Exploring Gaze-assisted and Hand-based Region Selection in Augmented Reality

RONGKAI SHI, Xi'an Jiaotong-Liverpool University, China
YUSHI WEI, Xi'an Jiaotong-Liverpool University, China
XUEYING QIN, Shandong University, China
PAN HUI*, The Hong Kong University of Science and Technology (Guangzhou), China
HAI-NING LIANG†, Xi'an Jiaotong-Liverpool Univeristy, China

Region selection is a fundamental task in interactive systems. In 2D user interfaces, users typically use a rectangle selection tool to formulate a region using a mouse or touchpad. Region selection in 3D spaces, especially in Augmented Reality (AR) Head-Mounted Displays (HMDs) is different and challenging because users need to select an intended region via freehand mid-air gestures or eye-based actions that are touchless interactions. In this work, we aim to fill in the gap in the design of region selection techniques in AR HMDs. We first analyzed and discretized the interaction procedure of region selection and explored design possibilities for each step. We then developed four techniques for region selection in AR HMDs, which leveraged users' hand and gaze for unimodal or multimodal interaction. The techniques were evaluated via a user study with a controlled region selection task. The findings led to three design recommendations and two proof-of-concept application examples.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Mixed / augmented reality**; **Empirical studies in interaction design**.

Additional Key Words and Phrases: Augmented reality, Eye-tracking, Gaze interaction, Multimodal interaction, Region selection, Head-mounted display

ACM Reference Format:
Rongkai Shi, Yushi Wei, Xueying Qin, Pan Hui, and Hai-Ning Liang. 2023. Exploring Gaze-assisted and Hand-based Region Selection in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 7, ETRA, Article 160 (May 2023), 19 pages. https://doi.org/10.1145/3591129

1 INTRODUCTION

With recent advancements, Augmented Reality (AR) Head Mounted Displays (HMDs) have become powerful and portable tools. They project virtual objects on see-through displays, allowing users to see both the virtual objects and the physical environment around them simultaneously. Current AR HMDs support a rich set of input modalities. Before eye trackers are standard features, freehand

*Also with The Hong Kong University of Science and Technology, Hong Kong SAR, and University of Helsinki, Finland.
†Corresponding author.

Authors' addresses: Rongkai Shi, Xi'an Jiaotong-Liverpool University, Suzhou, China, rongkai.shi19@student.xjtlu.edu.cn; Yushi Wei, Xi'an Jiaotong-Liverpool University, Suzhou, China, yushi.wei21@student.xjtlu.edu.cn; Xueying Qin, Shandong University, Jinan, China, qxy@sdu.edu.cn; Pan Hui, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, panhui@ust.hk; Hai-Ning Liang, Xi'an Jiaotong-Liverpool Univeristy, Suzhou, China, haining.liang@xjtlu.edu.cn.

2573-0142/2023/5-ART160 $15.00
https://doi.org/10.1145/3591129

mid-air gestures are the primary means of interaction. With eye-tracking capabilities, gaze-based interaction is gaining increasing attention as eye gaze has been considered as a fast, natural, and unobtrusive interaction modality [17]. It could be a complement modality when the user's hands are not available for a task. Both mid-air and gaze-based interaction are becoming common. Some recent work has explored their use separately or together for multimodal interaction in a wide range of scenarios, such as target selection [21, 39, 58] and manipulation [26, 63], text entry [29, 33], collaborative work [18], and games [55].

Two-dimensional (2D) region selection is a fundamental interaction task in interactive systems. Unlike a typical target selection task, region selection involves a longer and more complex interaction procedure for determining the region of interest. In contrast to selecting a single object, selecting a region often requires users to visualize in their minds aspects that are not visible and update the image continuously as the process progresses. As physical and virtual contexts co-exist, 2D region selection in AR HMDs opens more potential use cases, such as retrieving virtual and/or physical objects located at a particular region or area [42], or setting 2D windows for presenting the virtual information in a physical workspace [7, 22].

Region selection is typically made via a rectangle selection tool on desktop computers [59]. Users press and hold the mouse button to draw a rectangular region. Some domain-specific software also provides selection tools using other shapes or the lasso tool for free-form selection.[1] Other common 2D input interfaces, such as touchpads or touchscreen, also use a similar rectangle selection strategy that provides tactile feedback (e.g., [4, 61, 65]). The haptic feedback from these tools and the support provided to users' hands allow region selection to be precise and relatively easy to do. On the other hand, for AR HMDs, a 2D region selection task can be challenging because the interaction is typically touchless, such as via mid-air gestures or gaze-based interaction. Furthermore, the relatively small Field-of-View (FoV) in current AR HMDs can also make the task challenging, especially when the intended region is larger than the FoV—that is, the ending point lies outside of the users' view. This issue can prevent users from having an efficient and well-formed plan that they could get if they can have the whole view of the intended region. To the best of our knowledge, limited research has investigated interaction techniques for region selection in AR HMDs, especially involving regions larger than the HMDs' FoV.

This research aims to fill this gap in the exploration of region selection techniques in AR HMDs. To this end and as the first step in this exploration, we investigated three potential interaction metaphors: (1) **mid-air hand-based interaction**, a common and device-free metaphor for current AR HMDs; (2) **gaze-based interaction**, given its fast, unobtrusive, and hands-free input; and (3) **multimodal interaction combining the first two unimodal metaphors**, which may potentially mitigate common problems in hand-only or eye-only interactions (e.g., arm fatigue or impressive/unstable interaction). We first identified and discretized the region selection process and formulated a set of design considerations based on each step. This led to the design of four potential region selection techniques (**Gaze-Finger**, **Gaze-Pinch**, **Pinch-Only**, and **Eye-Only**, as shown in Figure 1), which were developed in the HoloLens 2 and evaluated via a user study. Gaze-Finger and Gaze-Pinch were multimodal techniques, while Pinch-Only and Eye-Only were unimodal. In the study, we measured the performance of these four techniques via a controlled region selection task and collected participants' feedback from different aspects, including perceived workload, usability, fatigue, social acceptability, and preference.

[1]Existing examples of different selection tools can be seen from GIMP (an image editor, https://docs.gimp.org/2.10/en/gimp-tools-selection.html) and Adobe Photoshop (a graphic editor, https://helpx.adobe.com/photoshop/how-to/selection-tools-basics.html). Accessed: 15th-Apr-2023.

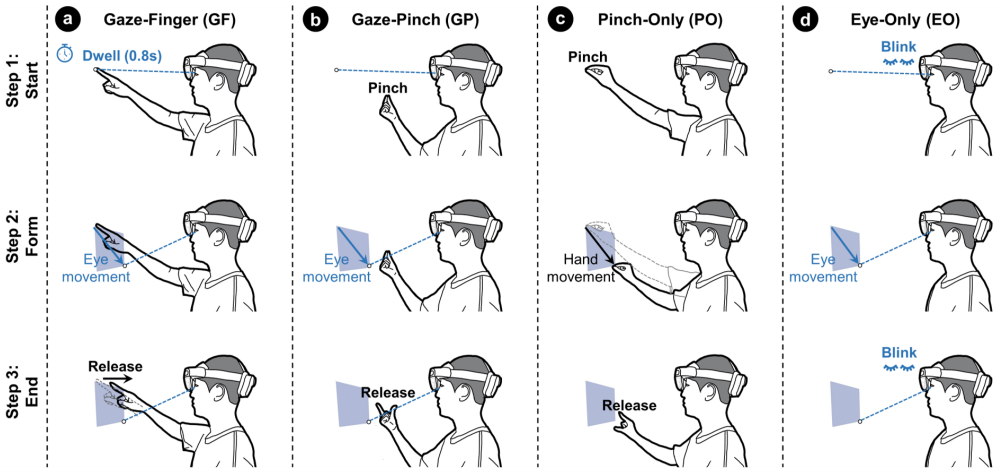

Fig. 1. 2D rectangular region selection in AR HMDs using four proposed techniques. (a) **Gaze-Finger**: dwell at the fingertip for 0.8s to start the region selection, remain the finger's depth distance and use eye gaze movement to form the region, and take the finger off to end the process. (b) **Gaze-Pinch**: perform a pinch gesture to start, hold the gesture, and use eye gaze movement to formulate the region, and un-pinch to end the process. (c) **Pinch-Only**: perform a pinch gesture to begin, hold it and select the region, and release the pinch gesture to end the process. (d) **Eye-Only**: use eye blink (0.2s) to indicate the start, use the gaze movement to determine the region, and perform another eye blink to end the process.

Our findings show that the two unimodal techniques (Gaze-Finger and Gaze-Pinch) outperformed the multimodal techniques (Pinch-Only and Eye-Only) in efficiency, accuracy, and users' subjective feedback. They also help us frame three design recommendations and choices for using these techniques in AR HMDs. Finally, we present two example applications based on the findings of the study. In short, the contributions of this work are:

- An evaluation of region selection strategies and interaction processes, and design considerations for developing gaze- and hand-based, or their combination for efficient and usable region selection techniques in AR HMDs.
- Design of four potential techniques for 2D region selection in AR HMDs.
- An evaluation of the four proposed techniques, leading to three recommendations for the choice and design of region selection for AR HMDs.

2 RELATED WORK

2.1 Region Selection in HMDs

Some target selection studies for Virtual Reality (VR) systems involve region selection as the first step of a multi-step task. Lucas [31] proposed a selection box and lasso tool to define three-dimensional (3D) volumes for selecting multiple objects in parallel in VR. Both techniques require users to define the intended region with a shape or volume constraint, which is similar to desktop selection tools [59]. We describe this strategy as *Draw a New Region*. Besides, the progressive refinement approach [20] for selecting targets in dense environments [49, 64] or distant targets [36] usually includes region selection as the first step to rearrange the objects in the region to then allow selecting the intended target with more ease and precision. This approach replaces the basic ray emitted from the controller with a cone or a spotlight [10, 11, 20, 25, 36, 49, 52, 64]. We describe this strategy as *Adjust a Predefined Region*. Instead of making a new region, users have a predefined

region (i.e., the cone) at hand. They need to adjust its size and place it in the intended position to cover the intended area.

Limited research has focused on region selection for AR HMDs. *SenseShapes* [40] is an early work to support object selection in AR HMDs via attached volumetric primitives, which used the *Adjust a Predefined Region* strategy. Wang et al. [57] implemented a lasso tool for spatial data selection in AR HMDs. However, as an extension tool of a PC setup, their users used mouse and keyboard input, which is neither common nor practical for AR HMDs. Lee et al. [23] proposed *TunnelSlice* for selecting 3D cuboid regions to support distant or occluded object acquisition in AR HMDs, which was based on the *Draw a New Region* strategy.

2.2 Gaze Interaction in HMDs

As stated by Tanriverdi and Jacob [54], the benefits of eye-based interactions in virtual environments include: (1) it requires less physical effort but increases interactivity, (2) it leverages users' natural eye behavior and pre-existing abilities, (3) it can be beneficial for interacting with distant objects, and (4) a pair of eye trackers only adds a little extra weight to the HMD. Many comparative studies have reported that eye-based interaction is faster than other input modalities, such as hand/controller input or head input, especially for pointing selection tasks [2, 21, 32, 38, 43, 54]. Given these benefits, researchers have applied gaze interaction to various scenarios in HMDs and utilized it in ways that benefit the task at hand [21, 29, 34, 37]. Eye gaze movement, as a continuous input signal, has been used to control operations [26], to draw a continuous path for determining candidate words in text entry [9], or to move the eye cursor for user authentication tasks [19]. On the other hand, the tasks that require discrete input, such as object selection [21, 39, 58], character selection for text entry [29, 30], and menu activation [34], used explicit eye fixations (i.e., dwelling on a point for a predefined period) or eye blinks instead.

Gaze interactions also suffer from several issues. One of the most prominent problems is the *Midas Touch* problem, which refers to the conflicting behaviors between the user's intentional actions and unintentional initiation of interaction [16]. One intuitive solution is to make the action deliberate. For example, the eye blink action for selection is designed to be different from spontaneous, natural blinking with specific requirements, such as multiple consecutive blinks [28] or longer eye-closed time intervals [30]. A dwell-based approach is commonly regarded as a solution to prevent the *Midas Touch* problem. However, a short dwell interval is still prone to the problem, while a long dwell time can induce eye fatigue and make the interaction inefficient [14, 15]. To better address this issue, more recent studies have combined gaze interaction with other modalities (e.g., [34, 63]), which we discuss more in the next section. Another problem of gaze interaction in HMDs is a lack of tracking precision and selection accuracy [13]. We envision the tracking technologies will progressively improve in the near future and more fine-tuning solutions will be available to improve stability and accuracy [8, 50] but until then we need to take into account the imprecise nature of gaze interaction.

2.3 Multimodal Interaction Combining Eye and Hand Input

Eye-hand coordination is a fundamental and natural motor control mechanism for interacting with the physical world. Eye-hand combination has been used for multimodal interactions because such a combination provides a richer and more efficient user experience. One domain application is gaze-supported selection. This combination has been implemented from mouse-based interaction for desktop systems [66] to mid-air gesture-based interaction for VR [35, 46, 47, 62] and AR systems [27, 33, 34]. Notably, Lystbaek et al. [34] introduced the *Gaze-Hand Alignment* principle, in which a selection event is triggered when the gaze and hand cursors are aligned. This concept is generally

affordable for AR HMDs and has been evaluated to be robust in menu selection [34] and character-based text entry tasks [33].

Researchers have also integrated gaze input into hand-based manipulation, following the principle of "gaze select, hands manipulate". This principle leverages the fast input from gaze movement for object indication and accurate input from hand gestures for object manipulation [5, 41, 51]. On the other hand, some researchers applied another strategy that uses hand input as trigger commands and eye gaze movement to control translations or positioning. In *GazeButton* [45], users tap and hold the button on the multitouch tablet device to activate the functions in text applications and use eye gaze to control the text cursor for highlighting or typing text. Turner et al. [56] developed *Gaze+RST* that utilized gaze and hand for concurrent manipulation tasks in multitouch displays. In *Gaze+RST*, the use of gaze is not limited to indicating and selecting the target object but also to controlling or assisting its translational movements. Similarly, Yu et al. [63] involved gaze input in the object manipulation process in VR HMDs.

3 DESIGN OF REGION SELECTION TECHNIQUES IN AR HMDS

In this section, we first analyze and discretize the interaction process of a 2D region selection task in AR HMDs and highlight the considerations in each step for designing possible solutions. Based on this, we propose four potential techniques for region selection in AR HMDs, including two multimodal techniques (Gaze-Finger and Gaze-Pinch), a mid-air hand-based technique (Pinch-Only), and a gaze-based technique (Eye-Only). Figure 1 demonstrates the interaction process and the proposed techniques.

3.1 Interaction Process

We summarized two interaction strategies for a region selection task in HMDs from the literature (see Section 2.1). Given that AR HMDs have a relatively restricted FoV, the intended region may have a larger height and/or width than the HMD's FoV. In this case, the *Adjust a Predefined Region* strategy is less workable as users cannot observe the whole region, as some parts could be out of sight, which makes the position and the size of the predefined region hard to determine. In contrast, the *Draw a New Region* strategy does not have this issue. It starts from a point in the region and because users are certain of its position, they can have it out of view to pursue the final part of the region. Thus, we focus on the *Draw a New Region* strategy. We assume at least one corner of the intended rectangular region is within the users' FoV. If this is not the case, users can always and easily navigate to find it. The rectangle selection in AR HMDs can be decomposed into three steps:

Start To select a 2D region, users first need to indicate a starting point, i.e., one of the four corners of the rectangular region. Then they perform an explicit trigger command to confirm the selection of the starting point. The trigger commands should be explicit and robust actions supported by the input modalities. For example, when only one modality is used, a time-based action (dwelling) may not be explicit, as they are not independent of the previous tracking state and may lead to the *Midas Touch* problem [16]. It is particularly risky in a region selection task because it does not involve a definite target object, and any points in the vision can be regarded as the starting point. Direct visual feedback should be given when the trigger command is executed to help users detect the transition between the states.

Form After selecting the starting point, users then need to draw the intended region. As we use a rectangular shape constraint, users only need to navigate toward the diagonal direction from the starting corner. During this process, the moving and starting points specify a rectangular region on the plane. This *Form* step represents a continuous input state. The generated region should be visualized in real-time to help users perceive the covered area. In addition, such a

state should be *kinesthetic* [53] when a gestural input is involved. Users should be able to hold the gesture so that they can get and remain aware of the forming state.

End Once the position and size of the region have been determined via the first two steps, users need to confirm to complete the region selection via another trigger command. The *End* step also requires *discrete input*, which shares the same considerations as the *Start* step.

One notable aspect is the depth distance where the above interaction process happens. We considered a 2D region selection in AR HMDs for general use, where the intended region would be invisible and conceived by users. To this end, there was no specific requirement on the depth of the region during the selection process. We tried to make the selection process of a region close to users within a suitable and comfortable depth for two reasons: (1) inherited from traditional 2D user interfaces, a front region would give a sense of covering further information; (2) it would be comfortable and efficient for users to interact at a depth within their arms' reach if a hand-based modality were involved. Such a front region would occlude users' hands. Thus, when a hand-based approach is used, we always render the hand mesh to ensure that users can notice the tracking states and the positions of their hands.

3.2 Techniques

3.2.1 Gaze-Finger (GF). GF is a multimodal technique inspired by the *Gaze-Hand Alignment* principle [34], as shown in Figure 1 (a). Users need to point to a corner of the intended region (the starting point) using their dominant hand's finger. They need to dwell on the fingertip for 0.8s to indicate the starting point [34]. We visualize the fingertip with a sphere to make the alignment easier. Then, they draw the intended region using their eye gaze movements. During this step, they need to maintain the lifted finger in a similar depth of the region. We set a 3cm depth range (1.5cm forward, and 1.5cm backward) to take into account users' unintentional movements. Once users move the eye cursor to the intended ending point, they confirm to complete by leaving the finger from the region's depth, or more often by withdrawing the finger naturally. The parameters were tested with six pilot users who confirmed their suitability.

3.2.2 Gaze-Pinch (GP). GP is also a multimodal technique that utilizes the eye-hand combination (see Figure 1 (b)). To start a region selection, users need to gaze at the intended starting point and perform a pinch gesture (pinch thumb and index finger together) to trigger the selection. They need to hold the pinch gesture and then move their eye gaze to the ending point to form the region. When the gaze reaches the ending point, they need to release the pinch gesture to confirm the selection of the ending point, which signals the conclusion of the process.

3.2.3 Pinch-Only (PO). PO is a unimodal technique that only uses hand gestures (see Figure 1 (c)). To perform a region selection, users need to pinch their thumb and index finger together, hold them and navigate to form the region, and release them to confirm the region. As users normally approach the index finger to the thumb for a pinch gesture, the thumb's position is more stable than the index's. This was also validated via the pilot users. Thus, we use the thumb's position for the formulation of the region.

3.2.4 Eye-Only (EO). This is also a unimodal but eye-based interaction approach (see Figure 1 (d)). Users first gaze at a corner of the intended region and blink their eyes to pin the starting point. Then they move their eye gaze to formulate the region. Once they position the eye cursor on the ending point, they blink again to confirm the completion of the region selection. Based on our pilot runs, a 0.2s eye-closed time can help avoid false positives induced by subconscious eye blinks. When designing the eye-only techniques, we also considered the *Gaze+Hold* technique proposed by Gomez et al. [44], which uses the closing of one eye for gaze input. However, at the time of

conducting this research, detecting eye-closing and opening events is not officially supported by eye-tracking-enabled AR HMDs, like HoloLens 2. More importantly, not everyone can perform wink gestures easily [30].

4 USER STUDY

In this study, we aimed to evaluate and compare the four proposed techniques via controlled experiments. We tested their performance in the given 2D region selection task in terms of speed and accuracy. In addition, we wanted to hear from participants about their subjective responses on perceived workload, usability, social acceptability, and preference for using the proposed techniques to complete the region selection tasks in AR HMDs.

4.1 Participants and Apparatus

We recruited 20 participants for this user study (5 females, 15 males). They were between 19 and 31 years old ($M = 22.70$, $SD = 2.99$). All of them were right-handed. Ten participants were near-sighted and wore glasses during the experiment. Fourteen reported no or little prior experience with AR HMDs. More than half reported no or little prior experience of using freehand gestures ($N = 12$) or eye gaze/blinks ($N = 14$) for interaction.

We used a Microsoft HoloLens 2 AR HMD, which has a 43°×29° FoV, a 60Hz refresh rate, and a 2K display resolution. In addition, HoloLens 2 enables 6 degrees-of-freedom eye tracking (with 1.5° visual angle accuracy) and hand tracking in real-time.[2] The program was developed using C# in Unity (version 2020.3.47f1) with Mixed Reality Toolkit (MRTK, version 2.8.2) and Mixed Reality OpenXR Plugin (version 1.5.1). Participants completed the experiments in a sitting position in a quiet room.

4.2 Study Design and Task

We used a 4×2 within-subjects design with Technique (GF, GP, PO, and EO) and Difficulty (Simple and Difficult) as the two independent variables. The task required participants to select a rectangular region (see Figure 2). As mentioned, users formulate a region based on their intentions in real use cases but for experimental purposes, the application would show the target region in the AR HMD for participants to select. The target region was presented in translucent blue and placed 0.5m away from the participants. In a Simple task, the target region was a 20°×15° or 15°×20° rectangle (approx. 145cm×105cm, or the reverse), and was randomly placed within the headset's FoV. In a Difficult task, the target region was a 45°×30° or 30°×45° rectangle (approx. 330cm×215cm, or the reverse), which was greater than the headset's FoV. We made one corner of the target region presented within the FoV for a Difficult task. This would ensure that the visual search process of the starting point, which could be a confounding factor, was not part of the task. Note that we used Difficulty but did not separate region size and region position relative to FoV as two independent variables because users could always fit a small target that was out of vision within the FoV, and a large target is at least partially out of vision.

We placed a "start button" at the center of the participants' view. The button was only shown between two trials and mainly for two purposes: (1) to allow participants to control the start of a trial, (2) to make participants move back to the initial position after each trial, which would guarantee a reset. Moreover, we placed this button at a 0.3m distance (closer than the target region), set a 1s pressing time, and encouraged using their non-dominant hand for activation to avoid false positives.

[2]Please refer to https://learn.microsoft.com/en-us/HoloLens/HoloLens2-hardware for other device specifications. Accessed: 15th-Apr-2023.

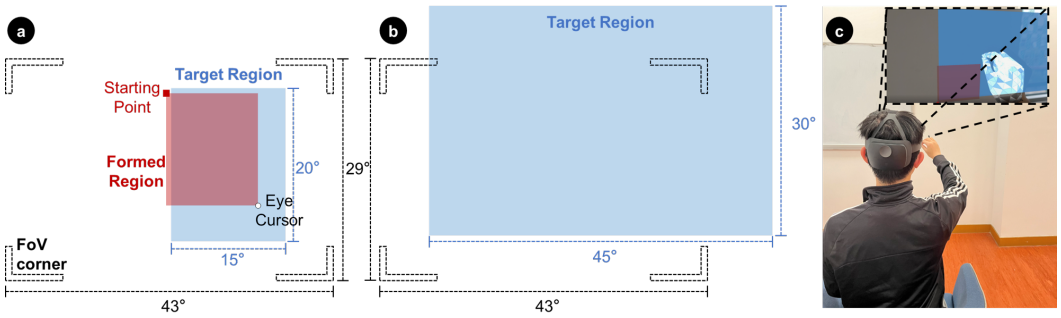

Fig. 2. Illustrations for the task (dotted lines and texts are for illustration only and were not shown to participants). (a) A Simple task, where the target region (15°×20°) is within FoV. (b) A Difficult task, where part of the target region (45°×30°) is out of FoV. (c) A participant completing a region selection task and a view of the AR HMD during a trial.

To minimize any learning effect, the order of the Technique conditions was counterbalanced with a Latin-Squared design, and in each Technique condition, the order of the Difficulty condition was randomized. Participants would complete ten trials in each condition, which led to a total of 1600 trials of data (=20 participants × 4 region selection techniques × 2 levels of task difficulty × 10 repetitions).

4.3 Evaluation Metrics

4.3.1 Objective Measurements. In each condition, we calculated the distance between the participants' triggered starting point and the corner of the target region that was the closest to it (denoted as DistanceS), and the triggered ending point to the corner that was the closest to it (denoted as DistanceE). If the two corners in these two measurements were not diagonal, we marked this trial as a failed trial. To better present the data collected from a 2D plane, we used the Cartesian coordinate system to record the distance rather than the angular system. In terms of time, we measured start time, navigation time, and total time. Start time was when the participants triggered the start button until they triggered the starting point, indicating the time consumed to complete the *Start* step. Navigation time would start counting immediately after, until they completed the region selection for the trial, representing the time for the *Form* and *End* steps. Total time was the sum of both as the total time spent for the region selection task in each trial.

4.3.2 Subjective Measurements. We used the raw NASA-TLX questionnaire [12], positive version of System Usability Scale (SUS) [24], Social Acceptability Questionnaire [1], and Borg CR10 questionnaires [3] to measure perceived workload, usability, social acceptance, and exertion/fatigue when using the techniques for AR region selection. Similar to the approach described by Lystbæk et al. [33], we used two Borg CR10 questionnaires to measure arm and eye fatigue separately. At the end of the experiment, we also asked participants to rank all four techniques according to their preferences and conducted a short interview asking for their comments on the techniques.

4.4 Procedure

The experiment was divided into four phases. First, participants were asked to fill out a pre-experiment questionnaire and were briefed about the AR HMD, its controls, tasks, experimental conditions, and procedures in this study. Second, after signing a consent form, participants wore the HMD and calibrated the eye tracker. Third, they went through each condition. Before the formal

trials, they were required to have a fixed 2-minute training to get familiar with the techniques. After each Technique condition, they filled out the subjective questionnaires as mentioned in the previous section. Last, they received a post-session interview about their experience and comments after they completed the tasks in all conditions. They were not allowed to take off the HMD before the end of the interview. The whole study lasted about 40 minutes for each participant.

4.5 Hypotheses

We tested four hypotheses (denoted by **H#**) in this study:

- **H1.** GF would achieve the highest region selection accuracy but lead to a lower speed because of the dwell time.
- **H2.** The perceived workload to complete the task would not differ among the four techniques. As all four techniques would apply a common region selection strategy and were expected to be used cost- and effort-effective.
- **H3.** Given the unimodal utilities, PO would induce the heaviest arm fatigue but lead to the least eye fatigue, while EO would be the reverse.
- **H4.** EO would receive the highest social acceptance among the four techniques in public places. Gestural input may lead to lower social acceptability ratings, especially when used in front of strangers.

5 RESULTS

5.1 Objective Measurements

In total, we identified 110 failed trials (6.88%). The reasons why participants failed to complete these trials varied (e.g., false positives, delay due to the device, or distractions). Thus, the analyses were based on the 1490 successful trials. We first removed outliers (51 trials, approx. 3.42% of the used trials) from the data where any of the measurements exceeded $M \pm 4SD$ in each condition. This helps limit the influence of tracking issues by the device and the participants' excessive concerns about speed or accuracy when completing the trials. All the performance measures were not normally distributed based on the results from Shapiro-Wilk tests ($p < .05$) and their Q-Q plots. Therefore, we applied Aligned-Rank Transformation [6, 60] to them before conducting two-way repeated-measure (RM-) ANOVA tests. Effect size was reported using partial eta squared (η_p^2). Pairwise comparisons were conducted with Bonferroni corrections if a significant difference was found. The performance results are summarized in Figure 3.

5.1.1 Start Time. Results from RM-ANOVA tests revealed both Technique ($F_{3,1412} = 203.887, p < .001, \eta_p^2 = .302$) and Difficulty ($F_{1,1412} = 51.321, p < .001, \eta_p^2 = .035$) had significant main effects on start time. We also found a significant interaction effect between Technique×Difficulty ($F_{3,1412} = 2.854, p = .036, \eta_p^2 = .006$) on start time. Post-hoc tests showed that GF ($M = 3.33s, SD = 1.10$) required a significantly longer start time compared to GP ($M = 2.36s, SD = 1.22$), PO ($M = 2.08s, SD = 0.93$), and EO ($M = 2.20s, SD = 1.02$) in Simple tasks (all $p < .001$). For Difficult tasks, start time for GF ($M = 4.08s, SD = 1.91$) was also significantly longer than GP ($M = 2.74s, SD = 1.40$), PO ($M = 2.23s, SD = 0.89$), and EO ($M = 2.34s, SD = 0.93$) (all $p < .001$). In addition, GP led to a significantly longer start time than PO ($p < .001$) and EO ($p = .022$). Besides, participants took significantly longer time to complete Difficult tasks than Simple tasks using GP ($p < .001$).

5.1.2 Navigation Time. We found a significant main effect of Technique ($F_{3,1412} = 13.861, p < .001, \eta_p^2 = .029$), a significant main effect of Difficulty ($F_{1,1412} = 141.558, p < .001, \eta_p^2 = .091$), and a significant interaction effect ($F_{3,1412} = 4.135, p = .006, \eta_p^2 = .009$) for navigation time. For Simple tasks, PO ($M = 2.77s, SD = 1.11$) required a significantly shorter navigation time compared

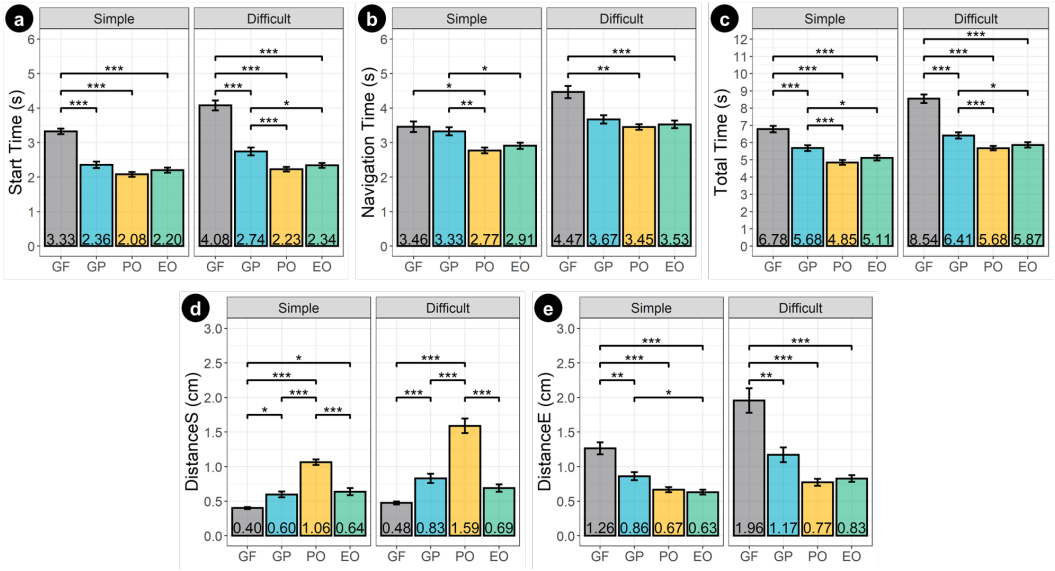

Fig. 3. Plots of objective measurements. (a-c) The time used to complete a region selection task: (a) start time, (b) navigation time, (c) total time. (d-e) The distance between participants' triggered points and their closest corner of the target region: (d) distance for the starting points (DistanceS), (e) distance for the ending points (DistanceE). The error bars represent standard errors. Significant differences in post-hoc tests are marked with *, **, and ***, representing a significance level of .05, .01, and .001 (Bonferroni-adjusted), respectively. The same scheme is used for the next figure also.

to GF ($M = 3.46s, SD = 2.13; p = .025$) and GP ($M = 3.33s, SD = 1.60; p = .004$); and navigation time by using EO ($M = 2.91s, SD = 1.23$) was shorter than GP ($p = 0.025$). For Difficult tasks, GF ($M = 4.47s, SD = 2.37$) required longer navigation time than PO ($M = 3.45s, SD = 1.10; p = .005$) and EO ($M = 3.53s, SD = 1.51; p < .001$). Besides, participants took significantly longer navigation time to complete Difficult tasks than Simple tasks regardless of the technique used (GF: $p < .001$, GP: $p = .002$, PO: $p < .001$, and EO: $p < .001$).

5.1.3 Total Time. There was a significant main effect of Technique ($F_{3,1412} = 127.017, p < .001, \eta_p^2 = .213$), a significant main effect of Difficulty ($F_{1,1412} = 151.799, p < .001, \eta_p^2 = .097$), and a significant interaction effect ($F_{3,1412} = 5.075, p = .002, \eta_p^2 = .011$) for total time. For Simple tasks, GF ($M = 6.78s, SD = 2.56$) led to a significantly longer total completion time compared to GP ($M = 5.68s, SD = 2.26$), PO ($M = 4.85s, SD = 1.85$), and EO ($M = 5.11s, SD = 1.94$) (all $p < .001$). On the other hand, GP required a longer time than PO ($p < .001$) and EO ($p = .013$). Difficult tasks had similar results. GF ($M = 8.54s, SD = 3.28$) led to a significantly longer total completion time compared to GP ($M = 6.41s, SD = 2.25s$), PO ($M = 5.68s, SD = 1.73$), and EO ($M = 5.86s, SD = 2.09$) (all $p < .001$) in Difficult tasks. In addition, GP required a longer time than PO ($p < .001$) and EO ($p = .011$). Same as navigation time, participants took significantly longer navigation time to complete Difficult tasks than Simple tasks regardless of the used techniques ($p < .001$ for all four techniques).

5.1.4 DistanceS. Results from RM-ANOVA tests revealed both Technique ($F_{3,1412} = 142.05, p < .001, \eta_p^2 = .232$) and Difficulty ($F_{1,1412} = 54.625, p < .001, \eta_p^2 = .037$) had significant main effects on DistanceS and revealed a significant interaction effect ($F_{3,1412} = 5.871, p < .001, \eta_p^2 = .012$).

For Simple tasks, PO ($M = 1.06cm, SD = 0.55$) had a significantly longer DistanceS compared to GF ($M = 0.40cm, SD = 0.21$), GP ($M = 0.60cm, SD = 0.56$), and EO ($M = 0.64cm, SD = 0.70$) (all $p < .001$). In addition, GP and EO both had a significantly longer DistanceS than GF ($p = .032$ and $p = .014$). For Difficult tasks, PO ($M = 1.59cm, SD = 1.41$) had a significantly longer DistanceS compared to GF ($M = 0.48cm, SD = 0.29$), GP ($M = 0.83cm, SD = 0.84$), and EO ($M = 0.69cm, SD = 0.72$) (all $p < .001$). And GP also had a significantly longer DistanceS than GF ($p < .001$). We also found a significantly longer DistanceS in Difficult tasks than in Simple tasks when using GP ($p = .003$).

5.1.5 DistanceE. There was a significant main effect of Technique ($F_{3,1412} = 39.819, p < .001, \eta_p^2 = .078$), a significant main effect of Difficulty ($F_{1,1412} = 36.357, p < .001, \eta_p^2 = .025$), and a significant interaction effect ($F_{3,1412} = 3.578, p = .013, \eta_p^2 = .008$) for DistanceE. For Simple tasks, GF ($M = 1.26cm, SD = 1.20$) led to a significantly longer DistanceE than GP ($M = 0.86cm, SD = 0.78; p = .003$), PO ($M = 0.67cm, SD = 0.50; p < .001$), and EO ($M = 0.63cm, SD = 0.49; p < .001$). In addition, GP had a longer DistanceE than EO ($p = 0.049$). For Difficult tasks, GF ($M = 1.96cm, SD = 2.37$) led to a significantly longer DistanceE than GP ($M = 1.17cm, SD = 1.34; p = .001$), PO ($M = 0.77cm, SD = 0.67; p < .001$), and EO ($M = 0.83cm, SD = 0.66; p < .001$). Compared between the difficulties, DistanceE in Difficult tasks was longer than Simple tasks for EO ($p = .026$).

5.2 Subjective Measurements

We applied non-parametric Friedman tests to NASA-TLX scores, SUS scores, and two types of Borg CR10 scores, and reported the effect sizes whenever feasible (Kendall's W). Cochran's Q test was applied to social acceptance rates due to its binary responses. Pairwise comparisons were also conducted with Bonferroni corrections. We mainly report results with significant differences here. Please refer to our supplementary materials for detailed reports.

5.2.1 NASA-TLX Workload. The results of NASA-TLX scores are summarized in Figure 4 (a). Friedman tests revealed significant main effects for NASA-TLX scores in Mental workload ($\chi_3^2 = 8.93, p = .030, W = .149$) and Physical workload ($\chi_3^2 = 7.95, p = .047, W = .132$). However, no significant differences were found in pairwise comparisons.

5.2.2 Overall Usability. Our analysis revealed significant main effects in SUS scores ($\chi_3^2 = 17.3, p < .001, W = .289$). Post-hoc analyses showed PO ($Mdn = 71.2$) received significantly higher SUS scores than GF ($Mdn = 52.5; p = .023$) and GP ($Mdn = 56.2; p = .014$), as shown in Figure 4 (b).

5.2.3 Arm and Eye Exertion/Fatigue. Figure 4 (c-d) show the results of Borg CR 10 scores for arm and eye fatigue, respectively. We found a significant main effect in arm fatigue ($\chi_3^2 = 34.1, p < .001, W = .568$) and a significant main effect in eye fatigue ($\chi_3^2 = 12.3, p = .006, W = .206$). EO did not induce arm fatigue and thus led to significantly lower perceived arm fatigue than GF ($Mdn = 3; p = .001$), GP ($Mdn = 3; p = .001$), and PO ($Mdn = 3; p = .002$). On the other hand, in terms of eye fatigue, PO ($Mdn = 1$) induced significantly lower eye fatigue than GP ($Mdn = 2.5; p = .018$) and EO ($Mdn = 3; p = .049$).

5.2.4 Social Acceptance. Overall, PO and EO were considered more acceptable than GF and GP in public locations or in front of strangers. In terms of locations, Cochran's Q tests showed significant main effects of Technique in acceptance rate on a sidewalk ($\chi_3^2 = 14.9, p = .002$), pub/café ($\chi_3^2 = 21.7, p < .001$), shop ($\chi_3^2 = 18.8, p < .001$), museum ($\chi_3^2 = 12.7, p = .005$), train/bus ($\chi_3^2 = 17.5, p < .001$), and workplace ($\chi_3^2 = 9.43, p = .024$). In terms of audiences, Cochran's Q tests showed significant main effects of Technique on acceptance rate in front of colleagues

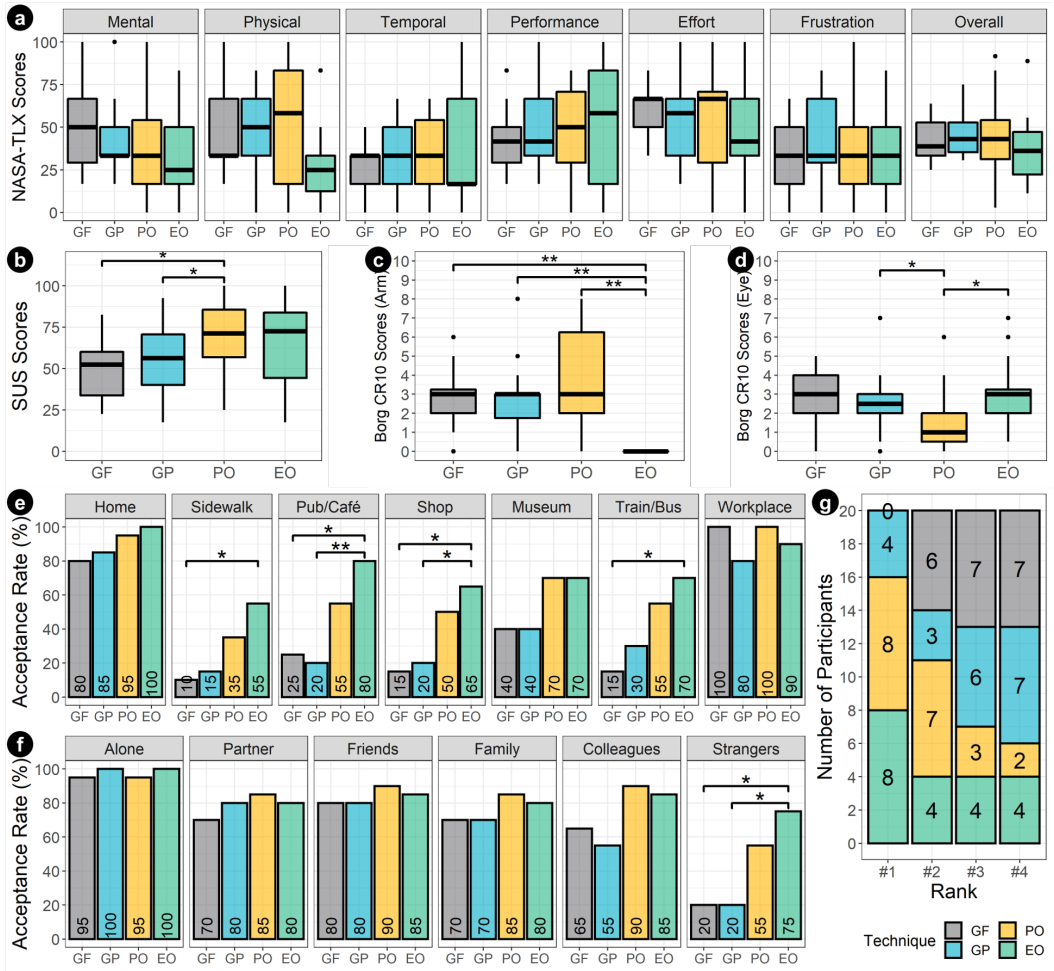

Fig. 4. Plots of subjective measurements. (a) NASA-TLX scores, (b) SUS scores, (c-d) Borg CR10 scores for (c) arm, and (d) eye exertion/fatigue. (e-f) Social acceptance rates in percentage for (e) locations, and (f) audiences. (g) Users' rankings of the four proposed techniques in terms of their overall preference.

($\chi^2_3 = 12.7, p = .005$) and strangers ($\chi^2_3 = 21.4, p < .001$). Figure 4 (e-f) summarize the acceptance rates in percentage and the significant results from the post-hoc McNemar tests.

5.2.5 Ranking. Figure 4 (g) shows the users' rankings of the techniques based on their preferences. There was a tendency towards favoring the unimodal techniques, with PO slightly more favorable than EO (in Rank#2). In contrast, the multimodal techniques—GF and GP, were not preferred. Considering the first and the second places, the preference of the techniques was PO (15 participants, 75%) > EO (12 participants, 60%) > GP (7 participants, 35%) > GF (6 participants, 30%). Participants' comments on the techniques are presented and discussed in the Discussion section as complementary insights to the above results.

6 DISCUSSION

6.1 Technique Evaluation

We found the two unimodal techniques (PO and EO) were faster than the two multimodal techniques (GF and GP) in the region selection tasks. GF had a significantly longer start time mainly because of its 0.8s eye fixation time for activation. On the other hand, it benefited to make the selection of the starting point more accurate (see Figure 3 (d)). The hand-based technique PO led to a significantly larger distance between the starting points than the other three techniques. During the trials, we observed that almost all participants had a preference for selecting the target region from the left-top corner to the right-bottom corner of the rectangular region (though we did not ask them to). Even if the left-top corner was not the closest one to the eye cursor or hand or was even out of vision in a Difficult task, participants still preferred to complete the task this way. PO required the pinch gesture to reach the location, which can be more error-prone for a distant starting point. For GP and EO, they used an eye cursor to locate the starting point, which was less affected. While for GF, the dwell-based alignment technique helped to improve accuracy, which supported **H1**. We found the multimodal techniques had a longer navigation time as well. This may be due to the lost hand tracking issue from the HMD. Based on our observation and participants' comments, they tended to perform the gesture in a lower but comfortable position but had to raise their hands to recover the tracking. This issue not only increased the navigation time but also left a longer distance between ending points (as the eyes involuntarily coordinated with the moving hand).

We did not find significant differences in the perceived workload of using the four techniques to complete the tasks; thus, **H2** was confirmed. As all four techniques used the same strategy for completion and a region selection task is common and fundamental in daily interaction with interactive systems, the workloads did not vary. **H3** was refuted—the results did not reveal any significant differences in perceived arm fatigue between PO and GF or between PO and GP, though we expected the multimodal techniques could help reduce arm fatigue to some extent. The reason can be twofold: (1) participants had to raise their hands to ensure the hand gestures were detected by the AR HMD when using GF and GP, and (2) a longer total time spent with GF and GP also increased fatigue. We hypothesize that if GF and GP could be captured by the HMD from a wider space in the future and users could thus perform the gestures in a more comfortable way, the perceived arm fatigue of using them would decrease. On the other hand, we found the perceived eye fatigue for EO was only significantly higher than PO but not GF and GP. This means the eye-based interactions for region selection would not bring extra eye fatigue compared to multimodal techniques with eye and hand.

For the social acceptability of the techniques, EO was more acceptable in public places and in front of strangers, which supported **H4**. EO is based on non-observable, hidden eye interaction, but all the other three involve explicit gestural interaction visible to others around the users, which can be more sensitive and uncomfortable to perform in public places. We found that PO's acceptance rates were higher than GF and GP. One possible reason is that PO involves continuous movements, which are more expressive and probably more understandable by other people. In contrast, hand gestures in GF and GP are for discrete input, which can be sudden, especially GF, as users may not want to point to someone in public places (something very impolite to do in many cultures).

The results from users' preferences are in line with the SUS scores. Participants preferred the unimodal techniques and thought they were more useful than the multimodal techniques. In the interviews, participants felt the two unimodal techniques were "*intuitive*", and "*similar to the existing region selection operations*". P7 commented "*I feel the Gaze-Finger and Gaze-Pinch are complex to use, the other two are more straightforward for region selection*".

6.2 Design Recommendations

Our findings allow us to distill three design recommendations (**DR#**) for choosing or designing AR region selection techniques.

DR1 When users' hands are available, a PO type technique is a suitable choice given that it is fast and preferred by participants. It can also be easily integrated with the workflow of other interaction processes.

DR2 When users' hands are not available, an EO type technique can be used. It allows fast and accurate selection without additional effort. It is also suitable when the AR HMD is used in a public place.

DR3 If an accurate starting point is essential, a GF technique type is a suitable choice but designers may need to consider improving how the selection ends.

6.3 Limitations and Future Work

We have identified the following limitations in this work which can serve as future research directions. First, the task we used was not perfectly controlled since the target region was generated randomly to avoid learning effects. More controlled studies could be conducted in the future to explore different tasks. Second, the study results were affected by the technique specifications. Though we conducted pilot trials to estimate the parameters, there is still space to optimize them and improve their usability. Last, as this work represents the first explorations of the topic, we only proposed and evaluated four interaction techniques. We plan to explore more possibilities in the future and evaluate their performance and user experience, especially when they are integrated into the workflow of other complex tasks (e.g., selection and interacting with groups of objects [48, 64]. We also plan to adapt the proposed techniques to region selection with other regular/non-regular or free-form shapes and extend the current region selection task with fine-grained processes.

7 EXAMPLE APPLICATIONS

In this section, we show two mock-up AR application scenarios based on our findings. The first application demonstrates a use case for retrieving information to check the main categories of books stacked on bookshelves in a library (see Figure 5 (a)). Selecting a region that contains the bookshelves in view, they can view information returned from a search. In this case, because the users have their hands occupied carrying some books or other personal belongings, they can be provided with the EO like technique (**DR2**) and perform the region selection without using their hands. The second scenario shows selecting a region as an image to be shared (see Figure 5 (b)). In this scenario, the user selects the region of interest to make an image for the newly brought item and shares the captured image with friends via a social media app. Given that the user's hands are available, she is given the PO type technique (**DR1**) and uses it to select the region. This scenario shows a good use for such PO technique given the continuous interaction flow and its integration with another related, follow-up task.

8 CONCLUSION

Region selection has broad use scenarios in Augmented Reality (AR) Head-Mounted Displays (HMDs). However, limited research has explored this topic to design and evaluate suitable techniques for the 2D region selection tasks. In this research, we proposed four techniques, including hand-based (Pinch-Only), eye-based (Eye-Only), and two multimodal techniques that leveraged the combined use of the hand and eye (Gaze-Finger and Gaze-Pinch). They were evaluated in a user study with region selection tasks with two levels of difficulties. Our results showed that the two unimodal techniques outperformed the multimodal techniques. Our results also led to three

Fig. 5. Two example application scenarios with region selection via an AR HMD. (a) Information retrieval for the book categories in a library. (b) Capturing an interesting part of a user's view and sending the captured part to another application.

recommendations for the choice and design of region selection techniques in AR HMDs. This work serves as an initial exploration of techniques for 2D region selection tasks in AR HMDs and can help frame future work.

ACKNOWLEDGMENTS

We thank the participants who joined the user study and the reviewers for their insightful comments that helped improve our paper. This research was partly funded by Xi'an Jiaotong-Liverpool University Special Key Fund (#KSF-A-03), the National Science Foundation of China (#62272396), and the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (#SZS2022004).

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* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 193–202. https://doi.org/10.1145/2628363.2628381

[2] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of Eye-Gaze over Head-Gaze-Based Selection in Virtual and Augmented Reality under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3206343.3206349

[3] Gunnar Borg. 1982. Psychophysical bases of perceived exertion. *Medicine and science in sports and exercise* 14, 5 (1982), 377–381.

[4] Peter Brandl, Clifton Forlines, Daniel Wigdor, Michael Haller, and Chia Shen. 2008. Combining and Measuring the Benefits of Bimanual Pen and Direct-Touch Interaction on Horizontal Interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Napoli, Italy) *(AVI '08)*. Association for Computing Machinery, New York, NY, USA, 154–161. https://doi.org/10.1145/1385569.1385595

[5] 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* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2818346.2820752

[6] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784

[7] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the World: 2D Windows for 3D Augmented Reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*

(Atlanta, Georgia, USA) *(UIST '93)*. Association for Computing Machinery, New York, NY, USA, 145–155. https://doi.org/10.1145/168642.168657

[8] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1118–1130. https://doi.org/10.1145/3025453.3025599

[9] Wenxin Feng, Jiangnan Zou, Andrew Kurauchi, Carlos H Morimoto, and Margrit Betke. 2021. HGaze Typing: Head-Gesture Assisted Gaze Typing. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. https://doi.org/10.1145/3448017.3457379

[10] Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. 1996. Aperture Based Selection for Immersive Virtual Environments. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 95–96. https://doi.org/10.1145/237091.237105

[11] Gerwin de Haan, Michal Koutek, and Frits H. Post. 2005. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In *Eurographics Symposium on Virtual Environments*, Erik Kjems and Roland Blach (Eds.). The Eurographics Association, 201–209. https://doi.org/10.2312/EGVE/IPT_EGVE2005/201-209

[12] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[13] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

[14] Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze Gestures or Dwell-Based Interaction?. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/2168556.2168602

[15] Howell Istance, Richard Bates, Aulikki Hyrskykari, and Stephen Vickers. 2008. Snap Clutch, a Moded Approach to Solving the Midas Touch Problem. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 221–228. https://doi.org/10.1145/1344471.1344523

[16] Robert J. K. 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* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[17] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (apr 1991), 152–169. https://doi.org/10.1145/123078.128728

[18] Allison Jing, Kieran May, Gun Lee, and Mark Billinghurst. 2021. Eye See What You See: Exploring How Bi-Directional Augmented Reality Gaze Visualisation Influences Co-Located Symmetric Collaboration. *Frontiers in Virtual Reality* 2 (2021), 17 pages. https://doi.org/10.3389/frvir.2021.697367

[19] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018. VRpursuits: Interaction in Virtual Reality Using Smooth Pursuit Eye Movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces* (Castiglione della Pescaia, Grosseto, Italy) *(AVI '18)*. Association for Computing Machinery, New York, NY, USA, Article 18, 8 pages. https://doi.org/10.1145/3206505.3206522

[20] Regis Kopper, Felipe Bacim, and Doug A. Bowman. 2011. Rapid and accurate 3D selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)* (Singapore). IEEE, 67–74. https://doi.org/10.1109/3DUI.2011.5759219

[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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173655

[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* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 356–366. https://doi.org/10.1145/3301275.3302278

[23] Hyeongmook Lee, Seung-Tak Noh, and Woontack Woo. 2017. TunnelSlice: Freehand Subspace Acquisition Using an Egocentric Tunnel for Wearable Augmented Reality. *IEEE Transactions on Human-Machine Systems* 47, 1 (2017), 128–139. https://doi.org/10.1109/THMS.2016.2611821

[24] James R. Lewis. 2018. The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction* 34, 7 (2018), 577–590. https://doi.org/10.1080/10447318.2018.1455307

[25] Jiandong Liang and Mark Green. 1994. JDCAD: A highly interactive 3D modeling system. *Computers & Graphics* 18, 4 (1994), 499–506. https://doi.org/10.1016/0097-8493(94)90062-0

[26] Chang Liu, Jason Orlosky, and Alexander Plopski. 2020. Eye Gaze-Based Object Rotation for Head-Mounted Displays. In *Symposium on Spatial User Interaction* (Virtual Event, Canada) *(SUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 4, 9 pages. https://doi.org/10.1145/3385959.3418444

[27] Xinyi Liu, Xuanru Meng, Becky Spittle, Wenge Xu, BoYu Gao, and Hai-Ning Liang. 2023. Exploring Text Selection in Augmented Reality Systems. In *Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Guangzhou, China) *(VRCAI '22)*. Association for Computing Machinery, New York, NY, USA, Article 35, 8 pages. https://doi.org/10.1145/3574131.3574459

[28] Feiyu Lu, Shakiba Davari, and Doug Bowman. 2021. Exploration of Techniques for Rapid Activation of Glanceable Information in Head-Worn Augmented Reality. In *Symposium on Spatial User Interaction* (Virtual Event, USA) *(SUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 14, 11 pages. https://doi.org/10.1145/3485279.3485286

[29] Xueshi Lu, Difeng Yu, Hai-Ning Liang, and Jorge Goncalves. 2021. IText: Hands-Free Text Entry on an Imaginary Keyboard for Augmented Reality Systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 815–825. https://doi.org/10.1145/3472749.3474788

[30] 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. https://doi.org/10.1109/ISMAR50242.2020.00061

[31] John Finley Lucas. 2005. *Design and evaluation of 3D multiple object selection techniques.* Ph. D. Dissertation. Virginia Tech.

[32] Francisco Lopez Luro and Veronica Sundstedt. 2019. A Comparative Study of Eye Tracking and Hand Controller for Aiming Tasks in Virtual Reality. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 68, 9 pages. https://doi.org/10.1145/3317956.3318153

[33] Mathias N. Lystbæk, Ken Pfeuffer, Jens Emil Sloth Grønbæk, and Hans Gellersen. 2022. Exploring Gaze for Assisting Freehand Selection-Based Text Entry in AR. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 141 (may 2022), 16 pages. https://doi.org/10.1145/3530882

[34] Mathias N. Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 145 (may 2022), 18 pages. https://doi.org/10.1145/3530886

[35] Diako Mardanbegi, Benedikt Mayer, Ken Pfeuffer, Shahram Jalaliniya, Hans Gellersen, and Alexander Perzl. 2019. EyeSeeThrough: Unifying Tool Selection and Application in Virtual Environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 474–483. https://doi.org/10.1109/VR.2019.8797988

[36] Daniel Mendes, Daniel Medeiros, Maurício Sousa, Eduardo Cordeiro, Alfredo Ferreira, and Joaquim A. Jorge. 2017. Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement. *Computers & Graphics* 67 (2017), 95–102. https://doi.org/10.1016/j.cag.2017.06.003

[37] Xuanru Meng, Wenge Xu, and Hai-Ning Liang. 2022. An Exploration of Hands-free Text Selection for Virtual Reality Head-Mounted Displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 74–81. https://doi.org/10.1109/ISMAR55827.2022.00021

[38] Katsumi Minakata, John Paulin Hansen, I. Scott MacKenzie, Per Bækgaard, and Vijay Rajanna. 2019. Pointing by Gaze, Head, and Foot in a Head-Mounted Display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 69, 9 pages. https://doi.org/10.1145/3317956.3318150

[39] 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. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 15, 7 pages. https://doi.org/10.1145/3448018.3457998

[40] A. Olwal, H. Benko, and S. Feiner. 2003. SenseShapes: using statistical geometry for object selection in a multimodal augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.* (Tokyo, Japan). IEEE, 300–301. https://doi.org/10.1109/ISMAR.2003.1240730

[41] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[42] 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 *Human-Computer Interaction – INTERACT 2021*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 544–565.

[43] Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/3131277.3132182

[44] Argenis Ramirez Ramirez Gomez, Christopher Clarke, Ludwig Sidenmark, and Hans Gellersen. 2021. Gaze+Hold: Eyes-Only Direct Manipulation with Continuous Gaze Modulated by Closure of One Eye. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. https://doi.org/10.1145/3448017.3457381

[45] Sheikh Rivu, Yasmeen Abdrabou, Thomas Mayer, Ken Pfeuffer, and Florian Alt. 2019. GazeButton: Enhancing Buttons with Eye Gaze Interactions. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 73, 7 pages. https://doi.org/10.1145/3317956.3318154

[46] Kunhee Ryu, Joong-Jae Lee, and Jung-Min Park. 2019. GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection. *Journal on Multimodal User Interfaces* 13, 4 (2019), 383–393. https://doi.org/10.1007/s12193-019-00305-y

[47] Robin Schweigert, Valentin Schwind, and Sven Mayer. 2019. EyePointing: A Gaze-Based Selection Technique. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 719–723. https://doi.org/10.1145/3340764.3344897

[48] Rongkai Shi, Jialin Zhang, Wolfgang Stuerzlinger, and Hai-Ning Liang. 2022. Group-Based Object Alignment in Virtual Reality Environments. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction* (Online, CA, USA) *(SUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 2, 11 pages. https://doi.org/10.1145/3565970.3567682

[49] Rongkai Shi, Jialin Zhang, Yong Yue, Lingyun Yu, and Hai-Ning Liang. 2023. Exploration of Bare-Hand Mid-Air Pointing Selection Techniques for Dense Virtual Reality Environments. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585615

[50] Ludwig Sidenmark, Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2022. Weighted Pointer: Error-aware Gaze-based Interaction through Fallback Modalities. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3585–3595. https://doi.org/10.1109/TVCG.2022.3203096

[51] Dana Slambekova, Reynold Bailey, and Joe Geigel. 2012. Gaze and Gesture Based Object Manipulation in Virtual Worlds. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology* (Toronto, Ontario, Canada) *(VRST '12)*. Association for Computing Machinery, New York, NY, USA, 203–204. https://doi.org/10.1145/2407336.2407380

[52] A. Steed. 2006. Towards a General Model for Selection in Virtual Environments. In *3D User Interfaces (3DUI'06)* (Alexandria, VA, USA). IEEE, 103–110. https://doi.org/10.1109/VR.2006.134

[53] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental Analysis of Barehand Mid-Air Mode-Switching Techniques in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300426

[54] Vildan Tanriverdi and Robert J. K. Jacob. 2000. Interacting with Eye Movements in Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 265–272. https://doi.org/10.1145/332040.332443

[55] 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* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3327–3336. https://doi.org/10.1145/2702123.2702214

[56] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4179–4188. https://doi.org/10.1145/2702123.2702355

[57] Xiyao Wang, Lonni Besançon, David Rousseau, Mickael Sereno, Mehdi Ammi, and Tobias Isenberg. 2020. Towards an Understanding of Augmented Reality Extensions for Existing 3D Data Analysis Tools. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376657

[58] Yushi Wei, Rongkai Shi, Difeng Yu, Yihong Wang, Yue Li, Lingyun Yu, and Hai-Ning Liang. 2023. Predicting Gaze-based Target Selection in Augmented Reality Headsets based on Eye and Head Endpoint Distributions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 14 pages. https://doi.org/10.1145/3544548.3581042

[59] G.J. Wills. 1996. Selection: 524,288 ways to say "this is interesting". In *Proceedings IEEE Symposium on Information Visualization '96*. IEEE, 54–60. https://doi.org/10.1109/INFVIS.1996.559216

[60] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[61] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[62] Wenge Xu, Xuanru Meng, Kangyou Yu, Sayan Sarcar, and Hai-Ning Liang. 2022. Evaluation of Text Selection Techniques in Virtual Reality Head-Mounted Displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 131–140. https://doi.org/10.1109/ISMAR55827.2022.00027

[63] Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-Supported 3D Object Manipulation in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 734, 13 pages. https://doi.org/10.1145/3411764.3445343

[64] Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Fully-Occluded Target Selection in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3402–3413. https://doi.org/10.1109/TVCG.2020.3023606

[65] Lingyun Yu, Konstantinos Efstathiou, Petra Isenberg, and Tobias Isenberg. 2016. CAST: Effective and Efficient User Interaction for Context-Aware Selection in 3D Particle Clouds. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 886–895. https://doi.org/10.1109/TVCG.2015.2467202

[66] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

Received November 2022; revised February 2023; accepted March 2023