze and touch input to locate objects and confirm actions. Eye Summon & Cast is split in to two rows for clarity. Eye Summon and Eye Cast each involve a single swipe gesture (down or up) combined with gaze.

|	Object Selection		Destination Selection	
	Locate	**Confirm**	**Locate**	**Confirm**
Eye Cut & Paste	Gaze	Tap	Gaze	Tap
Eye Drag & Drop	Gaze	Hold	Gaze	Release
Eye Summon	Gaze	Swipe	Swipe	Swipe
Eye Cast	Swipe	Swipe	Gaze	Swipe

3.2   Transfer Techniques

Eye Cut and Paste. The first of our techniques is Eye Cut & Paste; it adopts the familiar Cut & Paste semantic of desktop interaction. The steps of this technique are shown in Figure 2: To pull content, the user looks at an object, they then tap on their tablet to select and cut the object from view. A 'paste' is then performed by looking at the target device and a second tap inserts the object at the gaze location. To push content from a personal display, the same steps can be used, i.e., look at an object on the tablet, tap to select, look at the public display and tap again to drop.

Alternate semantics are possible for this technique, for example, once an object is cut, many copies can be pasted to a destination.
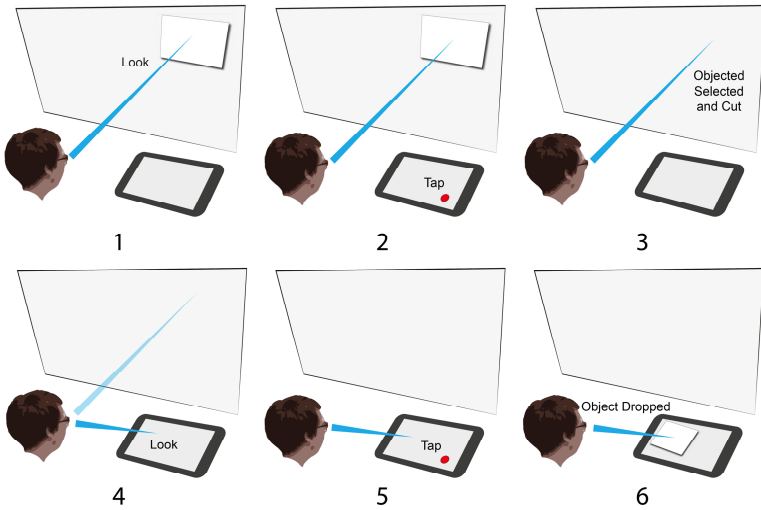
Fig. 2. Eye Cut & Paste: 1) Look at object, 2) Tap on tablet, 3) Object is selected and cut from view, 4) Look at tablet, 5) Second tap on tablet, 6) Object is dropped

Eye Drag and Drop. Our second technique, Eye Drag & Drop is likewise inspired by its desktop equivalent. Figure 3 shows, to pull content, an object is located by gaze and selected by a hold gesture. The object follows a users gaze for as long as they maintain holding with touch. As a user's gaze trajectory intersects the personal device, the object appears on the display. Once touch is released, the object is dropped. Similarly to Eye Cut & Paste, the steps of this technique can also be used to push content, i.e., look at an object on the tablet, hold touch, look at the public display and release touch.
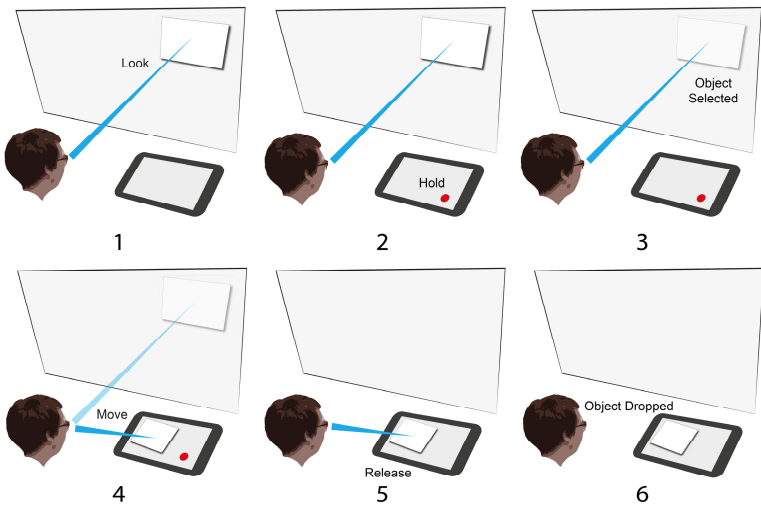
Fig. 3. Eye Drag & Drop: 1) Look at object, 2) Hold touch on tablet, 3) Object is selected and can be visibly moved, 4) Look at tablet, 5) Release touch from tablet, 6) Object is dropped
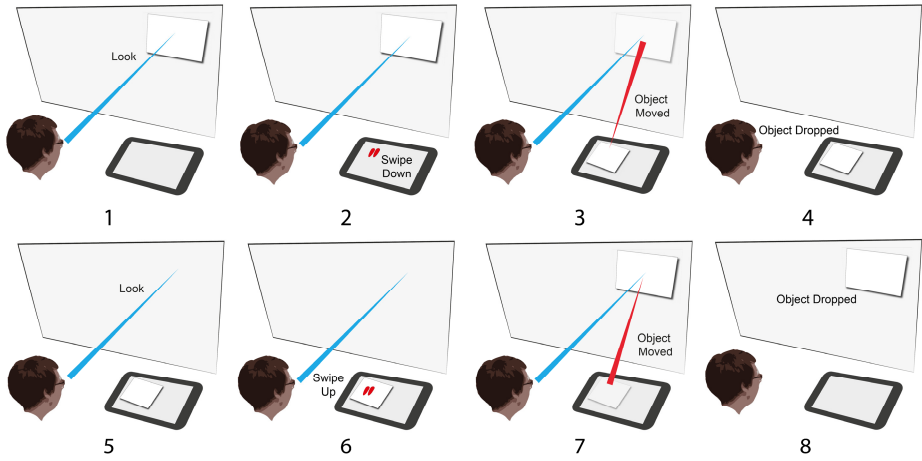
Fig. 4. Eye Summon & Cast. To summon: 1) Look at object. 2) Swipe down on tablet, 3) Object is moved to swipe location, 4) Object is dropped. To cast: 5) Look at destination, 6) Swipe up on object, 7) Object is moved to location of gaze, 8) Object is dropped

Eye Summon and Cast. Our final technique Eye Summon & Cast is based on a combination of gaze with a swipe gesture (see Figure 4). Unlike our other techniques, Eye Summon & Cast uses two differing methods (summon and cast) to pull and push content. An object on the remote screen can be located by gaze, and then summoned with a swipe down on the touch device. The swipe serves to confirm the object selection and simultaneously identifies the destination position on the target touch device. A cast is performed similarly: gaze now selects the destination, and a swipe up identifies the object to be transferred and implicitly confirms selection and drop.

Different semantics are possible for implicit identification, e.g., selecting the most recently 'pulled' object to be pushed back.

4   Application Scenarios

In the section we describe six application scenarios that demonstrate the versatility of Eye Pull, Eye Push. Each of our three techniques has been designed to complete the tasks, pull and push. As the flow of interaction differs between techniques, each can also be used for specialised tasks. Here we consider how each technique could be used in real-world scenarios to pull and/or push content. Table 2 outlines techniques, tasks and connected examples.

Table 2. Example application scenarios for each technique/task combination. Note for Eye Drag & Drop that the examples involve both tasks.

	Pull	**Push**
Eye Cut & Paste	Mid-Transfer Interaction	Duplicating for many users
Eye Drag & Drop	Sharing Read-only Content	Digital Form Filling
Eye Summon & Cast	On-the-go: Acquiring many objects	Sharing Content
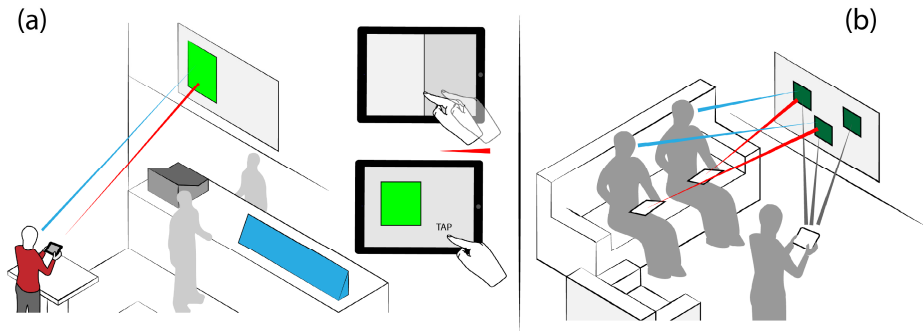
Fig. 5. (a) Mid-Transfer Interaction: A user pulls a flyer, they then switch to a suitable application before tapping to drop it. (b) Duplicating For Many Users: A user pushes three copies of an image using Eye Cut & Paste, two friends now have copies they can pull to keep

Mid-transfer Interaction. Eye Cut & Paste is analogous to desktop cut and paste. The advantages of this technique can be leveraged when pulling content. Traditional cut and paste allows for objects to be selected and temporarily stored on the clipboard. This allows for two further interactions, first it frees the user to perform other (usually navigation) tasks and second it allows for the duplication of content.

As an example, shown in Figure 5a: A user is typing up a document on a tablet pc in a café, a display above the café counter advertises weekly events. The user looks up at the display and notices a digital flyer about a music night at the café. To acquire a copy of this flyer, while still looking, the user taps on their personal device, the content is then held on the clipboard. Now, the user navigates to their calendar application, looking, taps to paste in the flyer and sets a reminder. Next the user switches to their social networking application and pastes in a second copy of the flyer to share with their friends.

Compared to our other techniques, Eye Cut & Paste is specialised to scenarios such as this, where interaction is required mid-transfer to allow content to be used for different purposes.

Duplicating for Many Users. As shown in the previous example, Eye Cut & Paste can be used for the duplication of content that has been 'cut'. The following example demonstrates how this can be leveraged when pushing content.

A user has cut and pasted a single photograph to a television to show to two other users. The users all like the picture and so want to obtain their own copies. The user performs the paste stage of the technique twice more to create additional copies on the television for the friends to pull to their own devices (see Figure 5b).

Digital Form Filling. Eye Drag & Drop is suited to tasks where changing context is part of the natural flow of interaction, where transfer is performed in a slow and continuous manner.

Figure 6a shows an art gallery, where paintings are displayed along a wall. Next to each art piece is a digital comments display containing the thoughts of gallery patrons and empty comment cards. To leave a comment, a user looks at an empty comment card and pulls it to their tablet. This is performed following the steps of Eye Drag & Drop. The user then fills in the card with their thoughts. The card is then pushed back to the comments display by the same method.
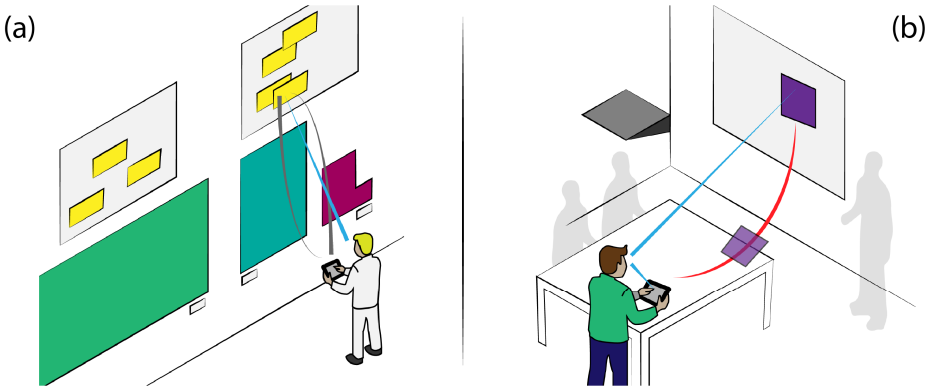
Fig. 6. (a) Digital Form Filling: A user shares their thoughts about artwork on a virtual comments board by pulling, completing and pushing a comment card. (b) Sharing Read-only Content: A user pushes and pulls an image for temporary viewing in a meeting.

As Eye Drag & Drop provides continuous visual feedback, content can be seen to visibly move as it follows a user's gaze. This allows the interaction to become analogous to physical tasks such as filling in and posting comment cards.

Sharing Read-Only Content. Users do not always want others to be able to obtain the content they share. Figure 6b shows how Eye Drag & Drop can be used to share in a read-only manner by maintaining control over content as it is displayed.

This technique allows a user to switch back and forth between large and personal display contexts in a steady and continuous manner: A user is in a meeting; they want to show a relevant image temporarily on a projected display without disturbing the current content. First they look at an image on their personal device and perform a touch hold, this attaches the image to the location of their gaze. The user then looks up at the larger display to show the picture, as they maintain holding their touch, the object does not drop. The user then reverts their eyes back to their personal device, removing the image from the large display. They then release their touch to drop the object.
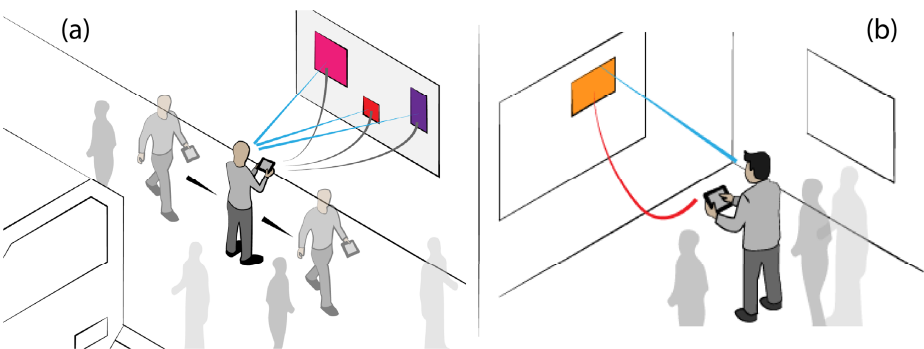
Fig. 7. (a) On-the-go: A user acquires many objects in quick succession at train a station. (b) Sharing Content: A user shares content to a display for viewing by a group of people.

On-the-go: Acquiring Many Objects. Eye Cut & Paste and Eye Drag & Drop require the user to change context between a large and personal display as they transfer an object. These two techniques are best suited to settings where the user's relative movement and schedule are not limited.

Eye Summon & Cast requires the eyes to identify a distant object, drop location is then defined by touch. This mechanism allows for the user to acquire an object without changing their visual context. This allows for the quick acquisition of many objects in sequence while on the go. Figure 7a demonstrates an example: a user has arrived in a busy train station and on the platform is a local information display. The display contains a wealth of tourist centric information on the local area. The user spots a train departure table, a local taxi number and a local map. The user swipes on their mobile device to grab each item in sequence as they pass by the display without having to change context.

Sharing Content. When browsing media on a personal device, users often want to share their experience with a large group on a bigger display. Figure 7b shows how Eye Summon & Cast can be used to allow for fluid interaction in this scenario: while browsing content, the user holds their finger on an image they wish to share. Now, looking at a larger display, the user can swipe upwards on their personal device to transfer the image for viewing. This interaction allows for a simple and natural method of choosing a public display, in particular in environments where more than one large display may exist.

5   User Study

In a user study we aimed to compare our three techniques to evaluate usability to understand which was better suited to each task and to users. We analysed performance and usability measures that were recorded as users pull and push a single object between a large display and a mounted tablet device.

5.1   Participants and Apparatus

We recruited 12 paid participants (11 male, 1 female, aged 22 to 41 (M = 25.4 S.D. = 5.1)), all had normal or corrected vision, and one was "colour-blind" but was able to distinguish the colours used in the experiment. Participants stood 150 cm from a 50" plasma display (whose base was 1 m from the floor). A tablet was mounted on a tripod at waist height. This decision was made to ensure eye-tracking accuracy remained constant throughout trials. This prevented parallax error that is inherent in monocular eye-tracking.

Participants wore a custom eye tracker that was calibrated with each participant at the beginning of the study. The eye tracker is based on SMI's iView X HED system but utilises an additional scene camera to detect personal device screens at close proximity using brightness thresholding with contour detection (see Figure 8) [24]. Contours were minimised to four points representing the rectangular surface of each screen. Gaze was then mapped to this rectangle using a perspective transformation to convert scene camera coordinates to on-screen coordinates. Although the system did
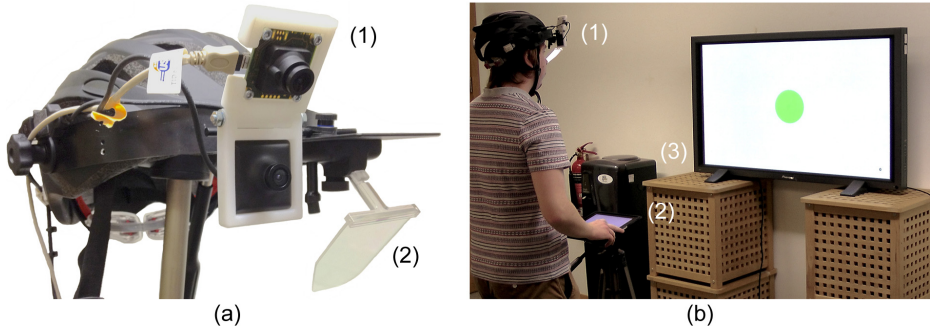
Fig. 8. System Setup: (a) Dual scene camera eye tracking system, (1) Additional scene camera (2) Eye camera. (b)(1) The system setup with head-mounted eye-tracker, (2) Touch tablet mounted on tripod (3) and Plasma TV

not use the commercial software provided, the system was accurate to within 1.5 degrees of visual angle, we found this to be sufficient accuracy for the target sizes used in this study. To compensate for parallax error, the system was calibrated twice, once for the public display and once for the tablet. The system switched between calibrations depending on which screen was in view.

5.2   Experimental Design and Procedure

The study followed a within-subjects repeated-measures design with two independent variables, *technique*, with three levels (1) Eye Cut & Paste (ECP), (2) Eye Drag & Drop (EDD), (3) Eye Summon & Cast (ESC) and *task*, with two levels (1) Pull, (2) Push. The dependent variables were task completion time and error rate. Users were asked to pull and push single objects between displays, this equated to one trial of the experiment.

For each technique participants performed 30 trials: one guided training, five practice, and 24 recorded trials. To begin a trial, participants fixated at a 175 px green circle on the public display and were asked to tap on the tablet. A red target would then appear on the public display. Targets had varying origins but were all located equidistant from the centre of the start point. This was to minimise anticipation when locating the next object. Participants pulled and dropped the object at arbitrary locations on the tablet. Upon dropping the object, its colour changed after a 5 sec delay to blue, prompting the participant to begin the push stage of the task. When pushing, the object had to be dropped within a target area double the size of the object (350 px in diameter) and in the same position from which it had been originally pulled. This was to ensure participants could complete the experiment without introducing a time penalty. All participants used the three techniques (order counterbalanced using a Latin square) and performed all trials with one technique before moving to the next. After completing all tasks with a particular technique, participants provided subjective feedback, including questions from the NASA Task Load Index (NASA-TLX). A final questionnaire gathered preference, task suitability, and general feedback.

All touch and gaze events, task completion times and errors were automatically logged. An error was logged under conditions where, selection failed on the first attempt, an object was dropped out of bounds of a target or an object was dropped out of bounds of a display.

6   Results

6.1   Task Completion Time

Participants completed a total of 864 (24 trials x 3 techniques x 12 participnats) trials. Figure 9 shows the mean completion times for each task. We compared these values in a 2 x 3 (task x technique) two-way repeated-measures ANOVA with Greenhouse Geisser correction. An interaction effect was found ($F_{1.721,18.933}$=5.178, p=.020). Further tests using a one-way repeated measures ANOVA with Greenhouse Geisser correction showed a significant difference for the pull task in completion time between the three techniques ($F_{1.992, 21.137}$=33.812, p<.0005). Further paired t-tests (Bonferroni corrected, new p-value=0.0083) showed that ESC was significantly faster than EDD (p<.0005) and ECP (p<.0005). ECP and EDD were not found to be significantly different (p=1.000).

For the push task, a significant difference was found across completion time ($F_{1.704, 18.749}$=19.235, p<.0005). Further post-hoc paired t-tests (Bonferroni corrected, new p-value=0.0083) showed that ESC was significantly faster than EDD (p<.001) and ECP (p<.001). A significant difference was not found between ECP and EDD when pushing objects.

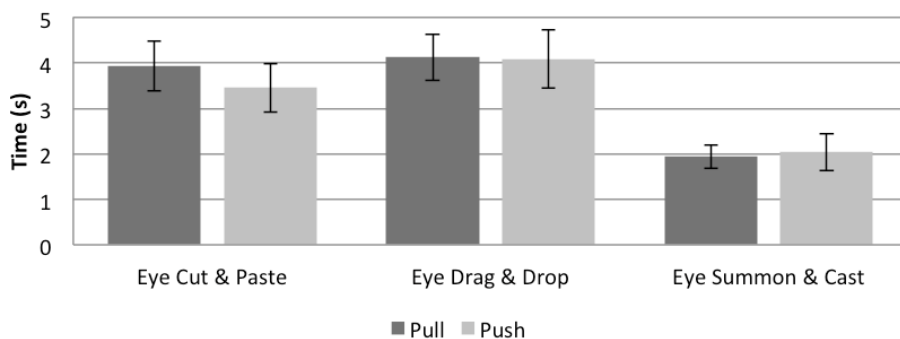No significant differences were found between tasks for each technique.

Fig. 9. Mean task completion time in seconds with 95% confidence intervals (CI)

6.2   Error Rate

The mean error rates for each technique are shown in Figure 10. In a 2 x 3 (task x technique) two-way repeated measures ANOVA with Greenhouse Geisser correction, we found no significant interaction or main effects. The means are calculated from 288 trials per technique per task. ECP showed a mean error rate of 1.58 for pulling

and 1.66 for pushing. EDD had a higher mean error for pushing than pulling, (2.25 and 0.83 respectively). ESC had a slightly higher mean error rate for pushing (0.83) than pulling (1.25) also.
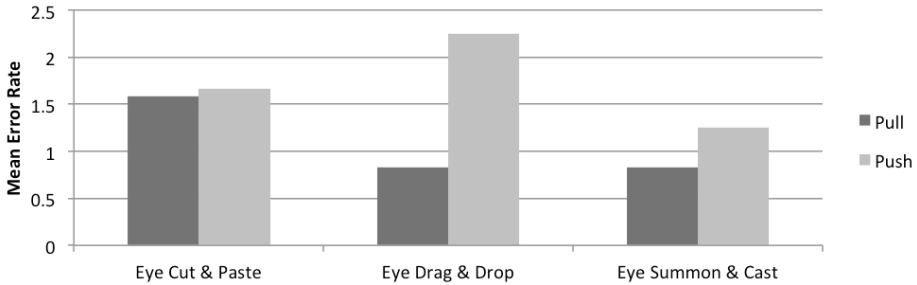

Fig. 10. Mean error rates, confidence levels omitted for clarity

6.3   Performance Perception

We recorded participant responses on a 7-point likert scale to questions regarding perceived speed, accuracy, ease of learning, suitability to task and preference. Friedman tests showed no significant differences in perceived speed for pulling or pushing objects, overall speed, accuracy or ease of learning. Participants were asked questions for each technique relating to their suitability and preference for the two tasks. No one technique was significantly suited to or preferred for pulling objects. For pushing objects (X^2(2)=9.500, P<=.009) ESC was significantly less preferred than EDD (Z=-2.756, P<.006) with no significant difference between other techniques and EDD.
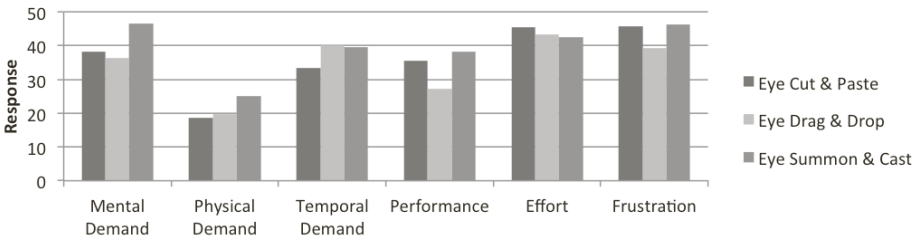

Fig. 11. NASA Task Load Index, scale 0-100, confidence levels omitted for clarity. Key: (ECP) Eye Cut & Paste, (EDD) Eye Drag & Drop, (ESC) Eye Summon & Cast.

Mean responses from NASA-TLX worksheets on a scale of 1-100 are documented in Figure 11. There were no significant differences for any factor. Overall, no one technique was significantly preferred.

6.4   Subjective Feedback

Participants provided subjective comments on the techniques they had just used. Participants commented on the perceived slowness of ECP: *It felt slow because it felt like*

I had to do twice as many actions and it required *a lot of tapping*. Several participants noted the techniques' similarity to its desktop counterpart saying *It's similar to copy and paste*. In comparison to ESC one participant said *I preferred that I didn't have to switch between selection techniques, I was always using my eyes* referring to varying swipe events used in ESC.

Participants perceived EDD to offer more control, stating *I felt I had more control moving objects* and that the *continuous feel of contact with the object* was *something that the other techniques lacked*. The sense of control also affected perceived speed and accuracy, *it felt slow, but it was definitely much more accurate because I could see the object in place before dropping it*. Similarly to ECP one participant found EDD similar to current desktop techniques saying, *It's just like moving windows around in an operating system*.

ESC was found to be difficult for participants: [it was] much harder than other techniques and I didn't know where to look. One participant found during the push task that it was frustrating that I had to look down to find the object, just out of peripheral vision. Finally the variations of swipe to perform summoning and casting were found to be confusing with participants saying, I didn't really like ESC because it had the addition of swiping in either direction.

7   Discussion

7.1   Results

Overall ESC was found to be the fastest but least preferred technique. Participants disliked ESC for two main reasons: (1) Confusion, the touch command used changed between swipe down and swipe up, this lead to confusion about which to use for each task. (2) Coordination, participants stated that they found it difficult to coordinate their hands and eyes. This result highlights an issue where eye-based input needs to correlate more naturally with a users need to use their eyes, to observe other actions they perform. This issue is specific to the requirements of pushing with ESC. The user must swipe up on a tablet-located object viewed in peripheral vision while simultaneously being required to fixate on a large display. A possible solution for this in further work would remove the need for simultaneous initial selection and targeting, and instead allow these to be performed in sequence, i.e., hold finger on tablet object to select, then look at large display, and finally perform a swipe up to transfer the object.

Participants responded well to EDD. In comparison to ECP, participants felt that being able to see the object moving gave them a greater sense of control. Although the system used a gaze cursor to provide continuous feedback to the user, it is clear that in EDD, this feedback is more obvious and familiar to users thus provoking a positive response.

As demonstrated in the example scenarios we outlined in section 4, it is possible to incorporate additional semantics in to our techniques. These can improve usability in more complicated scenarios. Users reported ECP felt slow due to the amount of tap commands required. To improve perceived speed, the paste behaviour can be leveraged in this technique to duplicate a selected object, thus reducing the need for context switching and quicker perception of transfer.

Furthermore, issues outlined above with ESC can be resolved by introducing an 'implicit object identification' semantic. In this case, the most recently pulled object would be pushed automatically, thereby removing the need to redirect visual attention to the touch modality.

7.2   Feasibility and Limitations

Eye Pull, Eye Push is dependant on the deployment of eye tracking as a pervasive technology. To realise such a vision there are several requirements and limitations: (1) Embedded or head-worn eye-tracking: it is imperative that users are always visible to the system. Current technology supports both, remote eye-tracking, where systems are embedded or situated below displays, and head-worn eye-tracking where users wear a personal eye-tracker. These are currently in the form of goggles but envisioned to become as small as standard glasses. (2) Calibration: current head-worn and remote eye-trackers require calibration before use. Calibration takes time and must be per-formed pre-interaction. More modern systems only require calibration that lasts less that 30 seconds but issues can still arise when interacting with displays at varying distances, this is due to a lack of robust parallax compensation. (3) Connection: users require a method to pair with displays as they interact. Do users implicitly pair with each display they look at? Are user's eye-tracking data globally broadcast for use? Or would authentication be required? To create seamless interaction, there would need to be a balance between privacy and functionality so that users are not inhibited by re-peated authentication.

8   Conclusion

In this paper we presented a novel interaction concept Eye Pull, Eye Push, gaze-supported cross-device content transfer. In our design we considered transfer between public and personal devices and how gaze and touch can be combined to create inte-raction techniques for this task. We outlined the following techniques: Eye Cut & Paste, Eye Drag & Drop, and Eye Summon & Cast. We presented and discussed sev-eral usage scenarios for these techniques.

Our three techniques were evaluated in a user study. Users were able to complete the basic tasks of pull and push, and responded most positively to our Eye Drag & Drop technique. The results of our user study showed that Eye Summon & Cast out-performed Eye Cut & Paste and Eye Drag & Drop in terms of speed but was least preferred by users due to its hand-eye coordination requirements. Eye Cut & Paste and Eye Drag & Drop performed similarly in terms of speed although Eye Drag & Drop was preferred due the more apparent continuous visual feedback it provided. In our discussion we outlined how additional semantics can be applied to each technique to extend functionality in differing scenarios. Furthermore we discussed the feasibility and limitations of Eye Pull, Eye Push in the real world.

In future work we aim to explore this design space further, to gain a full understanding of factors within it and the implications they have on this style of interaction, i.e., users proximity to content, display sizes and varying content-types.

References

1. Aliakseyeu, D., Nacenta, M.A., Subramanian, S., Gutwin, C.: Bubble radar: efficient pen-based interaction. In: Proc. AVI 2006. ACM (2006)
2. Ballagas, R., Rohs, M., Sheridan, J., Borchers, J.: Byod: Bring your own device. In: Ubi-Comp 2004 Workshop on Ubiquitous Display Environments, Nottingham, UK (September 2004)
3. Baudisch, P., Cutrell, E., Robbins, D., Czerwinski, M., Tandler, P., Bederson, B., Zierlinger, A.: Drag-and-pop and drag-and-pick: Techniques for accessing remote screen content on touch- and pen-operated systems. Proceedings of Interact, 57–64 (2003)
4. Bieg, H.J., Chuang, L.L., Fleming, R.W., Reiterer, H., Bülthoff, H.H.: Eye and pointer coordination in search and selection tasks. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA 2010, pp. 89–92. ACM, New York (2010)
5. Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P.: Touch projector: mobile interaction through video. In: Proc. CHI 2010. ACM (2010)
6. Bragdon, A., DeLine, R., Hinckley, K., Morris, M.R.: Code space: touch + air gesture hybrid interactions for supporting developer meetings. In: Proc. ITS 2011 (2011)
7. Carter, S., Churchill, E., Denoue, L., Helfman, J., Nelson, L.: Digital graffiti: public annotation of multimedia content. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, pp. 1207–1210. ACM (2004)
8. Dickie, C., Hart, J., Vertegaal, R., Eiser, A.: Lookpoint: an evaluation of eye input for hands-free switching of input devices between multiple computers. In: Proc. OZCHI 2006, pp. 119–126. ACM (2006)
9. Doeweling, S., Glaubitt, U.: Drop-and-drag: easier drag & drop on large touchscreen displays. In: Proc. NordiCHI 2010, pp. 158–167. ACM (2010)
10. Drewes, H., Schmidt, A.: The magic touch: Combining magic-pointing with a touch-sensitive mouse. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 415–428. Springer, Heidelberg (2009)
11. Greenberg, S., Boyle, M., Laberge, J.: PDAs and shared public displays: Making personal information public, and public information personal. Pers. and Ubiq. Comp. 3, 54–64 (1999)
12. Jacob, R.J.K.: What you look at is what you get: eye movement-based interaction techniques. In: Proc. CHI 1990. ACM, New York (1990)
13. Kumar, M., Paepcke, A., Winograd, T.: Eyepoint: practical pointing and selection using gaze and keyboard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2007, pp. 421–430. ACM, New York (2007)
14. Lankford, C.: Effective eye-gaze input into windows. In: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, ETRA 2000, pp. 23–27. ACM, New York (2000)
15. Mardanbegi, D., Hansen, D.W., Pederson, T.: Eye-based head gestures. In: Proc. ETRA 2012. ACM (2012)
16. Mardanbegi, D., Hansen, D.W.: Mobile gaze-based screen interaction in 3D environments. In: Proceedings of the 1st Conference on Novel Gaze-Controlled Applications, NGCA 2011, pp. 2:1–2:4. ACM, New York (2011)
17. Myers, B.A.: Using handhelds and PCs together. Comm. ACM 44, 34–41 (2001)
18. Rekimoto, J.: Pick-and-Drop: A direct manipulation technique for multiple computer environments. In: Proc. UIST, pp. 31–39 (1997)

19. Schmidt, D., Seifert, J., Rukzio, E., Gellersen, H.: A cross-device interaction style for mobiles and surfaces. In: Proceedings of the Designing Interactive Systems Conference, DIS 2012, pp. 318–327. ACM, New York (2012)
20. Stellmach, S., Dachselt, R.: Look & touch: Gaze-supported target acquisition. In: Proc. CHI 2012. ACM (2012)
21. Stellmach, S., Dachselt, R.: Still Looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets. In: Proc. CHI 2013. ACM (2013)
22. Stellmach, S., Stober, S., Nürnberger, A., Dachselt, R.: Designing gaze-supported multi-modal interactions for the exploration of large image collections. In: Proceedings of the 1st Conference on Novel Gaze-Controlled Applications, NGCA 2011, pp. 1:1–1:8. ACM, New York
23. Tani, M., Yamaashi, K., Tanikoshi, K., Futakawa, M., Tanifuji, S.: Object-oriented video: Interaction with real-world objects through live video. In: CHI 1992, pp. 593–598 (1992)
24. Turner, J., Bulling, A., Gellersen, H.: Extending the visual field of a head-mounted eye tracker for pervasive eye-based interaction. In: Proceedings of the 2012 Symposium on Eye-Tracking Research and Applications, ETRA 2012. ACM Press (2012)
25. Ware, C., Mikaelian, H.: An evaluation of an eye tracker as a device for computer input. In: Proc. CHI 1987, pp. 183–188. ACM (1986)
26. Zhai, S., Morimoto, C., Ihde, S.: Manual and gaze input cascaded (magic) pointing. In: Proc. CHI 1999, pp. 246–253. ACM, New York (1999)