

"This is the peer reviewed version of the following article: [Journal of the Society for Information Display, 2020, 28, (12)] which has been published in final form at [https://onlinelibrary.wiley.com/doi/10.1002/jsid.954] purposes in accordance with [Wiley Terms and Conditions for Self-Archiving](#)."

Gaze Window: A New Gaze Interface Showing Relevant Content Close to the Gaze Point

Seungwon Kim¹, Mark Billingham¹, Gun Lee¹, Weidong Huang²,

¹ University of South Australia
{Seungwon.Kim;Mark.Billinghurst;Gun.Lee}@unisa.edu.au

² University of Technology Sydney
Weidong.Huang@uts.edu.au

Abstract

In this paper, we propose a novel concept of the gaze window which follows the user's gaze to show relevant information close to their gaze point on a screen space. We explore the use of a gaze window, especially for two use cases: 1) a single user data entry task by displaying data-source image close to the gaze point where the data is typed in and 2) a teleconferencing setup showing a remote partner close to the gaze point. To investigate the performance of gaze window in these cases, we conducted two experiments. For the single user task, we compared the gaze window to a baseline non-gaze-window condition and found that the gaze window significantly reduced the user's gaze movement and mental effort and increased usability. In the teleconferencing task, we compared displaying partner close to gaze point in the gaze window to displaying partner at the corner of screen, when two remote users watch a shared and synchronized content together. We found that displaying the partner's face nearby the user's gaze helped the user look at the partner more frequently, and for a longer time, significantly increasing co-presence and emotional interdependence. As a result of this research, we also presented design guidelines for gaze-window systems.

Keywords: Gaze interface, Gaze window, Teleconferencing system, Data entry task

1. Introduction

With gaze tracing technologies getting increasingly advanced and affordable, gaze has been used for input and for interaction methods [4]. In this paper we present the Gaze Window which uses the user's gaze point to show the relevant content nearby the current gaze point. Using a traditional desktop interface, people often switch their gaze between two interesting objects or areas when they want to see both [19, 20]. For example, when comparing two on-screen images, when writing a document based on the information from another document, or when watching a shared content view and a video conferencing partner together in a video conferencing system. In these cases, users may put the two objects of interest close to each other to have a better look at both of them, especially by placing two relevant windows close to each other.

Placing windows close to each other requires additional inputs [10, 32], such as moving a window and reducing the size of the window, which may result in requiring more scrolling interaction (i.e. scrolling up/down). Based on this observation, we propose an idea of a gaze window for better looking at both by placing two interesting screen areas close to each other using the users' gaze point. Figure 1 shows an example of the gaze window in use when a user is typing numeric data in a spreadsheet. The user needs to keep switching their gaze between where the information is being acquired and where it is being typed into the spreadsheet. If the area of the data source is graphically copied and displayed near user's typing point, users will have less gaze movement in switching their gaze between acquiring and typing data (see Figure 1a). If a user changes his or her gaze point to another cell to type in the next

data (see Figure 1b), the system moves the copied data-source image using a gaze following function and the user can keep a close look of the data-source image (see Figure 1c).

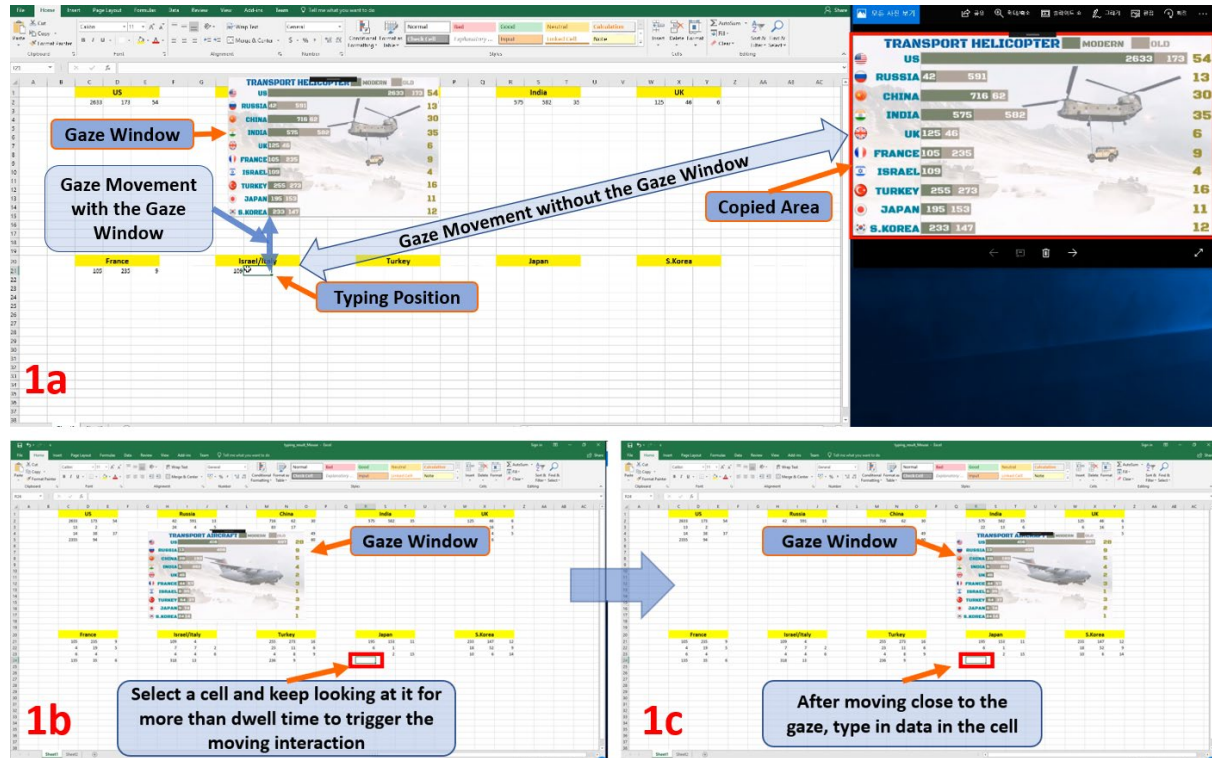


Figure 1. Screen shots showing the use of the gaze window interface. 1a) A user types into a spreadsheet base on the information from the right image. The red rectangle indicates the copied area of interest which is displayed on the gaze window positioned above the user's gaze point. 1b and 1c) A user changes his/her gaze point to the next cell and keeps looking at it to move the gaze window closer.

To investigate the benefit of the gaze window approach, we explore the use of the gaze window with a single user task (data entry) and a teleconferencing task. In the first study, we investigate how the gaze window helps a user complete the data entry task by comparing the gaze window condition (showing data image near the user's gaze point) to a baseline condition (without the gaze window). In the second study, we extend the use of the gaze window for a teleconferencing scenario where remote collaborators watch and discuss about a video. More specifically, we explore whether a gaze window that shows the remote collaborator close to the gaze point improves the feeling of co-presence and emotional interdependence when collaborators watch a video together by comparing the gaze window to the corner window that is statically placed at a corner of the display.

To the best of our knowledge, our work is the first to apply gaze interaction for a data entry task and showing a user's partner's headshot in teleconferencing system. Our research makes the following novel contributions:

- Proposing a novel interface concept of the gaze window that follows user's gaze to show relevant information close to the user's focus.
- Conducting and reporting on user studies comparing the use of the gaze window in data entry task and teleconferencing system.
- Conducting and reporting on a user study using the gaze window to display partner in both small and big displays.

This paper is an extension of its conference version [64]. In the remainder of the paper, we first review related work, then describe the gaze window interface. Then, we describe the user study design, report on and discuss the experimental results, and finally draw conclusions.

2. Related Work

In this section, we first review literature on the human visual system, user interfaces using gaze inputs, and teleconferencing system.

2.1 The Human Visual System

The sensory receptors in human eyes detect light in the visible spectrum, between violet and red [1]. The light enters through the cornea and is focused by the lens onto the retina [2]. The centre of the retina is called the fovea and it offers high visual acuity with ability to find details [3]. People mostly get visual information through the foveal vision, and it is limited to the central 2 degrees of visual angle, about the size of a thumbnail at arm's length [4]. Outside the foveal vision, visual acuity drops sharply to about 50 percent in the daytime [5].

Since the area of the foveal vision is limited, people use five eye movements to see: saccades, smooth pursuits, vergence, vestibular ocular reflex, and the optokinetic reflex [6, 7]. Between the eye movements, fixations occur to acquire information and people spend around 90% of their viewing time having fixations [8], where they maintain their gaze on a single location to hold a nearly still image on the retina. In short, people move eyes to coordinate or position them to acquire vision information. The gaze window interface is to reduce the time and effort for coordinating or positioning eyes by reducing the distance between two areas or objects of user interest in the screen area, so encouraging users to spend more time and effort for acquiring vision information.

Based on this understanding of the human visual system, we designed a gaze window interface which matched the performance of the human eye (more details in Section 3). The next section will describe previous studies of gaze interfaces.

2.2 User Interfaces using Eye Gaze

As real-time eye-tracking has become feasible, eye gaze has started to be used as a real-time input method [9, 10, 11, 63]. In this section, we discuss previous studies exploring the real-time gaze inputs.

2.2.1 Single User Interface

Generally, for human-computer interaction, gaze has the following benefits: 1) it is much faster at pointing compared to mouse input [12], 2) it can be used without using hands for impaired users [13] or when the user's hands are busy [4], and 3) it can be more natural if the human gaze activity is well integrated into the computer interaction activity [14]. On the other hand, gaze interaction may not be as accurate as mouse input [12], can be easily distracted by object movement in the peripheral vision, and sometimes have issues in input control with unconscious and involuntarily gaze moves.

In designing a gaze interface, Jacob [10] reported on the Midas Touch problem. Usually, people scan a scene for perception rather than for interaction purposes. If interactions are made with every eye fixation on the objects in the scene, a user cannot look anywhere without the triggering the interactions. To eliminate this Midas Touch problem, the system should distinguish between observing gaze and interaction gaze behaviour. To do this, Jacob [10] used the dwell time interaction which distinguishes the interaction gaze from observing gaze by the time the user gazes on an object or area. If the gaze dwelling time is greater than a certain amount, an interaction is triggered and the gaze is regarded as an interaction gaze. If not, the system considers the gaze as an observing gaze and the interaction is not triggered. In a user study, he found that button selection with a dwell time interaction was more convenient and faster compared to keypad input. A typical dwell time was between 600 and 1000 milliseconds [15]. In our study, we also used dwell time to distinguish between observing and interaction gazes.

Jacob [10] also conducted two more user studies with eye gaze input for moving objects and issuing menu commands. With moving objects, he compared mouse and gaze input for dragging, finding that gaze dragging was more convenient and faster for selecting the moving object. With the menu command task, a dwell time interaction was used with a pull-down menu and selecting an item from the menu (using a 1 second fixation). In a user study, dwell time interaction took longer compared to mouse button press interaction. These results show that gaze input would be better for a positioning task (such as dragging an object) compared to mouse interaction, but not for a task including a large amount of clicking (or selecting).

One of the interesting issues with dwell time interaction is the trade-off between dwell time and reliability [16]. If the dwell time is short, users might complete the task faster but there is a higher possibility of inadvertent commands. If the dwell time is longer, the possibility of error would be reduced but the task completion would be slower. Additionally, when measuring dwell time, the interface requires an area rather than a point because it is difficult to keep looking at the exactly same point (or pixels in a screen) for dwell time interaction even if a user keep looking at an object.

Some other researchers used gaze for controlling the view when looking at the edges of the screen or virtual arrow buttons [14, 17, 18]. This is similar to selection interaction with gaze because a user needs to position his/her gaze on an area or button, and the dwell time on the area or arrow button determines the panning amount. One of the issues of these systems was that a user cannot see other content in the view while looking at the edge of it or arrow buttons to trigger the panning interaction.

Ekman et al. [19] implemented a following gaze navigation. They used continuous gaze interaction to position an avatar at the gaze point in the task, continuously moving the avatar to escape from a monster. The interaction mapping was one to one between the avatar and gaze input and their study showed the possibility of using gaze to control character navigation. Similarly, Dorr et al. [20] introduced a continuous and one-to-one gaze mapping interface for the Breakout computer game and compared gaze and mouse-controlled interfaces. With the gaze interface, a user could continuously reposition the paddle while watching and following the moving ball. In a user study, they found that participants played better with the gaze control than with the mouse control. Vicker's study [21] also showed the benefits of gaze interaction compared to keyboard interaction in playing Frets on Fire. The task required continuously pressing one of the keys to play a song and their gaze interface supported continuous key input. Their interaction mapping was also one to one between the virtual piano keyboard and gaze input.

One interesting point in Dorr's and Vicker's gaze interfaces [20, 21] was that their system limited the area where the gaze inputs could be applied to. In Dorr's Breakout system, a user can look around and keep watching a ball over the entire screen, but the gaze inputs only apply to the horizontal movement of the paddle and the paddle is always at the bottom of the screen. Similarly, in Vicker's Frets on Fire system [21], a user can look around entire scene, but the gaze inputs only apply on a key on the keyboard displayed on the bottom of the screen. In our study for the data entry task, we also limit the vertical movement of the gaze window and use two horizontal alignments.

Since the gaze input cannot support diverse interaction like gesture input [15], other researchers used gaze together with voice, hand, mouse clicks, dwell time, or touch screen interaction. For example, Wilcox et al. [22] introduced a system using both gaze and voice: selecting an object by using dwell time interaction and performing appropriate action by using voice inputs such as 'pickup', 'walk', and 'use'. Yamamoto et al. [23] developed a computer game of "Hyakunin-Isshu" where a virtual character played a counterpart of a human player by reacting to the player's gaze when the player used hand gestures to select a card.

Pfeuffer et al. [24, 25] introduced a system combining gaze with touch screen input. In a normal touch-screen interaction, a user interacts based on finger touch point(s) (e.g. selecting an item by finger touch), but their system used a gaze point as the interaction point and a user could interact with the screen by touching anywhere on the screen. With the gaze and touch interaction, users could place their hands in the comfortable area while conducting interactions and also could select multiple objects which are far away each other without suffering from serious hand movement to directly point (or select) on the objects. While their studies mostly focused on the benefit of easier reachability of gaze on the touch

screen interface rather than direct hand touch, our study focuses on displaying an area or object of user interest close to the other of user interest.

Serim and Jacucci [26] explored the limitation of visually focusable area and developed a touch screen interface that shows selected relevant information close to the gaze point. This reduce the gaze and head movement for looking at two areas or objects of user interest and similar with our proposed gaze window interface, but their use cases were limited in touch screen interface as choosing one or two of user interests and(/or) accompanying interaction is done by touch screen interaction. In our single-user study, the target use case includes massive keyboard interaction such as typing task.

Fabian et al. [27] developed a system for information visualization within a thematic map where a user needed to extract relevant information from the legend. In their system, only the symbols in the recently focused (gaze fixation) area were fully rendered in the legend, so a user could easily find information of the symbols in the legend. In their studies, the gaze was mostly for providing position (or point) information to the interaction triggered and controlled by other input method.

2.2.2 Gaze Interface in the Teleconferencing System

In old studies, researchers simply displayed partner(s) and let user predict where partner is looking at. Ishii et al. [28] developed the Clearboard system which shared the screen space with a remote partner and displayed the partner live as a background video. They found that showing the partner's eyes helped the user to have better collaboration because it helped them to know where their partner was looking and reduced the effort to maintain the collaboration. Later, Billinghamurst and Kato [40] displayed remote partners without background and attached them to tracking markers, so a user can arrange the position of remote partner by positioning the markers and allow the users to predict where the partners are looking.

Instead of sharing their partners in a live video view and letting a user to predict where the partner is looking [65], several researchers directly displayed the remote partner's gaze on the synchronized content view and found that it increased co-presence and awareness of where others were looking. For example, Vertegaal overlaid the users' eye gaze in a 3D virtual meeting room including a live video of users [29, 30], and the gaze information helped to show who was talking to whom. Carletta et al. [31], Neider et al. [32], and Brennan et al. [33] also explored the effect of sharing gaze in a virtual environment and found that collaboration between users was faster with gaze cues. Stein et al. [34] worked on desktop setting and found that participants understood their partner's messages more clearly when sharing gaze cues.

Lately, researchers have started to use wearable eye trackers attached to a head worn camera [35, 36] or head mounted displays [36, 37, 38]. The wearable eye trackers increase the portability of the teleconferencing system and enable gaze sharing for remote collaboration on real-world physical tasks. In their studies, they found that users could complete a collaboration task quicker with gaze information and more clearly understand their partner's communication.

In contrast to these studies, our teleconferencing system is not sharing gaze point to remote partner(s) but use one's gaze point to display partner close to the gaze point for improving the users' feeling of co-presence and emotional interdependence. We especially focus on the teleconferencing system sharing a content view together with users' headshot.

In sharing both the content view and the partner's headshot in a teleconferencing system, many previous researchers focused on better system-configurations, such as display size [39], view perspective [40, 41], and portability [42]. These configurations were mostly for better sharing the content view to provide a similar level of user experience for the remote partner. However, only focusing on better sharing the content view could be limited in sharing emotion between users because the shared face is a one channel for sharing a user's emotion. Darwin [43] argued that our faces are the key to sharing emotion. Ekman and Friesen [44] followed Darwin's opinion and found that people could reliably recognize the six emotions (six basic emotions from face expressions: anger, fear, disgust, surprise sadness, and happiness) from the faces even if they came from very different cultures [45]. In our

teleconferencing study, we use the self-gaze point for positioning a gaze window which shows the partner's face near the user's gaze point for sharing emotion and feeling of co-presence.

3. Design the Gaze Window

The gaze window is a viewport in a graphical user interface which shows the user's selected area of interest. It follows the user's gaze to be near their gaze point where another area of user interest is displayed. Before developing the prototype, we discussed how to design the gaze window interface and finalized the seven design requirements as listed below:

1) *Pre-define gaze window content*: The content of the gaze window should be defined previously. This could be either defined manually by the user or automatically estimated by the system. In our system, the user needs to manually predefine the object or area of interest that will be displayed in the gaze window.

2) *Make gaze window to follow gaze points*: The gaze window should follow user gaze inputs. By following the user's gaze, the gaze window can be displayed close to the user's gaze point while the gaze window shows the predefined user interest. We used gaze fixation with dwell time to set the point that the gaze window moves to. By using dwell time interaction, the gaze window can be used regardless of the availability of other input methods and when other input devices are heavily used for other purpose (i.e. typing).

3) *Distinguish gaze interaction and gaze pause*: When using the dwell time method, the system should distinguish gaze inputs for moving the gaze window from gaze inputs for looking around. Since the looking around gaze mostly include very short gaze fixations [10, 46], a 600ms dwell time was chosen based on the comment by Majaranta [15]; "A typical dwell time is between 600 and 1000 milliseconds". With 600ms dwell time, the looking-around gaze can be ignored, reduce the possibility of malfunction of the gaze window due to the "Midas Touch" issue [10] (i.e. the gaze window always moving around even with the short gaze fixation for user looking around activity).

4) *Define area for gaze dwell*: The gaze window system should define an area to measure the time of gaze dwell for the dwell time interaction. We used a circle area to measure the dwell time with a radius of 3 degrees of visual angle from the centre of the gaze. We chose the 3 degrees value from the fact that foveal vision is less than 3 degrees and people could keep seeing foveal vision at a specific point when focusing on it [4]. The dwell time area should be dynamically generated to allow the following gaze function to be available in any area of the screen. When the system starts, the first gaze point is the centre of the dwell time area. If the dwell time is over 600ms in the area, the system triggers the gaze window movement. If the dwell time is less than 600ms, the system generates a new dwell time area with the first gaze fixation outside of the previous area. If an area is used for moving the gaze window (after triggering dwell time interaction), any gaze in the area does not trigger another movement of the gaze window even though it is longer than the dwell time (600ms) because the gaze window is already nearby the gaze point.

5) *Display gaze window close to but not cover the gaze point*: The gaze window should be displayed close to the user's gaze point but not covering it. Since the purpose of the gaze window is to show two areas of user interest close to each other (one of it is displayed in the gaze window and the other one is the one where a user looks at with gaze fixation in the screen), it should not occlude each other but should remain close enough to be glanced at when needed. The distance between the gaze point and the gaze window should be outside of the foveal view (around 2 degrees visual angle [4]) to eliminate possible distracting views. In our prototypes, the gaze window is at a distance of 3 degrees of visual angle from the gaze point that triggered the previous gaze window movement.

6) *Allow gaze fixations within gaze window*: The gaze fixations in the gaze window should not cause gaze window movement. This allows the user to look at the gaze window content. This can be achieved by not counting the dwell time in the gaze window.

7) *Determine the gaze window size according to the information amount displayed*: The gaze window should be big enough to show the displayed information, but also small to reduce the occluded area of

the screen. We give priority on showing information, so the size of the gaze window would be smallest one while supporting good vision of reading information. With this requirement, we note that the size of the gaze window could be dynamic according to the user's eyesight and type of information the gaze window contains (i.e. numerical text or partner's face).

Using these requirements, we developed two gaze window prototype systems: one for single user task and the other for teleconferencing collaboration tasks and conducted user studies to investigate how gaze window help users to manage tasks. We describe the details of our system development and user studies in the next sections.

4. Gaze Window for Single User Tasks

To investigate whether the gaze window helps single users to manage looking at multiple objects of interest, we implemented a gaze window system based on the seven requirements from Section 3 and conducted an experiment with a data entry task. In this task, the gaze window system generally operates in three steps (see Figure 1b and 1c): 1) the user selects a cell in a spreadsheet, 2) the user keeps looking at the selected cell to move the gaze window closer to it (using dwell time interaction), and 3) the user types in data. This requires the user to change focus between the image and spreadsheet. This experiment compared the gaze window condition (showing data image near the user's gaze point) to a baseline condition (without the gaze window).

4.1 The Single User System

The prototype system was implemented on a desktop computer with an Intel Core i7-7700K 4.2GHz quad core CPU, 16 GB RAM, and NVIDIA GeForce GTX 1070 graphics card driving two 21-inch monitors. To track the user's gaze and implement the following gaze interaction (design requirement 2), we used the Tobii EyeX¹, a desktop eye tracker and its SDK. The Tobii EyeX provides eye-tracking with about 0.5 to 1 degree accuracy, operates at a minimum 60 Hz update rate, covers up to a 27-inch screen, and supports easy gaze calibration. We used its SDK and customized the sample code [47] to capture gaze data in the screen space and implemented a function that copies a screen area of user selection with a red outlined box and displays it on a gaze window positioned near to the user's gaze point (see Figure 1a). The user can select the position and size of the copied area with the red outlined box (design requirement 1). To resize the copied area, the user can position the mouse cursor on the edges of the red box and drag while pressing the mouse left button. To change the position, the user can place the mouse cursor inside the red box and drag it to a new position while pressing the mouse left button.

The initial size of the gaze window is defined according to the screen size; its width was the screen width divided by 3.5 and the height was the screen height divided by 3. We designed the size of the gaze window with these simple formulas to make it easy to read text in the gaze window (design requirement 7). The initial size of the red box outline was the same as the gaze window, and its initial position was at the top left corner of the right monitor screen. The distance between the gaze point and gaze window was around 12% of the screen height (design requirement 5; this was around 3 degrees visual angle from the centre of fixation if a user looked at the display from 70 cm away). However, a user could change the distance and relative position between them by dragging the gaze window with a mouse. The last gaze point before dragging the gaze window (mostly inside the gaze window) and final gaze window position were used to define the distance and the relative position. To measure the time for the dwell time interaction (design requirement 4), we used a circle area with radius at 3 degrees of visual angle from the centre of fixation. We note that the circle of dwell time area dynamically changed its position (design requirement 4).

¹ <https://tobiigaming.com/product/tobii-eyex/>

To meet the design requirements 3 and 6, the system changed the position of the gaze window only when the user was not looking at the gaze window and the fixation dwell time was 600ms or more. The data entry task with a spreadsheet included a search activity to find a cell to type into. Since our study was not about searching but looking at multiple objects, we designed the task should have only simple searching by making the next cell to type to the right side of the previous cell or in the next line. With this left to right data typing task, we positioned the gaze window above or below the gaze point. The gaze window was placed above the gaze point if the gaze point was in the bottom half of the screen, and it was placed below the gaze point if the other way around. Moreover, in addition to positioning the gaze window above or below the gaze point, we horizontally align the gaze window by following the previous studies by Dorr's [20] and Vicker's [21].

4.2 User Study

To investigate the effectiveness of gaze window on single user tasks, we conducted an experiment to compare gaze window condition to non-gaze window baseline condition for data entry tasks. User answers from questionnaires, task completion time, and log data of user activities were collected for analysis.

4.2.1 Design

The within-subject design was employed by the study. This means that each participant performed the tasks in all the experimental conditions. For this study, there were two conditions for this experiment: (1) baseline and (2) the gaze window conditions. In the rest of this section, we use the abbreviations GW and BL to represent the gaze window and baseline conditions, respectively.

4.2.2 Participants

We recruited from the student population 12 participants (8 males and 4 females) aged 20 to 35 years old ($Mean = 28.3$; $Std.Dev. = 6.7$). All of them had heard about eye tracking systems, with five of them having prior experience of using eye trackers. None of them had prior experience of using any gaze window systems at the time of the experiment.

4.2.3 Procedure and the tasks

The experiment started with a participant signing an informed consent form, answering a demographic questionnaire and the experimenter describing the purpose of the study. The participant sat in front of a desktop and used a mouse and a keyboard to complete data entry tasks where they had to read data from an infographic image and enter the data into a spreadsheet displayed on the left screen with or without the gaze window (see Figure 2). The participants repeatedly selected cells, moved the gaze window, and typed in data.

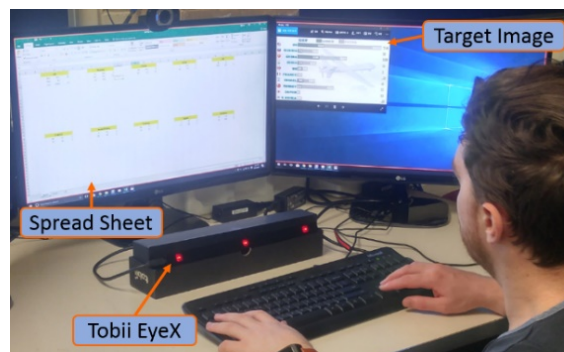


Figure 2. A user performs a task in baseline condition (without the gaze window). The user reads numerical data from the target image on the right screen and types it into the spreadsheet on the left screen.

More specifically in the experiment, participants went through training and experimental sessions for each condition. There were two training tasks including 62 and 63 numeric key inputs and 48 and 49 cell selections in the spreadsheet. Each training task had two data source images (participants could switch to the next image by clicking on the arrow button under the source image). The experimental tasks included 215 and 216 key inputs and 111 and 109 cell selections by using arrow keys on the keyboard. Each experimental task had four source images.

In the spreadsheet where the participants had to fill in the data, there were ten sections with names of ten countries, and participants were asked to type the number data under the relevant country. Since each source image included one type of information for the ten countries, participants typed in line by line. This data entry task included active keyboard typing, reading numbers, remembering the numbers (after reading a number until typing it in), and dozens of gaze-point changes. The source images were initially at the top left corner of the right screen (see Figure 4), but participants were allowed to change their position and size, as well as the spreadsheet window. The order of the conditions was counterbalanced between participants.

After completing each given condition, participants answered questionnaires asking for their impressions of the interface. These included the System Usability Scale (SUS) [48] and the Subjective Mental Effort Questionnaire (SMEQ) [49, 50], and two open-ended questions asking what they liked or disliked about the condition. Once the participant finished both conditions, the experiment was wrapped up with a final questionnaire asking about the participant's preference between the conditions and the reasons for their choice. Overall, the experiment took about 30 minutes for each participant.

In addition to the questionnaire, we collected objective measures such as the amount of gaze movement and task completion time. Note that we did not collect the gaze movement on the right screen due to a limitation of the Tobii EyeX, so the total amount of gaze movement in the baseline condition would be greater than the amount we recorded, as the gaze movement on the right screen was not counted.

4.2.4 Hypotheses

Based on the discussion in the Related Work and system development sections, we had the following hypotheses:

- (H1) The gaze window reduces the amount of gaze movement compared to the baseline condition.
- (H2) The gaze window has higher usability and induces less mental stress compared to the baseline condition
- (H3) The gaze window helps users to complete the data entry task more quickly compared to the baseline condition

4.3 Results

Since the data of the task completion time and the amount of user gaze movement were normally distributed according to Shapiro-Wilk tests (completion time: GW - $W(12)=.938$, $p=.468$, BL - $W(12)=.892$, $p=.126$; gaze movement: GW - $W(12)=.936$, $p=.448$, BL - $W(12)=.958$, $p=.758$), we analysed them using a paired t -test ($\alpha = .05$). We used the Wilcoxon Signed Rank test ($\alpha = .05$) to analyze the data from the questionnaires. Table 1 shows the results of the task completion time, the amount of gaze movement on the screen, and the questionnaire score: SUS and SMEQ.

Table 1. Results of the task completion time, amount of gaze movement on the screen, SUS, and SMEQ. The significant results are shaded in grey.

	<i>Mean (Std.Dev.)</i>		Method	Result
	BL	GW		
Completion Time	303 (76.4)	326 (74.5)	Paired t -test	$t(11)=-1.824$, $p=.095$

(in seconds)				
Gaze Movement (in pixels)	137814 (11908)	107670 (9025)	Paired <i>t</i> -test	$t(11) = 8.619, p < .001$
SUS	63.5 (11.5) - Okay	73.8 (14.7) - Good	Wilcoxon	$Z = -2.586, p = .01$
SMEQ	22.91 (20.61) - a bit hard to do	14.33 (15.77) - not very hard to do	Wilcoxon	$Z = -2.160, p = .031$

Interestingly, the gaze window did not produce a benefit for user task completion time (see Table 1). There was no significant difference in task completion time between the GW and BL conditions in a paired *t*-test (see Table 1, $t(11) = -1.824, p = .095$). On average, participants completed the task in 303 seconds ($SD = 76.43$) with the BL condition, and in 326 seconds ($SD = 74.52$) with the GW condition. They made very few typing mistakes, seven with the BL and five with the GW.

While the gaze window had no significant effect on task completion time, it did reduce the participants' gaze movement (see Table 1). During the experiment, the system collected the total amount of gaze movement in the screen space (in pixels with the resolution of 1920 by 1080). The participants' gaze moved 107670 ($SD: 9025$) pixels on average in the GW condition and 137814 ($SD: 11908$) pixels in the BL condition. A paired *t*-test showed that participants had significantly less gaze movement in the GW condition than in the BL condition ($t(11) = 8.619, p < .001$).

The questionnaire results also showed the benefit of the GW condition (see Table 1). From the SUS rating, the average score of the GW condition was 73.8 ($SD: 14.7$), which is in the range of 'GOOD' in the adjective rating by Bangor [51]. The BL condition had an average of 63.5 ($SD: 11.5$) which is in the range of 'OKAY'. In comparison (Wilcoxon signed ranks test), participants felt the GW condition was significantly more usable than the BL condition ($Z = -2.586, p = .01$). Moreover, the results of the SMEQ showed that participants felt that they needed significantly less mental effort for the data entry task with the GW condition than with the BL condition ($Z = -2.160, p = .031$). The average score with the GW condition was 14.3 ($SD: 15.7$) which is referred as 'not very hard to do' [49], and with the BL it was 22.9 ($SD: 20.6$), 'a bit hard to do' [49].

Confirming these results, participants gave relevant answers to the open-ended question asking about what they liked in the GW condition. Among the twelve participants, eight of them mentioned it was easy, convenient, or comfortable to use (e.g. "*It made it easier to look at the data, also made it easier to spot mistakes*"), and three reported they needed less physical effort to turn their head (e.g. "*It took less effort to physically turn my head consistently to copy the numbers shown on a different screen*"). In response to the question asking what they disliked about GW condition, five participants did not mention anything, four of them pointed out the late response (maybe from the 600ms dwell time). Overall, all participants preferred the GW condition over the BL condition for the data entry task.

4.4 Discussion and Limitation

Based on the experimental results, the gaze window had a higher level of usability and was less mentally stressful with less gaze movement compared to the baseline condition. Thus, **H1** (The gaze window reduces the amount of gaze movement compared to the baseline condition) and **H2** (The gaze window has higher usability and induces less mental stress compared to the baseline condition) were supported in the first study. The less gaze movement would influence participants' feeling in a higher level of usability. Additionally, we argue that the required time remembering numbers until typing them in would be less with the gaze window than with baseline condition because the required gaze movement was less. This less-time remembering numbers would influence less mental effort.

However, the benefits of the gaze window did not significantly affect the task completion time. Thus, the hypothesis **H3** (The gaze window helps users to complete the data entry task more quickly compared to the baseline condition) was not validated. There are three possibilities to explain in this regard. First,

we postulate that the gaze movement was fast enough [36, 37] so that it did not significantly affect the task completion time even though there was a significant difference in the amount of gaze movement between two conditions. Second, the gaze window movement included many dwell time interactions and the dwell time caused slower performance in the gaze window condition. Third, the participants managed to maintain the same level of task performance at the cost of gaze movement and mental stress in the baseline condition.

According to our observation, participants mainly looked at three areas during the experiment: 1) the target cell to type in numbers into a spreadsheet, 2) the data-source image, and 3) the keyboard. If a participant was good at typing numbers, he/she would hardly look at the keyboard and looked at the target cell less frequently to check whether their typing was correct or not. Instead, he/she would mostly look at the data-source image to read information from it. On the other hand, if a participant was not good at typing numbers, he/she would look at the keyboard more frequently.

Despite the positive effects of gaze window, we expect that there are situations where the gaze window could be relatively less or more beneficial. First, the benefit of using the gaze window would be marginal for data entry tasks with a small amount of data to type in as the gaze window requires initial set-up time. Second, if the copy and paste interactions are available in data entry and the task does not require too many of them, then the gaze window is not needed. On the other hand, there are certain cases where the gaze window can be useful for data entry. First, where the data source does not allow copy and paste interactions (e.g. images, as in our study). Second, even when the copy and paste interactions are available, if a lot of repetitive copy and paste are required, then simply typing in data with the gaze window would be useful. For example, typing on a keyboard would be faster if a lot of short words or numbers are needed to be copied and pasted one by one.

In the design and implementation of the system, we restricted the movement of the gaze window, which could be an issue for generalisability. For example, we placed the gaze window in the top half of a screen if the user was looking at the bottom half, and in bottom half if the other way around. However, sometimes the user may need to vertically scan through the text. However, this could be solved by allowing users to manually adjust the relative position between the gaze window and the user's gaze. For example, a user keeps the gaze window right or left, if the content requires him/her to vertically scan it.

Additionally, our system did not collect the gaze movement on the right screen due to the limitation of the Tobii EyeX. This might influence the amount of the gaze movement in the baseline condition but not in the gaze window condition because participants looked at the target image displayed on the right screen with the baseline condition but not with the gaze window. Therefore, the total amount of gaze movement in the baseline condition would be greater than the amount we recorded. Given this, our experiment still recorded significantly more gaze movements with the baseline condition than with the gaze window condition. Therefore, despite the limitation, we can still report that the gaze window condition reduces the amount of the gaze movement compared to the baseline condition, and that is our main result regarding the gaze movement.

5. Gaze Window for Teleconferencing

For better understanding of the gaze window interface, we explore the use of it in another user case: a teleconferencing setup especially when two remote users watch and talk about a shared and synchronized content together. We implemented a gaze window system. The system allows two users to share a synchronized content view (i.e. 360 seaside scenery video) while displaying the remote partner's headshot in the gaze window. We also conducted a user study to validate its usability and possible benefits.

5.1 The Teleconferencing System

The system prototype included two display systems, small and big, for two users, and the two displays were connected to each other through a wifi network. The system also captured the users' headshots to share in real-time and overlaid each of them on the partner's display. Since the main content in the gaze window was predefined as the partner's head shot, the participants in this study did not need to define the content of the gaze window (design requirement 1).

While developing the system, we also tried to meet the design requirements described in Section 3. For the small display system, we adopted the same desktop configuration as in the first study and used the Tobii EyeX (see Figure 3b) and the Tobii EyeX SDK for implementing the following gaze function (design requirement 2). However, instead of using two screens, the video and head shot were displayed in one screen (21 inches). The system obtained eye gaze data with Tobii EyeX SDK and customized sample code [47]. We kept the function of not moving the gaze window when the user's gaze dwelled on it (design requirement 6), and there was 600ms of dwell time to trigger the movement (design requirement 3). Since the task did not require any direction in performing the next step (unlike typing numbers from left to right in the single user task), the gaze window was always displayed on the right of the gaze.

The simplest way of capturing a headshot is by taking a live video of a partner and displaying it on the context view. However, the partner's headshot in an inset video looks like it is disconnected and unrelated to the context view with its background. To solve this issue, we followed Billingham's approach [40] that removes background of the head shot. We used a Kinect depth sensor and code from [52] to remove the background of the head shot (see Figure 3). Additionally, to remove the effect of size difference of the displayed face in each condition, our system keeps the size the same, regardless of how it was captured. The Kinect tracks the head and shoulder positions with the code from [53] and crops the user's face with the tracking information. For the design requirements 4 and 5, we followed the system set-up from the single-user study so the distance between the gaze point and gaze window was 12% of screen height (design requirement 5). These around 3 degrees of visual angle from the centre of fixation if a user looked at the display from 70 cm away for the small display and from 230cm away for the big display.

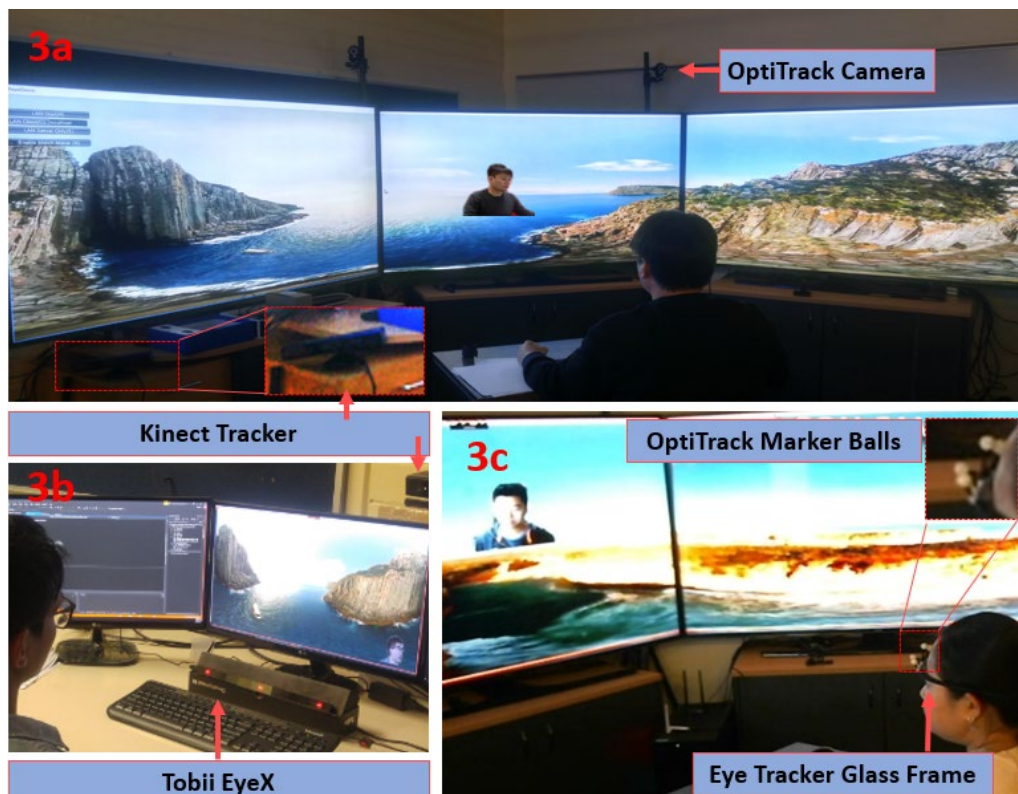


Figure 3. Teleconferencing system setup. The big display system included three large screens and calculated the user's gaze point with an OptiTrack and a Pupil Labs eye tracker. The small display system used one desktop screen and a Tobii EyeX for user gaze tracking. Both the big and small display systems extracted the users' head shots with Kinects and displayed them in the gaze window or the corner window on the other end while both users watched a shared synchronized video.

For the big display system, we used three large screens (75-inch wide each), a Kinect v2, a Pupil Labs wearable eye tracker, and an OptiTrack² motion capture system (see Figures 3a and 3c). Since there was no eye tracker supporting the large screens, we developed our own method for the gaze-following function. We created three virtual panels representing the three big screens with the previously saved position data of each screen corner measured with the OptiTrack system. The OptiTrack system tracked the user's head rotation and position data with the OptiTrack marker balls attached to a glasses frame which a user should wear (see Figure 3c).

With the head rotation and position, the system found a point on the large screen where the user's head was facing, using a ray casting method. The point that the user head was facing at was the centre point and the system calculated the gaze point on the large screens by adding the eye tracker gaze data (relative to the centre point in the wearable eye tracker coordinate) to the head facing point. This is possible because the added gaze data by the eye tracker (a Pupil Labs wearable eye tracker) was found in the large screen coordinate frame by calibrating the eye tracker with the big screen. For the live gaze data, we modified the sample code [54]. We note that the system did not share or transfer the gaze data to the other end, but used it for positioning the gaze window showing the extracted partner's head shot.

The software application (developed using the Unity game engine) played a shared video which was synchronized on both the small and big displays. The Kinect v2 extracted the user's head shot in the same way used for the small display system and sent it to the other end. This was done by using the sample codes from Background-remove example [55] and Body Basic example from kinect sdk V.2.0 development tool kit [56]. The gaze window was displayed on the right side of the gaze point (see Figure 3a and 3c), while the corner window stayed at the bottom right corner of the display (see Figure 3b).

We note that our system did not transfer the video but simply synchronized it by playing at the same time with signalling to reduce the network load. Moreover, the Kinect extracting the head shot was directly connected to the PC on the other end, so our system did not transfer the head shots through the wifi network as well.

5.2 User Study

To investigate the performance of gaze window on teleconferencing tasks, we conducted an experiment to compare gaze window condition to non-gaze-window condition for a video watching and talking task. User answers from questionnaires and log data of user activities were collected for analysis.

5.2.1 Design

In this user study, there were two independent variables: display size (large or small) and window type whether the system was using a gaze window or a corner window. When a corner window is used, it is statically placed at a corner of the display. The experiment took a within-subject 2x2 factorial design. Therefore, there were four conditions. As shown in Table 2, we use the acronyms SG (small display with gaze window), SC (small display with corner window), BG (big display with gaze window), and BC (big display with corner window) to represent the four conditions.

² <https://optitrack.com/>

Table 2. Two independent variables (display size, and window type for the user head shot) forming four experimental conditions, highlighted in grey

	Small display	Big display
Gaze window	Small display with the Gaze window (SG)	Big display with the Gaze window (BG)
Corner window	Small display with the Corner window (SC)	Big display with the Corner window (BC)

5.2.2 Participants

We recruited 22 participants (17 males and 5 females) aged 19 to 35 years old (*Mean* = 30.3; *Std.Dev.* = 7.7). All of them had experience of using video-conferencing applications or tools such as Skype, FaceTime before, However, none of them had any previous experience of using gaze window systems, despite that most of them had heard about eye tracking applications.

5.2.3 Procedure and the task

The experiment started with a pair of participants signing an informed consent form, answering a demographic information questionnaire and the experimenter describing the purpose of the study. Each participant sat in front of the small or the big display and calibrated their eye tracker.

Since the task was to watch a video and talk with their partner, participants did not have a training session, but we verbally explained and showed how the system worked. Participants watched a two-minute YouTube video that shows 360-degree seaside views for an immersive experience [57], and discussed about the video with their experimental partner. We recommended participants to talk about the objects they were looking at in the video, similar scenes they had seen before, experience of being in seaside places, or how they felt when watching the video. A pair of participants performed the task in all four conditions, and the order of the conditions was counter balanced based on a balanced Latin-square design. The participants switched places with their partners according to the order of the condition, so each participant in a pair tried all four conditions.

After each condition, participants answered Likert-scale rating questions adopted from the co-presence survey and perceived emotional interdependence sub-scales of the Harms' questionnaire [58], and two open-ended questions asking what they liked or disliked about the condition. The rating scale was from 1 (Strongly disagree) to 7 (Strongly agree). The experiment concluded with a final questionnaire asking participants to rank the condition according to their preference and stating the reason for their ranking. In addition to the questionnaires, we counted the number of times that the user looked at their partner's headshot, and the amount of time spent doing so from user activity log data. Overall, the experiment took about 60 minutes for each pair of participants.

5.2.4 Hypotheses

For this experiment, we had the following hypotheses:

(H4) the gaze window in teleconferencing encourages users to look at their partner more

(H5) the gaze window increases the level of co-presence and emotional interdependence

(H6) the effect of the gaze window is more prominent on a big display as it shows bigger picture of partner's head shot

5.3 Results

We analysed the participants' activity log using a repeated-measure ANOVA ($\alpha = .05$), and questionnaire results using a repeated-measure ANOVA with the aligned rank transform (ART) ($\alpha = .05$) [59]. Table 3 shows the results of the user gaze activities, and Table 4 shows the results of the questionnaire asking the level of co-presence and emotional interdependence.

Table 3. Results of the frequency and the amount of time spent looking at the partner's head shot (D_S: display size, W_T: window type, I: interaction between two factors). The significant results are highlighted in grey.

	<i>Mean (Std.Dev.)</i>				Factor	Result
	SG	SC	BG	BC		
Frequency (looking at partner's head shot)	28.5 (6.0)	13.7 (5.2)	20.3 (5.5)	4.9 (3.8)	D S	<i>F(1,21)=67.776, p<.001</i>
					W T	<i>F(1,21)=154.716, p<.001</i>
					I	<i>F(1,21)=.095, p=.761</i>
Amount of Time (looking at partner's head shot – in second)	20.4 (5.0)	9.2 (3.6)	15.2 (4.6)	3.6 (3.0)	D S	<i>F(1,21)=37.561, p<.001</i>
					W T	<i>F(1,21)=126.969, p<.001</i>
					I	<i>F(1,21)=.132, p=.719</i>

Table 4. Results of co-presence questions, and emotional interdependence questions (D_S: display size, W_T: window type, I: interaction between two factors). The significant results are highlighted in grey.

	<i>Mean (Std.Dev.)</i>				Factor	Result
	SG	SC	BG	BC		
Co-presence	5.8 (1.0)	4.3 (1.1)	5.4 (1.1)	3.7 (1.3)	D S	<i>F(1,21)=6.229, p=.020</i>
					W T	<i>F(1,21)=58.147, p<.001</i>
					I	<i>F(1,21)=.837, p=.371</i>
Emotional Interdependence	4.3 (1.4)	3.7 (1.4)	4.2 (1.4)	3.6 (1.2)	D S	<i>F(1,21)=.272, p=.608</i>
					W T	<i>F(1,21)=6.753, p=.017</i>
					I	<i>F(1,21)=.003, p=.959</i>

Both display size and window type factors had a significant main effect on the frequency and amount of time the participants looked at their partner (see Table 3). The data of the frequency and the amount of time looking at partner were normally distributed (Shapiro-Wilk test; frequency – SG: $W(22)=.969, p=.677$, SC: $W(22)=.956, p=.409$, BG: $W(22)=.966, p=.619$, BC: $W(22)=.921, p=.080$; amount of time – SG: $W(22)=.925, p=.096$, SC: $W(22)=.975, p=.823$, BG: $W(22)=.954, p=.377$, BC: $W(22)=.913, p=.055$). Participants looked at their partner's head shot significantly more frequently and for a longer amount of time with the small display than with the big display (frequency: $F(1,21)=67.776, p<.001$; amount of time: $F(1,21)=37.561, p<.001$). They also looked at their partner significantly more frequently and for a longer time with the gaze window than with the corner window (frequency: $F(1,21)=154.716, p<.001$; amount of time: $F(1,21)=126.969, p<.001$). There was no significant interaction between the two factors, display size and window type, in both frequency ($F(1,21)=.095, p=.761$) and amount of time ($F(1,21)=.132, p=.719$).

The co-presence results followed a similar trend as the results of participants' gaze activities, with a significant difference in both display size and window type conditions (see Table 4). The result showed that participants felt significantly higher co-presence with the gaze window than with the corner window ($F(1,21)=58.147, p<.001$), and with the small display than with the big display ($F(1,21)=6.229, p=.020$). However, the emotional interdependence results were different; only the window type had a significant effect but not the display size. Participants felt significantly better emotional interdependence with the gaze window condition than with the corner window ($F(1,21)=6.753, p=.017$), but there was no significant difference between the small and big display conditions ($F(1,21)=.272, p=.608$).

In the answers to the open-ended questions, we found that the big display provided more immersion compared to the small display, while the small display helped participants more easily focus on and

notice their partner and the video view. As the benefit of the big display, eight participants mentioned immersiveness (e.g. “*I preferred the large size (display) as I felt more immersed*”) or high level of the reality (e.g. “*The big monitor gives a much closer to reality view*”). However, some participants reported that with the small display it was easier to focus on and notice their partner and what was going on within the display (e.g. “*On the smaller screen, I was aware of what was going on*”; “*The smaller one (display) could be better because I could concentrate on my partner and the changes happening on the screen*”).

In terms of window types, participants reported that the gaze window made it easier to look at their partner, but also mentioned that one of the disadvantages of it: distracting for watching the video, especially by occluding the area behind the gaze window. Ten participants mentioned how easy it was to look at their partners (e.g. “*There were more chance to see his (partner’s) face*”; “*The gaze following window allows me to see my partner face easily*”). Two participants commented that the gaze window could block an area of the video behind it.

Participants ranked the conditions from 1 (best) to 4 (worst) according to their experience (see Table 5). On average, the BG condition was ranked as the best (*Mean*:1.64, *SD*: 0.85), the BC was the second (*Mean*: 2.59, *SD*:1.05), the SG was the third (*Mean*:2.77, *SD*:1.23), and the SC condition was ranked as the last (*Mean*:2.77, *SD*:1.23). Analysing the ranks using a Repeated-measure ANOVA with ART ($\alpha = .05$) [59] found that participants ranked the big display significantly better than the small display ($F(1,21)=8.877, p=.007$), and the gaze window better than the corner window ($F(1,21)=4.744, p=.041$). No significant interaction was found between the two factors ($F(1,21)=2.571, p=.124$).

Table 5. Participants’ average ranking of the conditions.

	<i>Mean (Std.Dev.)</i>				Factor	Result
	SG	SC	BG	BC		
Preference	2.77 (1.23)	3 (0.87)	1.64 (0.85)	2.6 (1.05)	D S	$F(1,21)=8.877, p=.007$
					W T	$F(1,21)=4.744, p=.041$
					I	$F(1,21)=2.571, p=.124$

5.4 Discussion and Limitation

In the study, we found that the participants looked at their partner’s headshot more frequently and for a longer time with the gaze window than with the corner window, and with the small display than with the big display. This would have affected the participants’ feeling of co-presence with their partner, so the gaze window and small display showed higher co-presence ratings than the corner window and big display. However, the participants preferred the big display significantly more than the small display. This contradiction could result from the factors that we did not measure by rating scale questions, such as immersiveness which was mentioned as the benefit of the big display in the open-end questions, similar to the many previous studies [60, 61].

Between the gaze and corner windows, the result of the emotional interdependence showed a similar trend with the co-presence. Participants felt a higher level of the emotional interdependence with the gaze window than the corner window. Watching videos aroused emotion [62] and participants more prominently shared the emotion through their facial expressions with the gaze window than the corner window as they looked at their partner’s head shot more frequently and for a longer time with the gaze window than with the corner window. Interestingly, the effect of more frequent and a longer time looking partner on the emotional interdependence was not found between the small and big displays as participants did not feel a significantly different level of emotional interdependence between them. This contradiction may come from factors we did not measure, such as the level of aroused emotion based on the display size. Participants might be emotionally more aroused with the immersive and big display

than with the small display, which might reduce the effect of frequent and longer times looking at their partner with the small display compared to the big display.

Another discussion topic is improving the gaze window interface for this task. Even though the gaze window did not cover the gaze point, some of the participants reported that the gaze window was distracting by occluding a part of video behind it. To solve this, we recommend changing the gaze window to make it semi-transparent when a user is not looking at it but becoming opaque again when a user is watching it.

The hypotheses of this study were that the gaze window encourages users to look at their partner more (**H4**), it increases the level of feeling co-presence and emotional interdependence (**H5**), and the effect is more prominent in the big display (**H6**). Two of these (**H4** and **H5**) were validated in the results. However, the effect of the gaze window was not that prominent in the big display compared to the small display (**H6** is not supported), although participants had a higher preference with the big display for its perceived greater immersion.

6. General Discussion

In this research, we investigated the use of the gaze window in two tasks: data entry and teleconferencing. With the data entry task, although participants did not complete the task earlier with the gaze window compared to the baseline, the gaze window showed benefits with reduced amount of gaze movements and reduced level of mental stress. This is likely because the participants focused on task performance in terms of completion time in the two conditions, but they achieved this at the cost of more gaze movements and higher mental stress. The limited benefits of gaze window in task completion time may also come from the waiting time for triggering the dwell time interaction (600ms) and diversity of user typing style. The dwell time (600ms) caused slower performance and some participants were not looking at the target cell for more than dwell time to move gaze window because of their typing style (mostly looking at the data-source image or/and the keyboard). If the participants were more familiar with the dwell time interaction, the results would be different. For the teleconferencing task, it was found that gaze window encouraged more facial contact and increased co-presence and emotional interdependence.

From the user studies, we gained following lessons: 1) the gaze window should be semi-transparent when moving and following the user gaze for a better view of the scene behind it, but becoming opaque when the user looks at it for certain amount of time, 2) since there are a variety of tasks requiring different eye movement, it would be better to provide the user with a configurable interface which allows them to define the gaze window anchor position relative to the gaze point, and 3) there should be proper feedback for the dwell time interaction, so that a user can easily know when the dwell time has started and its progress before triggering a command.

We understand that no experiment is perfect. But it should be noted that our studies have several limitations in the range of tasks and types of display devices used. If a user has more than two areas or objects of interest, our system would need to be extended to support such use cases. Since our studies involved 2D tasks on the screen space, it would be good to study 3D tasks in the future and the system should define proper depth position in the 3D tasks. Lastly, the experiment sample size was relatively small and participants were all from a student population. Particularly for the first study, we had only 12 participants. Although they were all from the same student population with similar background, to have more concrete results, it would better to recruit more participants from different backgrounds which would increase the reliability of the results and external validity.

The eye trackers used in the study also had limitations [16], such as not working properly when the user was outside of tracking area of the desktop eye tracker, or when the user accidentally touched the frame of the portable eye tracker. This did not happen during our studies because of the short use time, but in

longer term usage, it is likely that users will face such cases and the eye tracking data will become inaccurate and less reliable which is a common issue for eye-gaze applications.

7. Conclusion

In this paper, we proposed the novel concept of a gaze window interface that copies and displays an area of user interest close to the user gaze point and explored this in a single user task and a teleconferencing task. For the single user task, we conducted an experiment with a data entry task where the gaze window copied and showed the typing contents close to a user gaze point which was mostly at the typing point. The results showed that the gaze window significantly reduced the gaze movement and increased usability compared to baseline condition.

For the teleconferencing task, the system displayed the partner's headshot in the gaze window while users watched a synchronized video (i.e. 360 degrees scenery video) either on a small or big display. The results showed that users looked at their partner more frequently and for a longer time with the gaze window rather than with the corner window, and with a small size display rather than with a big size display. The level of co-presence was also increased in the conditions with the gaze window or with the small size display compared to the conditions with the corner window or with the big size display. In addition, users felt more emotional interdependence with the gaze window than with the corner window.

In the future, we will explore using the gaze window in a real world 3D task such as fixing a car while wearing an Augmented Reality based display such as Microsoft HoloLens. For example, an instruction video could be played in the gaze window while using an optical see-through display with an integrated eye tracker. We will also explore how users can share real-world experiences such as attending an evening party while sharing view of it through a videoconferencing, especially for investigating the effect of sharing emotion when the gaze window is showing the remote partner's bust shot.

References

- [1] Bruce, V, Green, PR, Georgeson, MA. Visual perception: Physiology, psychology, & ecology. Psychology Press; 2003.
- [2] D. Luebke, M. Reddy, J.D. Cohen, A. Varshney, B. Watson, and R. Huebner, Level of Detail for 3D Graphics, first edition, Morgan Kaufmann, 2002. Cited on page(s) 5, 6
- [3] J. A. Ferwerda, "Elements of Early Vision for Computer Graphics," IEEE Computer Society Press, vol. 21, no. 5, pp. 22–33, 2001. DOI: 10.1109/38.946628 Cited on page(s) 5, 6 2001
- [4] Duchowski, A. T. (2018). Gaze-based interaction: A 30 year retrospective. Computers and Graphics (Pergamon), vol.73, pp.59–69. <https://doi.org/10.1016/j.cag.2018.04.002>
- [5] Sundstedt V. Gazing at games: an introduction to eye tracking control. Syn- thesis lectures on computer graphics and animation. Morgan & Claypool Publishers; 2012. doi: 10.220 0/S0 0395ED1V01Y201111CGR014.
- [6] T. Vilis, The Physiology of the Senses Transformations for Perception and Action, Course Notes, University of Western Ontario, Canada, 2006. Cited on page(s) 6, 7
- [7] Duchowski A. T., Eye Tracking Methodology: Theory and Practice, Springer-Verlag, 2003
- [8] Snowden R., Thompson P., and Troscianko T., Basic Vision: an introduction to visual perception. Oxford University Press, 2006.
- [9] W. Huang. 2013. Establishing aesthetics based on human graph reading behavior: two eye tracking studies. Personal and Ubiquitous Computing 17(1): 93-105 (2013).
- [10] Jacob, RJK. The use of eye movements in human-computer interaction techniques: What you look at is what you get. ACM Trans Inf Syst 54 1991;9(2):152–169. 55
- [11] Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In: Proceedings of the ACM CHI'99 Human Factors in Computing Systems Conference (pp. 246–253). Addison-Wesley/ACM Press

- [12] Majaranta, P, Bulling, A. *Eye Tracking and Eye-Based Human–Computer Interaction*. London: Springer London; 2014. ISBN 978-1-4471-6392-3. URL: https://doi.org/10.1007/978-1-4471-6392-3_3. doi:10.1007/978-1-4471-6392-3_3.
- [13] Velloso, Eduardo, and Marcus Carter. "The emergence of eyeplay: a survey of eye interaction in games." *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. ACM, 2016
- [14] Poika Isokoski, Markus Joos, Oleg Spakov, and Benoît Martin. 2009. Gaze controlled games. *Univers. Access Inf. Soc.* 8, 4 (Oct. 2009), 323–337. DOI: <http://dx.doi.org/10.1007/s10209-009-0146-3>
- [15] Majaranta, P, Raïiha, KJ. Twenty years of eye typing: Systems and design issues. In: *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*. ETRA '02; New York, NY, USA: ACM; 2002, p. 15–22.
- [16] Mohamed Khamis, Florian Alt, and Andreas Bulling. 2018. The past, present, and future of gaze-enabled handheld mobile devices: survey and lessons learned
- [17] Emiliano Castellina and Fulvio Corno. 2008. Multimodal gaze interaction in 3D virtual environments. *Proceedings of the Conference on Communication by Gaze Interaction (COGAIN) 8* (2008), 33–37. <http://www.cogain.org/>
- [18] James Gips and Peter Olivieri. 1996. EagleEyes: An eye control system for persons with disabilities. In *The Eleventh International Conference on Technology and Persons with Disabilities*. 1–15
- [19] Inger M. Ekman, Antti W. Poikola, and Meeri K. Mäkäräinen. 2008a. Invisible Eni: Using Gaze and Pupil Size to Control a Game. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. ACM, New York, NY, USA, 3135–3140. DOI: <http://dx.doi.org/10.1145/1358628.1358820>
- [20] Dorr, M, Böhme, M, Martinetz, T, Barth, E. Gaze beats mouse: a case study. *Proceedings of COGAIN 2007*;16–19.
- [21] Stephen Vickers, Howell Istance, and Matthew Smalley. 2010. EyeGuitar: Making Rhythm Based Music Video Games Accessible Using Only Eye Movements. In *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology (ACE '10)*. ACM, New York, NY, USA, 36–39. DOI: <http://dx.doi.org/10.1145/1971630.1971641>
- [22] T. Wilcox, M. Evans, C. Pearce, N. Pollard, and V. Sundstedt, "Gaze and Voice Based Game Interaction: The Revenge of the Killer Penguins," in *Proceedings of International Conference on Computer Graphics and Interactive Techniques*, 2008.
- [23] Michiya Yamamoto, Munehiro Komeda, Takashi Nagamatsu, and Tomio Watanabe. 2011. Hyakunin-Eyesshu: A Tabletop Hyakunin-Isshu Game with Computer Opponent by the Action Prediction Based on Gaze Detection. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications (NGCA '11)*. ACM, New York, NY, USA, Article 5, 4 pages. DOI: <http://dx.doi.org/10.1145/1983302.1983307>
- [24] Pfeuffer, Ken, et al. "Gaze-touch: combining gaze with multi-touch for interaction on the same surface." *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 2014.
- [25] Pfeuffer, Ken, et al. "Gaze-shifting: Direct-indirect input with pen and touch modulated by gaze." *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 2015.
- [26] Serim, Baris, and Giulio Jacucci. "Pointing while looking elsewhere: Designing for varying degrees of visual guidance during manual input." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016.
- [27] Göbel, Fabian, et al. "Improving map reading with gaze-adaptive legends." *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, 2018.
- [28] Ishii, H., & Kobayashi, M. (1992, June). ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 525-532). ACM.

- [29] Vertegaal, R. The gaze groupware system: Mediating joint attention in multiparty communication and collaboration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '99; New York, NY, USA: ACM; 1999, p. 294–301.
- [30] Vertegaal, R, Ding, Y. Explaining effects of eye gaze on mediated group conversations:: Amount or synchronization? In: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work. CSCW'02; New York, NY, USA: ACM; 2002, p. 41–48
- [31] Carletta, J, Hill, RL, Nicol, C, Taylor, T, De Ruiter, JP, Bard, EG. Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods* 2010;42:254–265.
- [32] Neider M. B., Chen X., Dickinson C. A., Brennan S. E., Zelinsky G. J.: Coordinating spatial referencing using shared gaze. *Psychonomic bulletin & review* 17, 5 (2010), 718–724.
- [33] Brennan S. E., Chen X., Dickinson C. A., Neider M. B., Zelinsky G. J.: Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106, 3 (2008), 1465–1477.
2
- [34] Stein R., Brennan S. E.: Another person's eye gaze as a cue in solving programming problems. In Proceedings of the 6th international conference on Multimodal interfaces (2004), ACM, pp. 9–15. 2
- [35] Ou J., Oh L. M., Fussell S. R., Blum T., Yang J.: Predicting visual focus of attention from intention in remote collaborative tasks. *IEEE Transactions on Multimedia* 10, 6 (2008), 1034–1045. 2
- [36] Higuch K., Yonetani R., Sato Y.: Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2016), ACM, pp. 5180–5190. 1, 2, 3
- [37] Barathan S., Lee G., Billinghamurst M., Lindeman R.: Sharing gaze for remote instruction. In International Conference on Artificial Reality and Telexistence - Eurographics Symposium on Virtual Environments 2017 (2017), Eurographics. 2, 3, 4
- [38] Lee, G. A., Kim, S., Lee, Y., Dey, A., Piumsomboon, T., Norman, M., & Billinghamurst, M. (2017). Improving Collaboration in Augmented Video Conference using Mutually Shared Gaze. In Proceedings of ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, pp. 197-204.
- [39] Yarosh, S., Inkpen, K. M., & Brush, A. J. (2010, April). Video playdate: toward free play across distance. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(pp. 1251-1260). ACM
- [40] Kato, H., & Billinghamurst, M. (1999). Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99) (pp. 85-94). IEEE.
- [41] Kim, S., Billinghamurst, M., & Lee, G. (2018). The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration. *Computer Supported Cooperative Work (CSCW)*, 27(3-6), 569-607
- [42] Kim, S., Junuzovic, S., & Inkpen, K. (2014, November). The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. In Proceedings of the 18th International Conference on Supporting Group Work (pp. 167-177). ACM.
- [43] Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*.
- [44] Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875), 86-88.
- [45] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2)
- [46] Ware, C, Mikaelian, HH. An evaluation of an eye tracker as a device for computer input2. In: Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface. CHI '87; New York, NY, USA: ACM; 1987, p. 183–188.
- [47] <https://github.com/Tobii/interaction-samples>
- [48] J. Brooke. Chapter 21. SUS-A quick and dirty usability scale. In *Usability Evaluation in Industry*, (P. W. Jordan, B. Thomas, I. L. McClelland, B. Weerdmeester ed.), pp. 189-194, 1996.

- [49] Zijlstra, F. (1993). Efficiency in work behavior. A design approach for modern tools. PhD thesis, Delft University of Technology. Delft, The Netherlands: Delft University Press.
- [50] Zijlstra, F.R.H & Doorn, L. van (1985). The construction of a scale to measure subjective effort. Technical Report, Delft University of Technology, Department of Philosophy and Social Sciences
- [51] Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114-123.
- [52] <https://github.com/sakapon/Samples-2014/tree/master/Kinect%20SDK/v1.8/BackgroundRemovalBasics-WPF>
- [53] <https://www.microsoft.com/en-us/download/details.aspx?id=40276>
- [54] https://github.com/pupil-labs/pupil/blob/master/pupil_src/launchables/eye.py
- [55] <https://github.com/Vangos/kinect-2-background-removal>
- [56] <https://www.microsoft.com/en-us/download/details.aspx?id=44561>
- [57] Brown, A, Turner, J, Patterson, J, Schmitz, A, Armstrong, M, Glancy, M. Subtitles in 360-degree video. In: Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video. ACM; 2017, p. 3–8.
- [58] Harms, C., & Biocca, F. (2004). Internal consistency and reliability of the networked minds measure of social presence.
- [59] Wobbrock J. O., Findlater L., Gergle D., Higgins J. J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In Proceedings of the SIGCHI conference on human factors in computing systems (2011), ACM, pp. 143–146.
- [60] Keefe, D. F., Feliz, D. A., Moscovich, T., Laidlaw, D. H., & LaViola Jr, J. J. (2001, March). CavePainting: a fully immersive 3D artistic medium and interactive experience. In Proceedings of the 2001 symposium on Interactive 3D graphics (pp. 85-93). ACM.
- [61] Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., & Strehlke, K. (2003, July). blue-c: a spatially immersive display and 3D video portal for telepresence. In *ACM Transactions on Graphics (TOG)* (Vol. 22, No. 3, pp. 819-827). ACM.
- [62] Vermeulen, Jo, et al. "Heartefacts: Augmenting mobile video sharing using wrist-worn heart rate sensors." Proceedings of the 2016 ACM Conference on Designing Interactive Systems. ACM, 2016
- [63] Huang, W., 2007. Using eye tracking to investigate graph layout effects. *APVIS 2007*: 97-100
- [64] Kim, S., Billinghamurst, M., Lee, G.A., Norman, M., Huang, W., He, J. 2019. Sharing Emotion by Displaying a Partner Near the Gaze Point in a Telepresence System. *IV (2) 2019*: 86-91
- [65] Kim, S. Lee, G.A., Huang, W., Kim, H., Woo, W., Billinghamurst, M. 2019. Evaluating the Combination of Visual Communication Cues for HMD-based Mixed Reality Remote Collaboration. *CHI 2019*: 173