Eye Drop: An Interaction Concept for Gaze-Supported Point-to-Point Content Transfer

Jayson Turner
Lancaster University
School of Computing and
Communications
Lancaster, United Kingdom
j.turner@lancaster.ac.uk

Andreas Bulling
Max Planck Institute for
Informatics
andreas.bulling@acm.org

Jason Alexander
Lancaster University
School of Computing and
Communications
Lancaster, United Kingdom
j.alexander@lancaster.ac.uk

Hans Gellersen
Lancaster University
School of Computing and
Communications
Lancaster, United Kingdom
hwg@comp.lancs.ac.uk

ABSTRACT

The shared displays in our environment contain content that we desire. Furthermore, we often acquire content for a specific purpose, i.e., the acquisition of a phone number to place a call. We have developed a content transfer concept, Eye Drop. Eye Drop provides techniques that allow fluid content acquisition, transfer from shared displays, and local positioning on personal devices using gaze combined with manual input. The eyes naturally focus on content we desire. Our techniques use gaze to point remotely, removing the need for explicit pointing on the user's part. A manual trigger from a personal device confirms selection. Transfer is performed using gaze or manual input to smoothly transition content to a specific location on a personal device. This work demonstrates how techniques can be applied to acquire and apply actions to content through a natural sequence of interaction. We demonstrate a proof of concept prototype through five implemented application scenarios.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous

General Terms

Human Factors

1. INTRODUCTION

Despite the ubiquity of shared and personal displays, a significant gap still remains in the ability to seamlessly apply
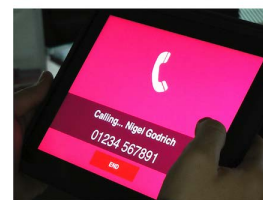
(a) Staff Directory with highlighted staff member using gaze

(b) Placing call   (c) Plotting route to room

Figure 1: Eye Drop - People Finding and Communication: A user selects contact information with gaze and touch to place a call or plot a route.

actions to content that is located remotely or out-of-reach. In this work we present Eye Drop, an interaction concept using gaze-supported techniques to transfer content. Eye Drop allows content to be used in a meaningful way as part of the interaction. In single-device interactions, content can be selected and immediately have actions applied to it. Throughout this interaction, all input is performed within the bounds of a single display. The applying of actions with content that is displayed remotely can often be tedious, i.e., copying an address to plot a route. Figure 1(a)-(c) demonstrates Eye Drop, it shows a user with a tablet device, standing in front of an information display. The user selects a contact number by fixating on it and holding their finger on the tablet device. The number is transferred to their touch location and dragged to the tablet's phone application icon. The user releases their finger, to drop the number and place a call. This example demonstrates Eye Drop's ability to hand-over con-
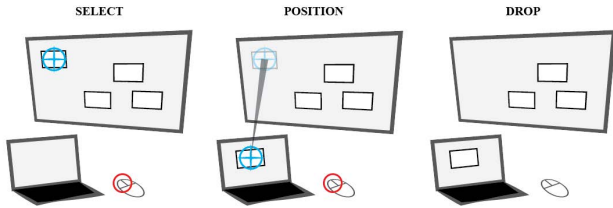
Figure 2: Gaze Positioning with Mouse: (Select) Look at content on shared display, confirm with mouse hold. (Position) Content is attached to gaze and can be positioned on the laptop display. (Drop) Mouse release drops the content.

Figure 3: Manual Positioning with 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.

tent to another modality for positioning, in this case touch.

Cross-device information transfer is well explored in human-computer interaction. Several techniques have been proposed including, the use of pen input acting as a token to transfer between displays [3]. Other methods are geared toward combining touch on a smartphone with pointing at a remote screen [1] or table top surface [4], facilitating data transfer to specific locations on the surface, and vice versa.

The eyes naturally look at what they desire, making pointing more implicit during remote selection. Additionally, the eyes can point when other modalities (i.e., the user's hands) are unavailable. Gaze offers the flexibility of interaction with devices over varying distances, from close-proximity to those that are out-of-reach, allowing selection of any content within sight. Early work in eye-based interaction identified the need for users to explicitly trigger actions when selecting content with gaze [2]. This has since been adopted to show that gaze can be combined with touch to support remote content acquisition [5]. In previous work, we investigated transfer between shared and personal displays using a combination of gaze and touch [6]. This work is distinct in that it supports transfer at a finer granularity, from point-to-point as opposed to display-to-display. Eye Drop provides techniques that allow gaze-acquired content to be transferred to a specific location and used to initiate further interaction.

This paper presents the Eye Drop concept, its requirements, and derived interaction techniques for gaze-supported point-to-point content transfer (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. Secondly we present an implementation and apply Eye Drop in two configurations, a tablet with touch input, and a laptop with mouse input. We then describe the ability of Eye Drop to acquire and use remote content in five application scenarios, *people finding and communication, meeting, lecture, photo sharing,* and *TV movie information.*

2. EYE DROP

Here we outline the requirements for Eye Drop alongside two techniques derived from its design.

2.1 Requirements

Eye Drop is accomplished by combining gaze and manual input, *Gaze* provides coordinates in display-space that indicate the location of a users visual attention and; *Manual*
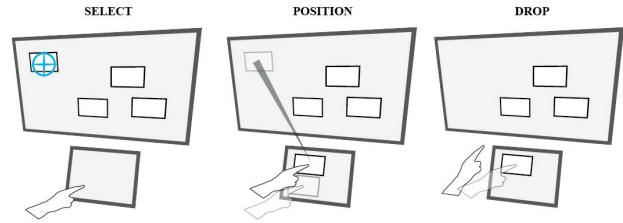
Input, provides explicit triggers operated by the user's hands to issue commands. These are elaborated below:

Gaze. Gaze input is required, as a minimum, to enable content selection on a public display. To utilise Eye Drop fully, gaze input should also be available on a users' personal device, to enable content positioning. To achieve this, the following setups can be used: a single head-worn eye-tracker, multiple remote eye-tracking systems attached to each display, or a combination of these. Remote eye-tracking requires displays to be augmented with this technology, whereas head-worn portable systems can, in principle, map users gaze to displays that are not "gaze-enabled" and simply in view of a user.

Manual Input. Manual input is required primarily to confirm actions and to position gaze-acquired content within a personal device display. The type of manual input used can vary depending on the available devices, this is demonstrated in our application scenarios.

Eye Drop encompasses three core interaction stages, *Select*: Content is targeted on a remote display and selected, *Position*: Content is transferred and positioned on a local device, *Drop*: Content is dropped at the specified local position.

2.2 Techniques

In the case of both techniques presented here, during selection, gaze is used to highlight an object, and a manual trigger confirms the selection. As outlined earlier, this approach has been utilised in several works previous [6, 5, 2]. Each technique is distinguished in its method of content transfer and final 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 object, 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. Figure 2 shows how gaze positioning is applied to a laptop with mouse input.
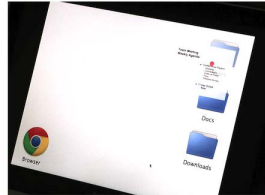
Manual Positioning. Here transferring content also constitutes transfer of modality, from gaze to manual input, with final positioning also performed by manual input. The user selects content using gaze and a manual trigger. Content is then instantaneously transferred to the close proximity device, held under the location defined by manual input, and positioned by the same means. A manual trigger then
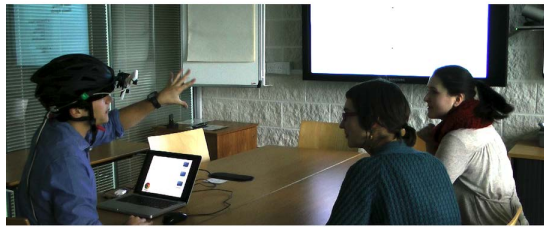
(a) Meeting with situated display and laptop

(b) Slide on display..

(c) ..transferred to the laptop

Figure 4: Meeting: A colleague acquires a slide from a presenters display to save to their laptop device.

(a) Lecture, with information on projecter, and student with laptop

(b) Information is transferred..

(c) ..to browser and searched for.

Figure 5: Lecture: A student selects specific text to quickly transfer and query on their laptop device.

drops the content. Figure 3 shows how manual positioning is implemented with a tablet device and touch input.

3. IMPLEMENTATION

We developed a prototype system implementing the Eye Drop concept. The system supports gaze input on both public and personal devices and consists of a customised head-worn portable eye tracker. Typical commercial eye trackers contain two cameras, one that observes the scene and another to track the orientation of the eye. These systems allow for a public display to be visible within its scene view but – due to viewing angle restrictions – it can be problematic to see close proximity displays. Similar to the system described in [7], our eye tracker overcomes this limitation by adding a second scene camera that is dedicated to the lower portion of a user's visual field.

We also developed a computer vision algorithm to detect both displays in the two scene cameras and map users' gaze to them automatically. To minimise error, both scene cameras were calibrated and undistorted to provide a rectilinear space. Public and personal displays are detected using brightness thresholding with contour detection. By minimising contours to four points, the rectangular surface of each display can be detected. Gaze is mapped from the scene camera to on-screen coordinates using a perspective transformation. To compensate for parallax error, and thus 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. The resulting system was accurate to within 1.5 degrees of visual angle.
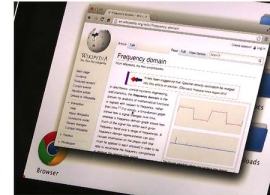
Our software consists four independent intercommunicating systems: (1) eye tracker, (2) application server, (3) public display GUI application, (4) personal GUI client application. Eye tracking data is transferred to the application server in the following format $(x, y, DisplayID)$, where x and y are gaze coordinates in display-space and $DisplayID$ denotes the detected display. The application server dis-

tributes this data to connected devices and applications, in this case a public display application and a personal device application. It also receives manual input events from the personal devices and delivers them to the public display application, these are in the form of *down*, *up* and *moved* events. Content (i.e. images and text) are serialised and transferred in the background over TCP upon selection. It is then up to the applications to determine how this transfer is visualised to the user. Our system is able to distinguish which display manual input is directed to based on which display has users' attention.

4. DEMONSTRATION APPLICATIONS

Here we describe five application scenarios, developed to demonstrate the versatility of the Eye Drop concept. We show how Eye Drop can be used in varying contexts, changing the personal device and manual input modality.

People Finding and Communication. A building foyer display contains information on its staff. (see Figure 1). A user wishes to contact and find a member of staff. The user looks at the desired staff information and selects it by touching and holding on their hand-held device. The selected information can be used in different ways: (1) Looking down at the tablet device, gaze can be direct to the "phone" application, then releasing touch, the information can be dropped and a call can automatically be placed. (2) Alternatively, the user can gaze on to the "maps" application, drop the contact information and plot a route through the building to the staff member's office. Eye Drop supports fine-grained positioning of gaze-acquired objects, allowing users to choose applications during transfer to utilise content.

Meeting. Three colleagues are discussing a slide on a large display (see Figure 4). One user is seated with a laptop and external mouse and wishes to obtain a copy of the slide. The user looks at the slide and holds down the mouse button to select it. Now, the selected slide follows the users gaze to their laptop screen and to a folder entitled "Slides". The user then lifts their finger from the mouse to drop and store the slide. The ability to perform fine-grained position-
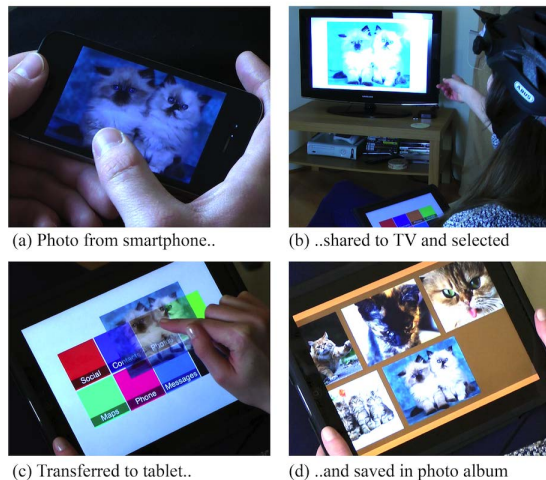
(a) Photo from smartphone..

(b) ..shared to TV and selected

(c) Transferred to tablet..

(d) ..and saved in photo album

Figure 6: Photo Sharing: A users selects a shared image and transfers it to the photo album on their tablet device.

(a) Eyes are watching a movie on TV...

(b) Touch to acquire viewed frame..

(c) ..display actors info on tablet

Figure 7: TV Movie Information: A user obtains a movie frame to view an actors biography.

ing on the laptop using gaze enables the user to store the acquired information in a specific location. Eye Drop allows for spontaneous interaction, users can gaze at content of interest and immediately obtain it, fluidly transferring between devices.

Lecture. During a lecture, attendees prefer not to disrupt the speaker to ask questions. A student sits with a laptop in a lecture theatre at University. The lecturer highlights terms shown on the projection (see Figure 5). To query the meaning of a term, the student fixates on it and holds down the mouse button on their laptop, instantly transferring the term to the laptop display. The user then drags the mouse and, in turn, the text to the browser icon to initiate a search for the acquired term. Here the user only needs to change their context once to acquire and use information as opposed to copying the text manually by typing.

Photo Sharing. At home, two users are browsing photographs on their TV (see Figure 6). Images are shared over Wifi from one user's smartphone. The second user is holding a tablet device. As the first user flicks through images, the second user spots an image that they would like to keep. While looking at the image, the user acquires it by touching down on their tablet device, transferring it to the location of the touch. Dragging the image to the "photos" application, the user releases their finger to drop the image. This opens the application and automatically adds the image to their album. In this scenario, information is transferred from one user to another using the TV as an intermediary. Although this is not part of the core interaction concept we present, it does touch upon how it can be extended.

TV Movie Information. A user is watching a film on TV, sitting on a sofa with a tablet in-hand (see Figure 7). To query information about a particular actor, the user acquires the current movie frame, while watching TV, by touching down on their tablet device. The transferred image contains the actors face, this is then dragged by touch to a "movie database" application, where the face is identified, loading a career history for perusal. This scenario demonstrates how data can be extracted from visual media that has users' attention. Transferring stills to touch allows specific informa-
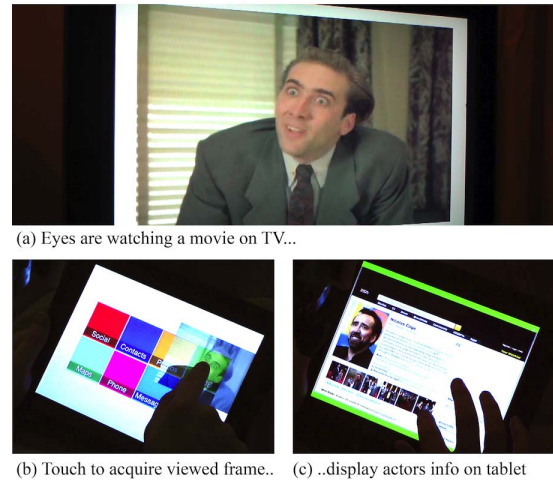
tion about media to be quickly obtained.

5. CONCLUSION

In this work we have presented Eye Drop, an interaction concept for content transfer between shared and personal displays. Eye Drop allows for content to be acquired remotely by gaze and positioned on personal devices, seamlessly enabling further interaction. We have shown how Eye Drop can be applied to touch tablet and laptop configurations, alongside five demonstrative applications for this concept.

6. REFERENCES

[1] A. Bragdon, R. DeLine, K. Hinckley, and M. R. Morris. Code space: touch + air gesture hybrid interactions for supporting developer meetings. In *Proc.*, ITS '11, 2011.

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

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

[4] D. Schmidt, J. Seifert, E. Rukzio, and H. Gellersen. A cross-device interaction style for mobiles and surfaces. In *Proc.*, DIS '12, 2012.

[5] S. Stellmach and R. Dachselt. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proc.*, CHI '13, 2013.

[6] J. Turner, J. Alexander, A. Bulling, D. Schmidt, and H. Gellersen. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Proc.* INTERACT '13, 2013.

[7] J. Turner, A. Bulling, and H. Gellersen. Extending the visual field of a head-mounted eye tracker for pervasive eye-based interaction. In *Proc.* ETRA '12, 2012.