Exploration of Techniques for Rapid Activation of Glanceable Information in Head-Worn Augmented Reality

Feiyu Lu
feiyulu@vt.edu
CHCI, Dept. of Computer Science,
Virginia Tech
Blacksburg, VA, USA

Shakiba Davari
sdavari@vt.edu
CHCI, Dept. of Computer Science,
Virginia Tech
Blacksburg, VA, USA

Doug A. Bowman
dbowman@vt.edu
CHCI, Dept. of Computer Science,
Virginia Tech
Blacksburg, VA, USA

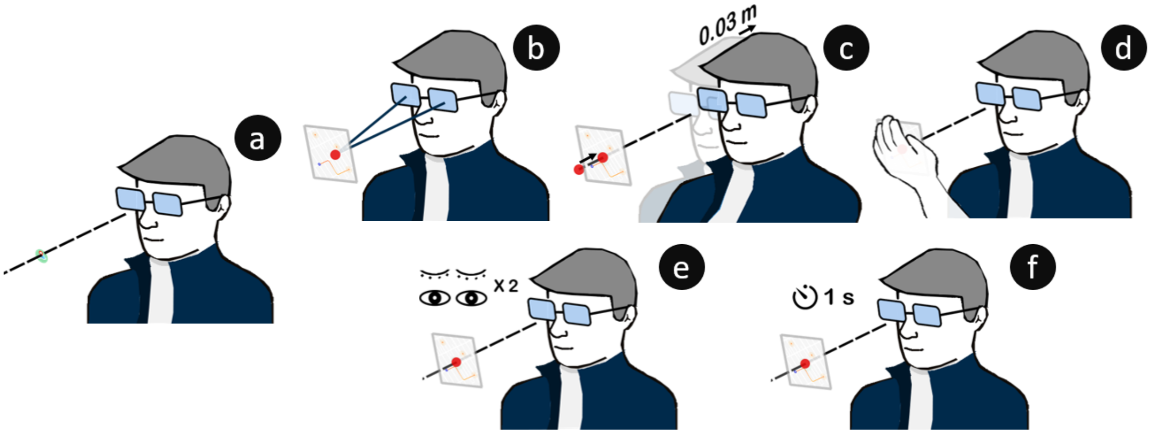

Figure 1: (a) When the user is looking in the direction of minimized virtual content, we propose five interaction techniques to activate the virtual information: (b) Fixation-Glance, in which users converge their gaze at the depth of the content; (c) Head-Depth, in which users lean backward three centimeters; (d) Hand-Overlay, in which users put their hand slightly behind the virtual content; (e) Blink, in which users blink the eye twice within one second; and (f) Dwell, in which users maintain their gaze on the virtual content for one second.

ABSTRACT

Future augmented reality (AR) glasses may provide pervasive and continuous access to everyday information. However, it remains unclear how to address the issue of virtual information overlaying and occluding real-world objects and information that are of interest to users. One approach is to keep virtual information sources inactive until they are explicitly requested, so that the real world remains visible. In this research, we explored the design of interaction techniques with which users can activate virtual information sources in AR. We studied this issue in the context of Glanceable AR, in which virtual information resides at the periphery of the user's view. We proposed five techniques and evaluated them in both sitting and walking scenarios. Our results demonstrate the usability, user preference, and social acceptance of each technique, as well as design recommendations to achieve optimal performance. Our findings can inform the design of lightweight techniques to activate virtual information displays in future everyday AR interfaces.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Empirical studies in interaction design**; **Mixed / augmented reality**.

KEYWORDS

interaction technique, adaptive interface, glanceable information, augmented reality, user study

ACM Reference Format:
Feiyu Lu, Shakiba Davari, and Doug A. Bowman. 2021. Exploration of Techniques for Rapid Activation of Glanceable Information in Head-Worn Augmented Reality. In *Symposium on Spatial User Interaction (SUI '21), November 9–10, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3485279.3485286

1 INTRODUCTION

With the recent advancements in display, computing, and tracking technologies, augmented reality head-worn displays (AR HWDs) are becoming increasingly lightweight and powerful. In the recent

---

SUI '21, November 9–10, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-9091-0/21/11…$15.00
https://doi.org/10.1145/3485279.3485286

years, more and more AR HWDs (or AR glasses) have been announced in the industry (e.g., nreal glasses[1], Spectacles by Snap[2], ThinkReality by Lenovo[3]). Different from conventional AR headsets that are bulky and designed for special-purpose applications, these displays are close to achieving the form factor of a normal pair of eye-glasses, and are targeting personal uses and everyday consumers.

Recent research in AR are gradually moving outside research labs to the fields [19, 26, 38]. In the near future, AR glasses are likely to be general-purpose and worn all-day, displaying all sorts of virtual content to provide always-on access to information [8, 11, 40]. However, when virtual content becomes available continuously and pervasively, it could be distracting or overwhelming, and it could occlude real-world objects of importance to users. To address these challenges, Lu et al. proposed Glanceable AR, in which virtual content was displayed as glanceable widgets in the periphery of the users' view [26, 27], so that users can perform a quick glance to access the information they need.

Although Glanceable AR is a promising solution to enable always-on information display, whether it completely addresses the occlusion challenge remains questionable. Virtual information in the periphery can still block users' view when they intend to glance at the real world behind the virtual content. To make glanceable information more unobtrusive, an adaptive information display strategy can be applied, in which virtual content is minimized when not needed, but can be activated through interaction techniques to show more detailed information [6, 24, 29, 33–35]. In this way, the real world is "prioritized", and users can explicitly control what they see depending on their needs [4, 33].

However, effective activation techniques for glanceable content must address several challenges. First, the techniques have to be *rapid*, as users need to be capable of performing them within the duration of a quick glance. Second, they need to be *lightweight*, so that users can allocate their limited cognitive and attentional resources to their primary tasks and to events in the environment, rather than diverting cognitive effort to the interface. Third, they must be *reliable*, since unintentional activation could occlude important information, and failure to activate on demand could lead to distraction and frustration.

In this research, we first conducted a user study to validate the occlusion and distraction issues in Glanceable AR interfaces. We then addressed the design challenges by proposing a set of design considerations for developing techniques to explicitly activate virtual content in Glanceable AR interfaces. Next, we proposed five techniques for activating the minimized glanceable widgets, and evaluated them in controlled experiments. Two common everyday scenarios were replicated in the experiments: (1) sitting in front of a desktop computer; and (2) walking in an indoor environment.

The contributions of the paper include: (1) validation of occlusion and distraction issues in Glanceable AR, (2) design considerations for developing techniques to activate minimized glanceable information, (3) design and evaluation of five promising techniques, and (4) design recommendations for future implementations of these techniques.

2 RELATED WORK

2.1 Everyday information acquisition with AR HWDs

ARWin was an early prototype that explored AR for everyday information access by augmenting a physical workspace with virtual calendar, clock and weather information [5, 6]. Recent research highlighted an important issue to be addressed for everyday AR interfaces, which is how to handle the potential occlusion of virtual information overlaying the real-world environments that are of interest to the users [4, 22]. Lu et al. proposed Glanceable AR, in which virtual information resides at the peripheral vision to avoid occluding user's view when not needed [26, 27]. However, it remains unclear if such an approach can completely address the occlusion problems. In this research, we attempted to validate the occlusion problem for glanceable information display with AR HWDs, and proposed lightweight user interactions to solve the issue of unwanted occlusions caused by virtual information overlaying the real-world environments that we are interested in.

2.2 Adaptive Information Display in AR/VR

One way to alleviate the occlusion issues caused by augmented virtual information is adaptive information display. In 2003, Vertegaal first proposed Attentive User Interfaces (AUI), a system that was continuously aware of what users pay attention to and adapted its services [39]. In 2019, Lindlbauer et al. proposed an online context-aware method to automatically adapt the position and level of detail (LoD) of virtual information [24]. Such systems usually require a large amount of knowledge about users, their preferences, real-time cognitive load, the surrounding physical environment, and the tasks they are engaging in. In most cases, external sensors and cameras are required to obtain these data, which could increase costs and induce privacy concerns [11, 20, 44]. Another way of displaying the information adaptively is through user's explicit interactions [6, 12, 33, 35]. Such methods put more efforts on the user side as compared to automatic detection on the system-level, but could be more predictable and accurate, especially in everyday situations when contextual information changes frequently. Similarly, in this research, we focused on situations where user adapt the information display level through explicit, but lightweight, interactions in head-worn AR systems.

2.3 Explicit interactions in AR/VR

Existing research has mainly explored three categories of input methods for explicit activation of virtual content: controller, head, and gaze-based interactions. Davari et al. explored pointing and clicking-based interactions with a hand-held controller [4]. However, controllers are less likely to be always available in daily situations. In 2004, Diverdi et al. proposed LoD interfaces, in which virtual content was displayed with different levels of detail (LoDs) based on the distance between the head position and the content [6]. Yu et al. explored a similar concept by calculating the distance of head movement in the depth dimension [43]. Gaze-adaptive interfaces have been extensively explored in the literature. Kim et al. explored displaying relevant information based on user gaze point in a 2D screen space [18]. Pfeuffer et al. and Piening et al. explored

---

[1] https://www.nreal.ai/
[2] https://www.spectacles.com/
[3] https://www.lenovo.com/us/en/thinkrealitya3

adapting the transparency and information level in the virtual content through gaze directions [33, 34]. Lu et al. explored using gaze depth to distinguish users' attention between virtual information and the real world to modify the size of the virtual information [26]. Hand input is gaining popularity with the advancements in tracking technologies, but most of the existing work focused on special-purpose use cases such as text-entry, target selection, and games [30, 37, 42]. There has been a lack of research regarding how hand-based input could be utilized to explicitly activate minimized virtual information in everyday situations. In this research, we focus on the integration of head, gaze and hand-based input specifically for lightweight activation of virtual content in AR HWDs.

3 PRELIMINARY STUDY: VALIDATING OCCLUSION ISSUES

Before diving into the solution space, we conducted a preliminary study to validate whether occlusion issues exist when virtual content is placed unobtrusively in the periphery. We varied both the location and the LoD of the virtual content to understand the effects of these factors on occlusion.

3.1 Interface conditions

In the study, participants were asked to wear a Magic Leap One AR headset. The headset displayed four pieces of virtual information (weather, calendar, email, to-do list) in different forms that varied in LoD and location. The virtual information was not interactive—it was always displayed, so that there were opportunities for the virtual content to occlude the real world. We used two ways of locating the information in the periphery (heads-up display (HUD) or head-glance (HG)) and two LoDs (Full-app or Icon), yielding four interfaces in total (see Figure 2). In the HUD condition, content was display-fixed at the edges of the field of view (FoV) of the AR display and always visible. Such HUDs are commonly used in commercial products without head tracking such as Google Glass[4] and North Focals[5]. The HG condition used a glanceable AR interface proposed by Lu et al. that showed potential for everyday uses [26]. In HG, content was fixed to the torso of the user's body and stayed outside the central view of the users. It became visible only when the user turned their head.

Participants were asked to wear a belt with the Magic Leap controller placed in a 3D-printed case to track their body orientation. The four interface conditions included were : (1) HUD-Full; (2) HUD-Icon; (3) HG-Full; (4) HG-Icon; and finally (5) None condition, in which no virtual content was displayed, as a baseline condition.

3.2 Tasks

In the study, participants were asked to walk in a 6.22 by 9.95 meter indoor environment casually and identify certain paper signs in the real world (see Figure 4 (b)). This task was designed to represent common scenarios in which people walk around in the real world and search visually for a target or destination (e.g., a conference room in a building with a certain room number). Twenty signs were distributed randomly in the space on walls, whiteboards, floors, and tables. Participants were asked to wear the AR headset, walk in an figure-eight path at a comfortable speed, and search for signs that contained a certain letter, which varied with every trial. Only some of the signs contained the target letter. Participants were asked to speak out loud the word in each sign that contained the letter as soon as they noticed it. Interface condition was the only independent variable for the study. We used a within-subjects design, in which each participant experienced the five interface conditions one by one. Latin square counterbalancing was applied to the order in which conditions were experienced. The task took about one minute to complete for each condition.

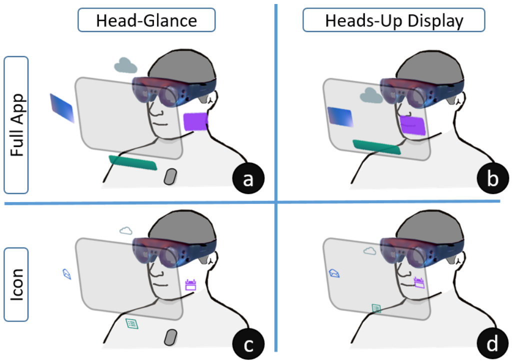

Figure 2: The four interfaces in the first study with information at the periphery: (a) HG-Full; (b) HUD-Full; (c) HG-Icon; and (d) HUD-Icon.

3.3 Participants and Procedure

Eight participants (7M/1F) between 19 and 33 years old (M=24.37, SD=1.64) were recruited for the study. The participants were all college students. Four participants did not have experience with AR before the study.

The experiment was divided into five phases. First, participants were welcomed upon arrival, and were asked to read and sign the consent form (the study was approved by the Institutional Review Board of the university). Second, participants were asked to fill out a background questionnaire. Third, participants were given a brief introduction to the experiment background, hardware, five conditions, and the tasks involved in the study. Fourth, after participants had no further questions, they were instructed to put on the AR headset and go through the conditions one by one. After each condition, participants were asked to fill out a questionnaire containing Likert-scale questions that asked about the level of distraction and intrusiveness of the AR interface (see Figure 3). After finishing all conditions, a brief interview was conducted, in which we asked participants about their preferences and comments about the AR interfaces. The study took 30 minutes in total.

3.4 Results and Discussions

Results in the Likert ratings showed that all four interface conditions caused some visual interference (occlusion) when users were visually searching the signs (see Figure 3 (a-c)). The HUD-Full condition was found to be the most annoying and got in the way the most often.

In the interview, ignoring the baseline condition, six out of eight (75%) participants ranked the HG-Icon condition the most preferred

---

[4] https://www.google.com/glass/start/

[5] https://www.bynorth.com/

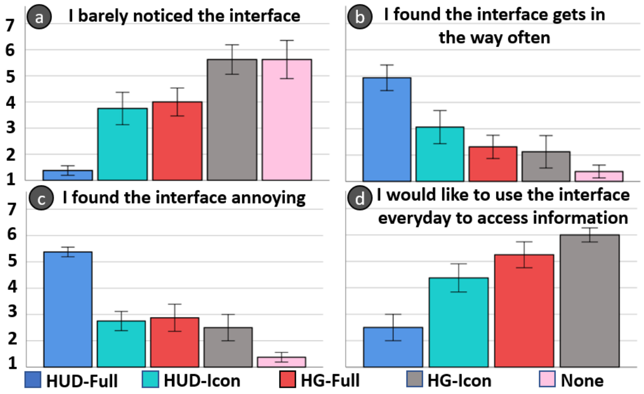

Figure 3: 7-point scale ratings on four statements (a-d).

and found it the least distracting. Seven out of eight (87.5%) participants disliked the HUD-Full condition, and found it the most distracting. For the HUD-Full condition, participants commented: *the icons being fully displayed greatly reduced my vision and distracted my focus*, and *I needed to pause to read the sign because I was trying to look past the items.* Similar comments were made for the HUD-Icon condition: *The icons were obtrusive, they felt like it was crowding my field of vision.* For the HG-Icon condition, participants commented: *It fits comfortably and doesn't limit my view too much*, and *Icons are small and don't take up much of the visibility.* For the HG-Full condition, participants commented: *I liked the interface because it wasn't directly in my line of sight.* However, four participants (50%) felt that it was still distracting to display the full app in HG-Full. They commented: *It still gets in the way sometimes when I look for the signs,* and *I would prefer the icon one [HG-Icon] better.*

In general, our results indicate that even when content is placed in the periphery, it can still be noticeable and annoying, and it can still visually interfere by occluding the real world during visual search tasks. However, these effects can be reduced by placing the content outside the central visual field (using the HG technique) and by reducing the LoD to a less obtrusive icon. Using an icon to represent the virtual content leads directly to the need for techniques that can activate the full virtual content when the user wishes to access that information.

4 DESIGN CONSIDERATIONS FOR ACTIVATION TECHNIQUES

Having established that Glanceable AR interfaces can still lead to occlusion and distraction when examining real-world information, we next brainstormed possible methods for accessing virtual content while minimizing occlusion problems. Based on the results of the preliminary study, we decided to use a real-world prioritized strategy, where virtual content is attached to the body and represented as small, semi-transparent icons by default (similar to the HG-Icon condition) [4, 33]. We then need methods for activating the virtual content in order to display the full version, and for deactivating the virtual content when the user is finished with it. We highlight five aspects below that need to be carefully considered for designing the solution interfaces.

Default visualization: When virtual information is minimized by default, it becomes critical for the users to be aware of where the information is so that they can activate and look at the information. Results of the preliminary study shows that representing minimized virtual information by an icon could be an effective strategy to prevent the virtual information from blocking users' view, while still maintaining some level of awareness towards where to look at to activate the content. As such, we apply the same strategy and set icons as the virtual information's default inactive state. Upon activation, the icon would go away and be replaced by the corresponding full-version virtual content.

Deactivation: Virtual information needs to be deactivated (minimized) automatically when users stop paying attention to it. A straightforward method is to use user's real-time gaze direction. Gaze direction provides a good indication of where users' current visual attention is located [9]. When they are no longer gazing in the direction of a piece of virtual information, we can assume that they do not need the information anymore, and thus automatically minimize the information.

Activation: Different from deactivation, detecting the intent of activation is challenging. Noticing that users are looking in the direction of a piece of virtual content is not enough to imply that they want to activate the information. It is possible that their intention is to see the real-world environment behind the content. To prevent false activation of the virtual content and occluding the user's view, we need to design suitable techniques for users to disambiguate their interest layer (i.e., real world or virtual content) explicitly and rapidly. Existing research has mainly explored explicit interactions in selection tasks [3, 21, 31, 43]. To our knowledge, our work is the first that looks at these interactions to disambiguate users' interest between AR content and real-world environments, which we believe is crucial if AR information is to be integrated pervasively in daily situations.

Input modality: which input modalities have been applied for activating the virtual information. In this research, we explore controller-free interactions, which include gaze, hand, and head-based techniques to activate the glanceable content.

Depth dimension: whether depth information is considered as part of the input dimension. Occlusion is one of the strongest depth cues [7]. When virtual information occludes the real world, it becomes intuitive to the users that the virtual information is at a closer layer of depth while the real-world is at a farther layer of depth. Enabling users to indicate the target depth layer of interest with the techniques could be intuitive in activating the virtual content. However, indicating the target depth could also increase user workload.

5 ACTIVATION TECHNIQUES

Based on the design considerations, we propose five techniques for explicit activation of virtual information: Dwell, Bink, Fixation-Glance, Hand-Overlay, and Head-Depth. Table 1 shows a summary of the five techniques.

5.1 Dwell

The most basic technique to prevent false activation of virtual information s the gaze-based dwell technique [16]. It was developed to

Table 1: The five techniques to activate or deactivate virtual information and the hypothesized trade-offs.

Technique	Activation			Deactivation	Depth	Accessibility	Robustness
Dwell	Gaze at the direction of the content	+	Maintain the gaze for 1 second	Stop gazing at the direction	No	High	Low
Blink			Blink twice within 1 second		No	Medium	Medium
Fixation-Glance			Converge gaze at the content's depth		Yes	High	Medium
Hand-Overlap			Put hand slightly behind the content's depth		Yes	Low	High
Head-Depth			Move the head backward in the depth dimension for 0.03m		Yes	Low	High

avoid the effects of eye tracking jitter and the "Midas Touch effect." To activate the virtual information, users gaze in its direction and maintain their gaze for one second. The content does not appear if users look away from the target before the end of the one-second period (see Figure 1 (f)). We consider Dwell as the baseline technique, since it is a widely adopted method to avoid false-positive activation of virtual content.

5.2 Blink

Blinking has been explored in the literature in object selection tasks and assistive technologies [3, 10]. Recently, Lu et al. found that blink is advantageous as compared to dwell for text-entry in VR [28]. It could be an effective strategy to explicitly activate virtual information. With the Blink technique, users activate the virtual information by looking in its direction and blinking their eyes. We need to prevent content from being activated by involuntary blinks; therefore, we use multiple consecutive blinks. However, the increased number of required blinks could induce eye fatigue and increase the activation time. We initially experimented with three blinks in a row within 1.5 seconds. which was found to induce eye fatigue with prolonged use. Through iterative testing, we found that two consecutive blinks within a one-second window produced a good balance between speed and accuracy (see Figure 1 (e)). A white dot is displayed below the virtual content when the first blink from the user is successfully registered. Then the user can blink again to activate the virtual information.

5.3 Fixation-Glance

Gaze-depth has been explored in object selections, visualizations, and information acquisition [14, 26, 32]. The Fixation-Glance technique utilizes gaze as the input modality with the depth of gaze being considered as an extra input dimension. To activate the virtual information, users need to not only look in its direction, but also converge the gaze at the depth of the information (see Figure 1 (b)). As such, if users intend to look at the real-world environment behind the virtual content, virtual information does not appear. The design of Fixation-Glance closely follows the idea of natural user interfaces (NUI), in which the user operates through intuitive interactions related to everyday natural human behavior [25]. In everyday life, our gaze naturally fixates in the direction and at the depth of objects that are of interest in order to see them clearly. The Fixation-Glance interface takes advantage of this process so that mental workload is minimized. However, due to technological limitations, current binocular eye-trackers only provide accurate estimation of gaze depth when it is within two meters. We used approximately 0.4 meters as the depth to activate the AR content because real-world objects are less likely to appear at this depth, and it was easy for users to converge at this depth without discomfort.

5.4 Hand-Overlay

Previous research on eye-hand coordination has found that eye and hand benefit from each other in selecting and manipulating objects [15, 17]. The Hand-Overlay technique utilizes the hand as the input modality with the hand depth being considered as an extra input dimension. To activate the virtual information, the user looks in its direction and puts their hand slightly behind the virtual content, blocking the real-world environment to indicate their interest in the depth layer in front of the hand (i.e., the virtual information). Virtual information activates only when the gaze ray intersects with the virtual information and the user's hand simultaneously (see Figure 1 (d)).

5.5 Head-Depth

The Head-Depth technique utilizes the head as the input modality with the head depth as an extra input dimension. To activate the virtual content, the users looks in its direction and then leans backward to activate it (see Figure 1 (c)). Following the design guidelines proposed by Yu et al., we used 3 cm (~1.18 inch) as the distance threshold to activate the virtual information [43]. The idea of using head depth was first explored in adaptive interfaces such as the LoD interface [6] and the proximity-aware user interface [12]. It was proved to be faster as compared to gaze-based dwell interactions to select objects in 3D environments [43].

5.6 Trade-offs among interfaces

Table 1 summarizes the five activation techniques, as well as our hypothesized trade-offs related to accessibility (ease of access) and robustness (avoidance of false activation). The Dwell technique offers high accessibility, as the user only needs to keep looking at the content without performing other actions. However, it would be low in robustness because looking in the direction alone does not necessarily indicate that users intend to look at the virtual content. Blink improves robustness by adding two consecutive blinks as a security layer when user is looking in the virtual content's direction. However, we blink involuntarily all the time, so false activations can still occur. Intentional blinks could also induce eye strain when used frequently in a short period of time, reducing the accessibility. Similarly, Fixation-Glance improves robustness by requiring users to converge or diverge their gaze. We are positive about its accessibility because of natural eye fixations. However, Fixation-Glance may also be more susceptible to false activation because of involuntary eye movements. Hand/head-based input modalities offer high robustness, because they require actions that are less likely to be performed unintentionally by users. However, that also lowers the accessibility and potentially makes them more intrusive to co-present others.

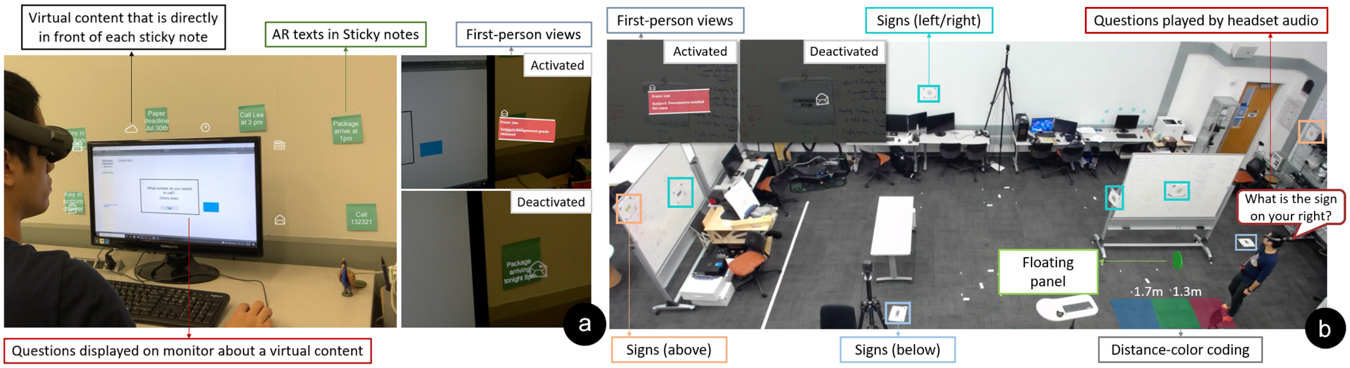

Figure 4: Illustrations of the two task contexts: (a) the sitting context; (b) the walking context.

6 PRIMARY STUDY: TECHNIQUE EVALUATION

A primary study was conducted with two research goals: (1) to validate the trade-offs we hypothesized and (2) to explore what factors could influence the usability, user preference, and acceptance of these interfaces.

6.1 Participants

Fifteen participants (7M/8F) between 19 and 35 years old (M=23.73, SD=4.48) were recruited for the study. Four of them did not have prior experience with AR before the study. Participants all had corrected or uncorrected near-perfect vision.

6.2 Tasks

Two types of task contexts were included in the study: sitting in front of a desktop computer and walking in an indoor environment. A 5 × 2 within-subject design was used for the study with two independent variables: five *interfaces* (Dwell, Blink, Fixation-Glance, Hand-Overlay, and Head-Depth), and two *task contexts* (Sitting and Walking). For both contexts, participants were asked to wear an AR headset, and access information in either the virtual content or in the real-world environment behind the virtual content using each of the five techniques. Participants were instructed to access the information as fast as they could (see Figure 4).

6.2.1 Sitting task context. In the sitting task, participants were asked to sit in front of a desktop monitor placed in front of a wall surface. The wall was approximately 0.7 meters away from the participant's head. Six blank sticky notes were attached to the wall. Text was rendered on each sticky note using the AR display (although the sticky notes represented "real-world" content, we used AR rendering so that the text could be changed for each trial) (see Figure 4 (a)). Around 0.3 meters in front of each sticky note, a piece of virtual content was displayed. The virtual content and the sticky note always overlapped with each other no matter where the user's head was located. In real-world use cases, it is unlikely that users would choose to place the virtual content and the sticky notes so that they overlap. However, we considered a sticky note to be analogous to an attention-grabbing event in the real-world environment that happens to share the same direction with the virtual information. If the virtual content is activated inadvertently by the user, it would expand and occlude users from seeing information in the sticky notes. Thus, this setup allowed us to evaluate how robust each interface condition was towards false positives.

Participants were asked to answer questions displayed on the computer screen. Participants saw four questions regarding the virtual content, and four questions regarding information in the sticky notes, yielding a total of eight questions. After participants finished answering a question, they used a mouse to click on a "Next" button to proceed to the next question on the computer screen. The AR headset and the computer were connected via network, so when users hit the "Next" button, information in the virtual content and the sticky notes in the AR headset automatically updated. As such, users always had to look at the information in order to provide the correct answer. Participants were instructed to answer the questions verbally as soon as they figured out the answer. In cases when participants failed to answer a question (e.g., when a technique did not work well for the participants), they were allowed to skip the question and proceed to the next one.

6.2.2 Walking task context. In the walking task, participants were asked to walk around in an indoor environment while following a virtual floating panel rendered in AR. The panel moved along a predefined path in the room with a constant speed. Based on its distance to the AR headset, the panel also changed color. Participants were asked to maintain a safe distance from the panel during the walking task (1.3-1.7 meters) so that the color of the panel stayed green (see Figure 4 (b)).

During walking, the AR display rendered four pieces of virtual content using the HG metaphor 0.4 meters away from the AR headset. The content was body-fixed. Each of the four pieces of virtual content was located in one direction of users' peripheral vision (i.e., up, down, left, right), which was approximately 26 degrees horizontally and 19 degrees vertically away from the central view. The reason that we chose four instead of six as in the sitting scenario was that walking is a more cognitively heavy task than sitting. Too much information at the periphery could limit the user's awareness of the surrounding environment. By reducing the amount of virtual content, we hoped to alleviate the problem of information overload and ensure a safe walking environment for participants.

Similar to the preliminary study, in the walking task, eight physical signs were distributed in the indoor environment so that the user would encounter two signs on their left, two on their right, two above, and two below (see Figure 4 (b)). During walking, the AR display asked questions verbally about information in the virtual content or the signs behind the content. Eight questions were asked in total for each trial (4 directions × 2 source (signs/virtual content)). Similar to the sitting task, when questions were asked about a physical sign in a particular direction, the virtual content would be rendered so that it overlaid the corresponding physical sign exactly. The signs were placed at depths of 1.2 - 2 meters depending on the user's speed and location during walking.

On average, each piece of activated content occupied around 13.75 degrees of visual angle horizontally and 7.35 degrees vertically, and each minimized icon occupied around 2.35 degrees horizontally and 2.24 degrees vertically on the display.

6.3 Apparatus

The study used a Magic Leap One AR HWD. The device has 1280 × 960 resolution with 50-degree diagonal FoV. The built-in head, hand, and binocular eye-tracking sensors of the Magic Leap were used to implement the activation techniques. The display of the Magic Leap is connected to an external battery/processing unit. For the walking task, participants wore a body strap to comfortably carry the battery pack with them during walking. To realize the HG interface metaphor, they also wore a waist strap with a 3D printed case to hold the Magic Leap controller to track the body torso during the walking task. The experimental software was developed via Unity 2019.3.7f1 with the SDK provided by Magic Leap.

6.4 Measures

Quantitative: We used the System Usability Scale (SUS) questionnaire [2], Social Acceptability Questionnaire [1] and NASA TLX workload questionnaire [13] to gauge the usability and workload of each interface. We also recorded the time it took for participants to answer the questions prompted by the computer screen/AR display. We calculated the time it took for participants to start verbally saying the answer after the question was displayed on the computer screen for the sitting task, and after a question audio finished playing in the AR headset in the walking task. We also recorded the percentage of false positives (i.e., false activation of the virtual content) when the intention was to access information in the real-world environment.

Qualitative: As for qualitative data, participants were asked to comment on what they liked or disliked about using each technique. For the walking task specifically, participants were video-recorded by a Logitech C930e camera mounted high on a wall for later analysis of walking behaviors while using the interfaces.

6.5 Hypotheses

We tested three hypotheses in the study:

- **H.1.** In both tasks, Fixation-Glance will be the most preferred, with faster information acquisition speed and lower workload as compared to other techniques, because our eyes naturally fixate in the direction and at the depth of objects that we are interested in, which will reduce the occurrences of conscious user input and inadvertent activations.
- **H.2.** Dwell will be the least preferred while accessing information in the real-world, because the user will have to obtain the information within the one-second dwell time to avoid inadvertent activation of the virtual content and occlusion of the real world.
- **H.3.** The Head-Depth technique will be the least favored technique in the walking condition, because it is challenging to lean backward accurately while walking at the same time.

6.6 Procedure

The experiment, which was approved by our university ethics board, was divided into six phases. In the first phase, participants were asked to read and sign the consent form. In the second phase, they were asked to fill out a pre-study questionnaire to collect demographic information and prior experience with AR. In the third phase, participants were given a detailed introduction to the experiment background, hardware, five interfaces, and the tasks involved in the study. When participants had no further questions, in the fourth phase, we helped participants to put on the AR HWD, and participants were asked to complete two calibration processes: (1) the fitting guide program of the Magic Leap One to determine the ideal size of the forehead-pad and nosepad; and (2) the visual calibration program of the Magic Leap One to ensure proper functioning of eye-tracking. Fifth, participants experienced each of the five conditions one by one, first in the sitting task, then in the walking task. The order of the five interfaces was counterbalanced using a Latin Square design. Before completing the experimental task in each condition, a training session was provided to get participants familiar with the interactions. After each interface condition, participants were asked to fill out the SUS and the NASA TLX workload questionnaires, and questions about what they liked or disliked about the interface. Each condition took about six minutes. Last, after finishing all five conditions in both sitting and walking scenarios, participants were asked to fill out a post-study questionnaire, in which we asked them to rank the interfaces based on their own preferences. The entire experiment took about 90 minutes. Participants were allowed to take a break anytime in between trials.

6.7 Results

We conducted a series of analyses to test our hypotheses and explore the trade-offs between the interfaces. We decided not to compare between the sitting and walking task because they involved different procedures and setup. As such, we separated the data based on the task, and used a one-way repeated-measures ANOVA (RM-ANOVA), with interface condition as the only independent variable for all the analyses. Shapiro-Wilk tests were applied to test the normality of the data. Friedman tests were applied when data failed the normality tests. A Greenhouse-Geisser correction was applied for violations of sphericity. For qualitative data gathered from questionnaires and recordings, Wilcoxon signed-rank tests were conducted. We applied Bonferroni corrections for all pair-wise comparisons. We used an α level of 0.05 in all significance tests. In the results figures, pairs that are significantly different are marked with * when $p \leq .05$, ** when $p \leq .01$, and *** when $p \leq .001$. For simplicity, we will use the abbreviations "DW", "BL", "FG", "HO", and "HD" for "Dwell", "Blink", "Fixation-Glance", "Hand-Overlay", and "Head-Depth" interfaces for the rest of the paper.

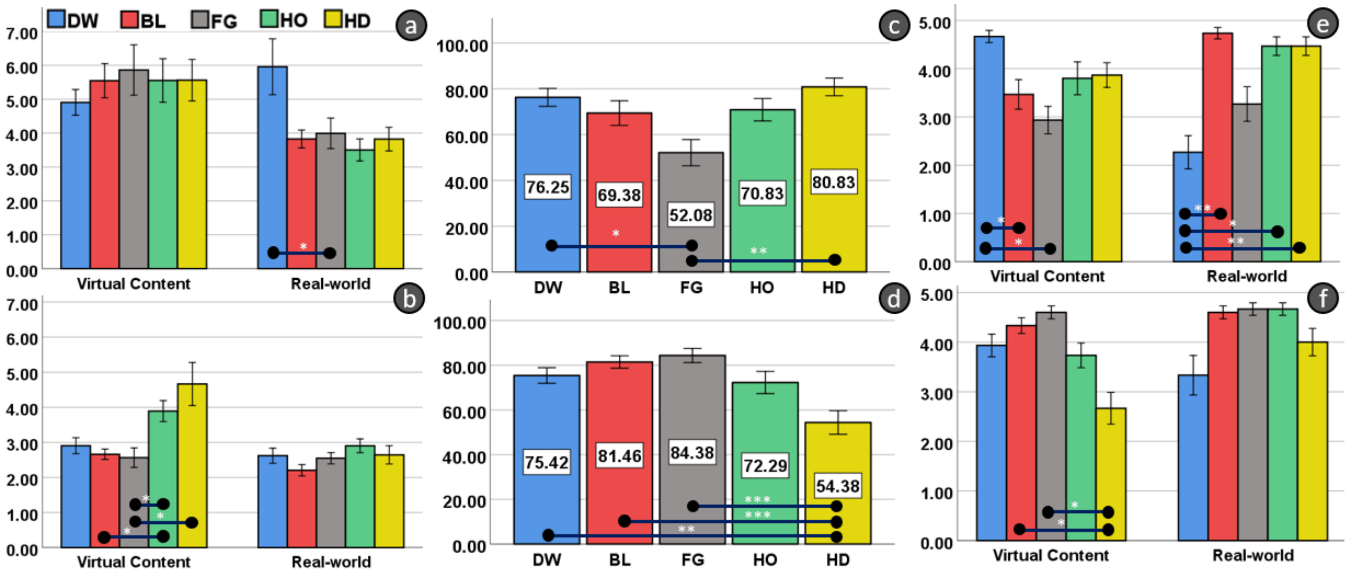

Figure 5: (a-b) The time it took for participants to answer the questions in the virtual content/real-world for the (a) sitting task, (b) walking task; (c-d) the SUS score for the five interface conditions in the (c) siting task and (d) walking task; and (e-f) Participants' response in 7-point Likert scale to the question "How do you like using the interface to access information in the virtual content/real-world" for the (e) sitting task and (f) walking task ($\pm S.E.$).

6.7.1 Task Performance. In each task, we collected data for 4 (number of questions asked) × 2 (source of content: either virtual content or real-world) × 5 (number of interface conditions) × 15 (number of participants) trials, yielding a total of 600 trials. We separated the data based on virtual-content and real-world and averaged the values for each participant in each condition, and used those averages in our analyses, leading to a total of 75 data points per source.

Figure 5 (a) shows the time it took for participants to answer the questions in each condition about each source in the sitting task. For accessing information in the virtual content, RM-ANOVA revealed no significant difference among interface conditions in the sitting task ($F_{(4,52)} = .454, p = .769$). As shown in Figure 5 (a), for accessing information in the real-world sticky notes, a significant main effect was found on interface conditions with a large effect size ($F_{(1.716,24.026)} = 6.936, p = .006, \eta_p^2 = .331$). Pairwise comparisons showed that DW (M = 5.96s, SD = 3.19) led to significantly longer time to access information in the sticky notes than FG (M = 3.99s, SD = 1.74) ($p = .041$). No other significant differences were found.

As shown in Figure 5 (b), for accessing information in the virtual content in the walking task, RM-ANOVA revealed a significant difference among interface conditions with a large effect size ($F_{(1.997,27.954)} = 8.711, p = .001, \eta_p^2 = .384$). Pairwise comparisons showed that FG (M = 2.56s, SD = 1.08) was significantly faster than HO (M = 3.89s, SD = 1.16) ($p = .032$) and HD (M = 4.66s, SD = 2.38) ($p = .011$). BL (M = 2.66s, SD = .57) was also significantly faster than HO ($p = .016$). For the time it took to access information in the real-world signs, RM-ANOVA found no significant differences among interfaces ($F_{(4,56)} = 1.590, p = .190$).

6.7.2 Overall usability. As shown in Figure 5 (c-d), we found significant main effects of interface conditions on the SUS score for both the sitting (($F_{(4,56)} = 5.040, p = .002, \eta_p^2 = .265$) and walking task (($F_{(4,56)} = 9.826, p < .001, \eta_p^2 = .412$). Pairwise comparisons show that FG (M = 52.08, SD = 22.15) received a significantly lower SUS score than DW (M = 76.25, SD = 15.16) ($p = .017$) and HD (M = 80.83, SD = 14.98) ($p = .008$) in the sitting task. In the walking task, HD (M = 54.39, SD = 20.38) received a significantly lower score as compared to DW (M = 75.42, SD = 13.51) ($p = .02$), BL (M = 81.46, SD = 10.73) ($p = .001$), and FG (M = 84.38, SD = 12.33) ($p < .001$).

6.7.3 User preferences. Figure 5 (e-f) shows participants' responses to the question "How do you like the interface for accessing information in the virtual content / real-world?" on a 5-point Likert scale. For virtual information acquisition, DW (M = 4.67) received the highest rating while FG (M = 2.93) received the lowest average rating in the sitting task. Friedman test shows significant differences among interfaces ($\chi^2(4) = 19.730, p = .001$). Wilcoxon signed-rank test showed that DW received significantly higher rating than BL (M = 3.47) ($Z = -3.025, p = .025$) and FG ($Z = -3.088, p = .020$) for accessing virtual information in the sitting task. As for the walking task, FG (M = 4.60) received the highest average rating for accessing information in the virtual content, while HD (M = 2.67) received the lowest average rating. Friedman test shows significant differences among interfaces ($\chi^2(4) = 22.673, p < .001$). HD interface received a significantly lower rating as compared to BL (M = 4.33) ($Z = -2.989, p = .028$) and FG ($Z = -3.190, p = .014$).

For accessing information in the real-world, BL (M = 4.73) received the highest ratings for information acquisition in the sitting task, while both FG and HO (M = 4.67) received the highest rating for the walking task. DW received the lowest rating in both sitting (M = 2.27) and walking (M = 3.33). For the sitting task, Friedman test found a significant main effect of interface on ratings ($\chi^2(4) = 27.983, p < .001$). DW received a significantly lower score than BL ($Z = -3.313, p = .009$), HO (M = 4.47)

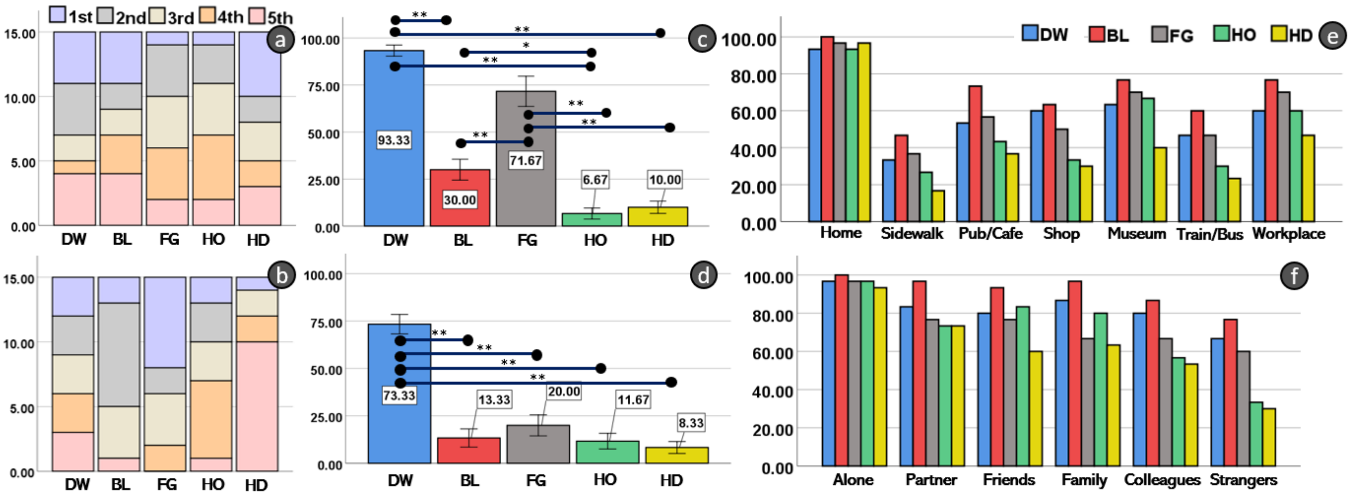

Figure 6: (a-b) The ranking of each interface under (a) sitting and (b) walking task; (c-d) The false activation rates of virtual content for (c) sitting task (d) walking task ($\pm S.E.$); and (e-f) social acceptance rate in percentage for (e) locations and (f) audiences.

($Z = -3.092, p = .020$), and HD (M = 4.47) ($Z = -3.352, p = .008$). Friedman test found a significant main effect of interface for the walking task ($\chi^2(4) = 14.373, p = .006$). However, no differences were identified in pairwise comparisons.

Figure 6 (a-b) shows participants' ranking of the five interfaces. For the sitting task, participants' preferences were very scattered. There was no clear tendency towards favoring or disliking a specific interface. In contrast, for the walking task, 8 participants (53.33%) ranked FG as the most favored interface, and 8 participants (53.33%) ranked BL as the second place. Meanwhile, 10 participants (66.67%) ranked HD as the least favored interface.

6.7.4 NASA TLX Workload. Our analysis revealed several significant differences in the NASA TLX workload ratings. For the sitting task, FG was considered significantly worse on Mental workload ($Z = -2.779, p = .050$) as compared to HD. For the walking task, HD was rated significantly worse than FG ($Z = -3.008, p = .026$) on Physical workload. HD was also rated worse on the Effort category than BL ($Z = -3.119, p = .018$) and FG ($Z = -2.817, p = .048$).

6.7.5 Percentage of false activations. We counted the percentage of false-positives (i.e., a piece of virtual content was activated falsely by participants when a question was asked about information in the real-world). As shown in Figure 6 (c), in the sitting condition, DW (M = 93.33%, SD = 11.44) and FG (M = 71.67%, SD = 31.15) resulted in the highest false activation rate, followed by BL (M = 30.00%, SD = 21.55), HD (M = 10.00%, SD = 12.67), and HO (M = 6.67%, SD = 11.44). For the walking task (Figure 6 (d)), DW (M = 73.33%, SD = 19.97) resulted in the highest false activation rate, followed by FG (M = 20.00%, SD = 16.90), BL (M = 13.33%, SD = 18.58), HO (M = 11.67%, SD = 16.00), and HD (M = 8.33%, SD = 12.20).

6.7.6 Social acceptability. Figure 6 (e-f) shows participants' responses towards the social acceptability questionnaire. In general, all interfaces can be accepted by almost all participants in private situations such as home environment or when alone. BL was considered the most acceptable to use in all locations and in front of all kinds of audiences. In contrast, HD and HO were considered less acceptable to use as compared to other interfaces.

6.7.7 Comments on interfaces. When asked what they like or dislike about each interface, FG was praised for being *fast, easy to use*, and *intuitive*. Participants commented: *I was able to activate and close widgets very easily*, and *the content just appeared naturally when I look there*. However, they also commented *more practice is needed for this technique, it would make me lose focus of my surrounding*, and *it was much harder to use in sitting than walking*. HO was praised for being *natural* and *intuitive*, but participants commented that *(I would be) worried I might bump into people*, and *it takes time to pull up my hand*. Participants liked DW because it is *easy and straightforward*. However, participants disliked it for always appearing and blocking the real-world objects: *the content always expands and block my vision... I had to look away to reset so that I could read the sticky notes*. Some participants also found it challenging to gaze at the fixed location while moving: *walking and dwelling on the widget simultaneously was hard*. As for BL, participants commented that *It was very easy to trigger the widgets... it requires little physical movements*. However, due to involuntary blinks, sometimes the content was triggered unintentionally: *the widgets sometimes appear without my intention*, and *getting the blinks to register sometimes is difficult in the sitting condition*. Participants liked HD because of its robustness: *the widgets never pop out unless I wanted them to*. However, participants disliked it for requiring too much physical movement. Participants especially considered it awkward and annoying to use in the walking task: *it required too much movement, I had to stop walking in order to look at the widgets, leaning forward and backwards was somewhat awkward, especially during walking*.

6.8 Discussion

In **H.1**, we hypothesized that FG would be the most favored interface with faster access speed and lower mental workload due to its natural and intuitive interaction. Our results partially supported this hypothesis by showing that most participants preferred FG over the other interfaces in the walking condition. It also resulted in lower physical workload and effort compared to HD. However, there was no clear tendency of favoring FG in sitting task. FG received higher SUS score in walking than sitting. In the sitting task, FG was considered more mentally difficult than HD. Participants' comments showed that FG was considered more challenging to use in sitting than walking, which was surprising to us since walking requires continuous attention to the surrounding environment. The false-positive rates in Figure 6 (c-d) of FG between sitting and walking were also highly different. After revisiting participants' comments and the playback, we surmise that the major reason was the different distance in the depth dimension between the virtual content and real-world. Due to limitations in eye-tracking technology, the gaze depth estimations could be inaccurate for some users. For the sitting task, the virtual content was only 0.3 meter away from the sticky notes, while for the walking task, the virtual content was 0.8-1.6 meters away from the signs. A larger distance allowed a larger safe region when the depth estimation was inaccurate or jittering.

Our results also supported **H.2** by showing that HD was the least favored when accessing virtual content during walking. HD received the highest SUS score in the sitting task, but it had the lowest SUS score and was ranked last by most participants in walking. Participants commented that they had to fully stop in order to perform the interaction, which reflected that it was hard to use during mobile situations.

Interestingly, a few participants also commented that BL was easier to use in the walking task than sitting. They mentioned *Sometimes I trigger the wrong widget when I blink in the sitting condition, which did not happen during walking*. We speculate that this was due to the difference in the number of pieces of virtual content. One issue with the BL technique was that the eye-tracking results would become inaccurate during blinks, which was also found in a recent study [28]. As such, if the safe region around virtual content was not big enough, blinks might be registered to the virtual content nearby. We had more pieces of virtual information that were much closer to each other in the sitting task. The safe region was smaller and more cluttered, so virtual content close to each other was more likely to be triggered falsely. As such, algorithms need to be applied to stabilize the gaze direction when a blink occurs to increase the scalability of the technique. Despite this problem, our results demonstrate great potential of BL. It received good SUS scores for both tasks, was ranked in the first/second place by most participants in the walking task, and received the highest ratings on almost all social acceptability categories.

Our results support our **H.3** by showing that DW was the lowest-rated technique for accessing information in the real-world. It was not robust given its highest false-positive rates in both task conditions.

On the contrary, HO and HD interfaces were very robust given their low false activation rates. However, they were also less accessible in the walking condition, and were considered less socially acceptable with longer access time and greater physical workload given the amount of physical movement required. During walking, FG and BL techniques could be a good balance between accessibility and robustness.

In general, our results did not show significant benefits of including depth as a input dimension. Techniques with depth input could still have high workload and low efficiency depending on the input modality being used.

7 DESIGN RECOMMENDATIONS

We distill the following design recommendations based on our findings:

- While using FG, it would be optimal to have the virtual content in a close depth layer and the real-world information in a far-away depth layer.
- When the goal is to optimize robustness (for example, when attention to the real-world environment is required), HO is a good option because they induce low false activation rates. HD is also a good option, but only in stationary scenarios.
- When the goal is to optimize accessibility of virtual content, FG and BL could be suitable interfaces to consider, since they have fast access times in scenarios where false activations are unlikely.
- When the scenarios of use are likely to change dynamically in terms of activity (staionary/mobile), depth difference between virtual and real information, and location (private/public), BL would be a good option because it achieves good performance with high acceptance across multiple scenarios.

8 LIMITATIONS & FUTURE WORK

There are a few limitations to our work. First, although the AR HWD we used in the study has multiple focal planes, vergence-accommodation conflict may still affect the results of using Fixation-Glance [36]. Second, we did not use machine-learning models to improve gaze depth estimations. Recent research highlighted such possibilities, which could make Fixation-Glance more usable in both tasks [23, 41]. Last, the study setups for the sitting and walking condition were not fully controlled. On the other hand, we attempted to make the task scenarios ecologically valid and representative of everyday scenarios. More controlled studies could be performed in the future to explore different tasks and contexts.

9 CONCLUSIONS

In the near future, AR glasses could support everyday information acquisition by displaying information to the users through the lens. However, virtual content could occlude users from seeing objects of interest in the real-world. In this research, we proposed and evaluated five techniques to help address such occlusion issue through explicit activation of virtual content in the context of Glanceable AR, a real-world prioritized information display metaphor. Our results demonstrated the trade-offs of gaze, hand, and head-based techniques in sitting and walking contexts. Our results could inspire future implementations of lightweight techniques for explicit activation of virtual content in AR HWDs.

ACKNOWLEDGMENTS

The authors would like to thank the study participants and anonymous reviewers. This work has been supported by the Office of Naval Research.

REFERENCES

[1] David Ahlström, Khalad Hasan, and Pourang Irani. 2014. Are you comfortable doing that? Acceptance studies of around-device gestures in and for public settings. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 193–202.

[2] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[3] Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+ gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 131–138.

[4] Shakiba Davari, Feiyu Lu, and Doug A Bowman. 2020. Occlusion Management Techniques for Everyday Glanceable AR Interfaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 324–330.

[5] Stephen Di Verdi, Daniel Nurmi, and Tobias Hollerer. 2003. ARWin-a desktop augmented reality window manager. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* IEEE, 298–299.

[6] Stephen DiVerdi, Tobias Hollerer, and Richard Schreyer. 2004. Level of detail interfaces. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 300–301.

[7] David Drascic and Paul Milgram. 1996. Perceptual issues in augmented reality. In *Stereoscopic displays and virtual reality systems III*, Vol. 2653. International Society for Optics and Photonics, 123–134.

[8] Steven K Feiner. 2002. Augmented reality: A new way of seeing. *Scientific American* 286, 4 (2002), 48–55.

[9] Alexandra Frischen, Andrew P Bayliss, and Steven P Tipper. 2007. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin* 133, 4 (2007), 694.

[10] Kristen Grauman, Margrit Betke, Jonathan Lombardi, James Gips, and Gary R Bradski. 2003. Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society* 2, 4 (2003), 359–373.

[11] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2016. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE transactions on visualization and computer graphics* 23, 6 (2016), 1706–1724.

[12] Chris Harrison and Anind K Dey. 2008. Lean and zoom: proximity-aware user interface and content magnification. In *Proceedings of the sigchi conference on human factors in computing systems*. 507–510.

[13] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[14] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A design space for gaze interaction on head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[15] Annette Horstmann and Klaus-Peter Hoffmann. 2005. Target selection in eye–hand coordination: Do we reach to where we look or do we look to where we reach? *Experimental Brain Research* 167, 2 (2005), 187–195.

[16] Robert JK Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 11–18.

[17] Roland S Johansson, Göran Westling, Anders Bäckström, and J Randall Flanagan. 2001. Eye–hand coordination in object manipulation. *Journal of Neuroscience* 21, 17 (2001), 6917–6932.

[18] Seungwon Kim, Mark Billinghurst, Gun Lee, and Weidong Huang. 2020. Gaze window: A new gaze interface showing relevant content close to the gaze point. *Journal of the Society for Information Display* 28, 12 (2020), 979–996.

[19] Pascal Knierim, Paweł W Woźniak, Yomna Abdelrahman, and Albrecht Schmidt. 2019. Exploring the potential of augmented reality in domestic environments. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[20] Sarah Krings, Enes Yigitbas, Ivan Jovanovikj, Stefan Sauer, and Gregor Engels. 2020. Development framework for context-aware augmented reality applications. In *Companion Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. 1–6.

[21] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. 2018. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[22] Wallace S Lages and Doug A Bowman. 2019. Walking with adaptive augmented reality workspaces: design and usage patterns. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 356–366.

[23] Youngho Lee, Choonsung Shin, Alexander Plopski, Yuta Itoh, Thammathip Piumsomboon, Arindam Dey, Gun Lee, Seungwon Kim, and Mark Billinghurst. 2017. Estimating gaze depth using multi-layer perceptron. In *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*. IEEE, 26–29.

[24] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 147–160.

[25] Weiyuan Liu. 2010. Natural user interface-next mainstream product user interface. In *2010 IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design 1*, Vol. 1. IEEE, 203–205.

[26] Feiyu Lu and Doug A Bowman. 2021. Evaluating the potential of Glanceable AR interfaces for authentic everyday uses. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 768–777.

[27] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A Bowman. 2020. Glanceable ar: Evaluating information access methods for head-worn augmented reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 930–939.

[28] Xueshi Lu, Difeng Yu, Hai-Ning Liang, Wenge Xu, Yuzheng Chen, Xiang Li, and Khalad Hasan. 2020. Exploration of Hands-free Text Entry Techniques For Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 344–349.

[29] Ann McNamara, Katherine Boyd, Joanne George, Weston Jones, Somyung Oh, and Annie Suther. 2019. Information placement in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1765–1769.

[30] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality. (2021).

[31] Hyung Min Park, Seok Han Lee, and Jong Soo Choi. 2008. Wearable augmented reality system using gaze interaction. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, 175–176.

[32] Thies Pfeiffer, Marc Erich Latoschik, and Ipke Wachsmuth. 2008. Evaluation of binocular eye trackers and algorithms for 3D gaze interaction in virtual reality environments. *JVRB-Journal of Virtual Reality and Broadcasting* 5, 16 (2008).

[33] Ken Pfeuffer, Yasmeen Abdrabou, Augusto Esteves, Radiah Rivu, Yomna Abdelrahman, Stefanie Meitner, Amr Saadi, and Florian Alt. 2021. ARtention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics* 95 (2021), 1–12.

[34] Robin Piening, Ken Pfeuffer, Augusto Esteves, Tim Mittermeier, Sarah Prange, Philippe Schröder, and Florian Alt. 2021. Looking for Info: Evaluation of Gaze Based Information Retrieval in Augmented Reality. In *IFIP Conference on Human-Computer Interaction*. Springer, 544–565.

[35] Radiah Rivu, Yasmeen Abdrabou, Ken Pfeuffer, Augusto Esteves, Stefanie Meitner, and Florian Alt. 2020. StARe: Gaze-Assisted Face-to-Face Communication in Augmented Reality. In *ACM Symposium on Eye Tracking Research and Applications*. 1–5.

[36] Dieter Schmalstieg and Tobias Hollerer. 2016. *Augmented reality: principles and practice*. Addison-Wesley Professional.

[37] Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y Chen. 2015. User-defined game input for smart glasses in public space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3327–3336.

[38] Leena Ventä-Olkkonen, Jonna Häkkilä, and Kaisa Väänänen-Vainio-Mattila. 2014. Exploring the augmented home window: user perceptions of the concept. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia*. 190–198.

[39] Roel Vertegaal et al. 2003. Attentive user interfaces. *Commun. ACM* 46, 3 (2003), 30–33.

[40] Roy Want. 2010. An introduction to ubiquitous computing. In *Ubiquitous computing fundamentals*. 1–36.

[41] Martin Weier, Thorsten Roth, André Hinkenjann, and Philipp Slusallek. 2018. Predicting the gaze depth in head-mounted displays using multiple feature regression. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 1–9.

[42] Wenge Xu, Hai-Ning Liang, Anqi He, and Zifan Wang. 2019. Pointing and selection methods for text entry in augmented reality head mounted displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 279–288.

[43] Difeng Yu, Hai-Ning Liang, Xueshi Lu, Tianyu Zhang, and Wenge Xu. 2019. Depthmove: Leveraging head motions in the depth dimension to interact with virtual reality head-worn displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 103–114.

[44] Özgür Yürür, Chi Harold Liu, Zhengguo Sheng, Victor CM Leung, Wilfrido Moreno, and Kin K Leung. 2014. Context-awareness for mobile sensing: A survey and future directions. *IEEE Communications Surveys & Tutorials* 18, 1 (2014), 68–93.