An Exploration of Freehand Crossing Selection in Head-Mounted Augmented Reality

STEPHEN UZOR and PER OLA KRISTENSSON, Department of Engineering,
University of Cambridge

Crossing, or goal crossing, has proven useful in various selection scenarios, including pen, mouse, touch, and virtual reality (VR). However, crossing has not been exploited for freehand selection using augmented reality head-mounted displays (AR HMDs). Using the HoloLens, we explore freehand crossing for selection and compare it to the state-of-the-art "gaze and commit" (head gaze) method. We report on three studies investigating freehand crossing in multiple use cases. The first study shows that crossing outperforms head gaze in selection scenarios of varying target arrangements. The second explores crossing, head gaze, and hand pointing in radial menu and dynamic interface scenarios. The third explores crossing as a function carrier for a variety of basic interaction functions in a drawing application. This work builds on existing knowledge on the goal-crossing paradigm by demonstrating its potential as a useful interaction method in 3D AR HMD interfaces.

CCS Concepts: • **Human-centered computing** → **User studies**; **Mixed/augmented reality;**

Additional Key Words and Phrases: Augmented reality, crossing, HoloLens, selection

1 INTRODUCTION

Object selection in user interfaces is one of the most basic operations in **human–computer interaction** (**HCI**). Due to the rapidly evolving devices and modalities available today, there is an increased need for the investigation of new interaction techniques and the adaptation of established ones to facilitate effective user interaction. Given the availability of current handheld smart devices (e.g., smartphones and tablets), touch interaction currently rivals the mouse in importance for interacting with today's portable computers. However, recent years have seen the proliferation of **three-dimensional** (**3D**) interfaces, with the commercial availability of depth sensors (e.g., Microsoft's Kinect), **virtual reality** (**VR**)—for example, Oculus Rift and HTC Vive—and **augmented reality** (**AR**), **head mounted displays** (**HMDs**), for example, Microsoft HoloLens.

Authors' address: S. Uzor and P. O. Kristensson, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK; emails: steven.uzor@gmail.com, pok21@cam.ac.uk.

1.1   Selection in HCI

Selection, in one (1D) and **two-dimensional** (**2D**) interfaces, has been extensively explored in the HCI literature [1, 2, 4, 21, 55]; however, we are only just scratching the surface in terms of establishing effective methods for 3D selection [22, 55]. In the research space, 3D interfaces have been widely studied. Notable works have included investigations into volumetric displays [23], depth sensing with freehand selection [49], and VR [51, 55].

In most commercial VR devices, such as the HTC Vive [12] and Oculus Rift [20], users select objects primarily with physical controllers [51, 55]. Selection is mostly achieved in VR using either a "virtual hand" (a cursor that represents the user's hand position or VR controller [6, 7, 16]) or a technique called raycasting—projecting a ray from the controller, usually toward a flat 2D image plane in the 3D environment [16, 55, 58]. Selections, in this case, are then confirmed with a button press when the ray intersects an interface element. In **optical see-through** (**OST**) AR HMDs, raycasting is often used. However, the ray is usually cast from the user's head (the "head gaze" technique). Unlike in VR, the user's physical hands are visible to the user in OST AR HMDs, creating an opportunity for the freehand selection of virtual objects registered in physical space.

1.2   Optical See-Through (OST) Augmented Reality (AR) Displays

AR has seen increasing use in both desktop and mobile devices, and this has opened up new possibilities [56] for interaction between humans and the digital world, including in learning (e.g., Voinea et al. [59]), product design (e.g., Purdy and Choi [45]), and collaborative data analytics (e.g., Yim et al. [64]).

As is the case in mobile AR (e.g., on smart devices), AR HMDs have been extensively researched in the past [33]. However, the proliferation of Microsoft's HoloLens [14] has led to the availability of a commercial standard AR tool for researchers and consumers alike. Consequently, there is an increased need for a standardization of interaction techniques in AR. This need becomes more urgent as we proceed toward a potential future where smartphone technologies are replaced with interactive near-eye-displays as the dominant portable personal computing device [62]. The "**gaze and commit**" (**GC**) technique—the default selection method in the HoloLens and many other HMDs—uses *head gaze* for directional and rotational movement in 3D space. In some sense, this type of interaction seems unintuitive, considering that (in the real-world) users interact with objects primarily with their hands. Indeed, most HoloLens apps do not use previously studied hand-based selection techniques, such as raycasting pointing [48, 55] or the *virtual hand* technique [7, 16], despite the headset having the built-in functionality for hand tracking. One of the goals of our work is to explore both of these hand selection techniques, in a crossing-based context, and how they compare to the current *head gaze* HoloLens technique.

1.3   Crossing-Based Selection

Crossing, also referred to as goal-crossing, is an established method of selecting a target by intersecting one, or more, of its boundaries [1, 2, 41, 51, 55]—see Figure 1. The method has been used successfully in a variety of studies on 2D interfaces (e.g., Accot and Zhai [1, 2], Apitz and Guimbretiére [4]). Crossing has been employed in many modern interfaces [2, 4], for example, when hovering over a user interface element, such as a hyperlink. Some of the main advantages of crossing, include: (a) the ability to select objects without an explicit selection confirmation action [55], e.g., a button press; and (b) the ability to select multiple items, e.g., brush attributes in a drawing application, in one selection action [4].

Most of the existing work on crossing has been on 2D interfaces, using the mouse [18, 25, 61], pen [2, 4], and finger touch [37, 38]. Recent work in HCI has also explored crossing in 3D
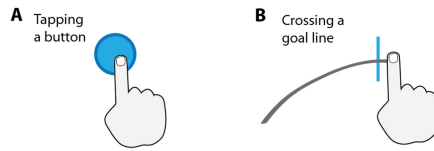
Fig. 1. In a typical tapping task, the user moves a cursor to a target and selects it using a tapping action (A). In crossing-based selection, the user "crosses" a target's goal line to select it (B).

environments [49, 55], highlighting its benefits for selection efficiency versus similar techniques, such as pointing with a VR controller [55] and *virtual hand* selection, using a wand, for example, Barrera Machuca and Stuerzlinger [6], Batmaz et al. [7]. Despite these studies, there is little work on the use of freehand crossing selection in AR environments.

One might assume that previous work on crossing in VR (e.g., Tu et al. [55]) would suffice, in this regard. However, for a number of key reasons, the differences in interactive spaces, between AR and VR, warrant additional work in the AR freehand space. First, compared with a VR controller, freehand in-air selection can be jittery and relatively inaccurate [5, 49], leading to potential differences in interaction performance. Second, users can still see the real environment, and their actual hands, in AR HMDs, which is not possible in an entirely virtual environment, that is, in VR. Therefore, users may perceive target depth and hand occlusion differently, resulting in potential differences, in selection performance, between both displays [7, 8, 43].

1.4 Contributions

We expect that AR HMDs will see increasing use as these technologies move toward a form factor that can be embraced by a wider user base [62]. Therefore, it is necessary to explore how hand-based selection techniques compare to *head gaze* in terms of efficiency and performance, to understand their potential application in this space. Furthermore, it is also important to conduct such an investigation using a wide range of target arrangements (e.g., target depth and direction) and interface styles to understand the effects of these factors on performance [10, 40, 47, 55]. In this article, we describe such an investigation, and we report on three user studies to explore the efficiency, practicality, and potential use cases of freehand crossing in AR HMDs.

In Study 1 (Section 5), we explore freehand crossing selection in a variety of target configurations that are reasonable in a raycasting context—in 2D planes with different target depths, amplitudes, and directions (angles). We find that crossing outperforms *head gaze* for **selection time (ST)** and error rate, demonstrating its potential as a viable selection technique in the freehand AR space.

In Study 2 (Section 6), we investigate the benefits and drawbacks, with regard to user experience and performance, of freehand crossing, *head gaze,* and hand pointing (all raycasting-based techniques) in a variety of practical multiple selection use cases—including static, radial menu, and dynamic (or moving) target selection, which are all commonly encountered in HCI. We find that crossing can be just as efficient, and even preferable to *head gaze*, in static and dynamic scenarios. However, crossing can be less accurate than *head gaze* in radial menu scenarios, where subsequent target locations are not necessarily anticipated by the user.

In Study 3 (Section 7), we explore freehand crossing in a variety of selection operations, using a drawing application, in both a raycasting context, for 2D plane selection, and in a *virtual hand* context for performance (efficiency) and user experience. We find that raycasting-based crossing is comparable to *head gaze* in terms of performance. However, *virtual hand*-based crossing in 3D selection is less efficient and requires more effort than *head gaze* and raycasting crossing.

Through all of these investigations, we posit that raycasting-based freehand crossing can be a viable and efficient selection method for use in AR HMD interfaces. Finally, in Section 8, we

discuss the HCI challenges and how these can be mitigated for both an enhanced user experience and improved performance. The lessons learned and the resulting design recommendations help to expand the emerging literature on freehand selection in AR HMD design spaces.

2 BACKGROUND

Due to the vast literature on selection in HCI, it is not possible to discuss all of the relevant work in this article. Hence, we discuss key studies on selection in 3D environments, VR/AR and crossing.

2.1 Crossing-Based Selection

Crossing is a selection technique involving the deliberate intersection of a target's boundary [1, 2, 4, 51, 55]. Previous work has highlighted the benefits of crossing in both 2D (e.g., Accot and Zhai [1, 2], Apitz and Guimbretiére [4]) and 3D interfaces (e.g., Ren and O'Neill [49], Susu et al. [51], Tu et al. [55], Wolf et al. [62]). For instance, several works have found that crossing can encourage more efficient selections in 2D touch interfaces [2, 3, 24, 37, 38]. Accot and Zhai [1], in one of the earliest studies on crossing, found that the technique performs just as well as, and in some cases better than, pointing on touch surfaces. Further benefits of crossing are apparent in its application in aiding users with motor impairments [61].

Since, crossing does not require an explicit selection confirmation action, users are able to select targets easier. In contrast, established techniques, such as pointing and gaze, often require a specific confirmation action for this purpose; for example, a button press, an "air-tap" with the index finger, or a dwell timeout (hovering over a target for a period of time) [9, 50]. These confirmation actions are intended to reduce errors, for example, the *Midas touch* effect [57, 62], which results in unintentional target selection. Since crossing requires no explicit confirmation action, the *Midas touch* effect is one potential disadvantage of crossing, for example, in interfaces with multiple densely packed targets [49].

2.2 Selection in 3D User Interfaces

2.2.1 Selection Techniques in 3D Environments. Three-dimensional interfaces have opened up avenues for additional selection techniques that extend beyond the traditional 2D mouse and touch input. However, designing 3D interfaces that allow effective and efficient target selection is more challenging compared with 2D [29]. Hence, numerous selection techniques have been explored in this domain and have given rise to a variety of interfaces [19, 23, 30, 32, 34, 42, 44].

Most selections in 3D [51] are implemented using either raycasting (e.g., Grossman and Balakrishnan [23], Lehikoinen and Röykkee [34])—a technique where a ray is projected from an input device (akin to pointing with a laser pointer), or a "virtual hand"—a technique that places a cursor at the hand position [6, 7]. Machuca et al. [6] introduced a variation of Fitts' Law to determine movement depth with a *virtual hand* in stereoscopic displays; this was implemented using a marker wand. Batmaz et al. [7] also implemented a similar *virtual hand* technique and found that movement toward and away from users in AR HMDs and VR was slower than similar movement in the lateral direction (side to side).

2.2.2 Raycasting Pointing. Two-dimensional image planes have been shown to foster relatively simpler 2D selections in 3D environments [36, 48, 53, 55]. However, selecting 3D objects using both freehand and controllers can be difficult compared with selecting 2D targets with a raycasting technique [51, 55]. Part of this difficulty stems from the challenges associated with perceiving depth when looking, or pointing, at a target [7]. Ramcharitar and Teather [48] attempted to address this problem with their *EZCursorVR* technique, which presents a 2D cursor in front of the user in order to enable easier 3D selections beyond the image plane. Zhai et al. [67] developed the

Silk Cursor, a volumetric semi-transparent cursor to facilitate easier selections in a dynamic 3D environment. Tu et al. [55] used a pointing technique (based on raycasting) as a foundation for crossing selection in VR. The technique used a forward-projected ray from a VR controller—e.g., as described in Ramcharitar and Teather [48].

Raycasting has also been used in freehand selection scenarios. For example, Ren and O'Neill [49] investigated freehand crossing, using the Kinect, for marking menu selections. We use raycasting-based freehand selection in our work, in a similar way to Ren and O'Neill [49], but in the AR OST HMD space. The implementation of our freehand crossing technique is described later in Section 5.2.

2.2.3 Gaze-Based Selection. Gaze-based methods have also been explored in 3D selection and target acquisition. Such a method seems ideal for HMDs, since a user can leverage the natural movement of the head to acquire objects in 3D space [46, 47, 50]. We make the distinction between "eye gaze" and "head gaze" as these are two different gaze methods presented in the literature. *Eye gaze* uses eye trackers to enable the user's eyes to acquire targets [52] and has been extensively studied in the past [9, 15, 46, 47, 50, 52].

Head gaze is the current dominant method for target acquisition in AR HMDs. The literature suggests that *head gaze* outperforms *eye gaze* in similar virtual environments. For instance, Qian and Teather [46] found that *head gaze* outperforms *eye gaze* in terms of error rate, ST, and throughput in VR. The disadvantage for *eye gaze* lies in the saccadic movement [50] of the eyes, making *eye gaze* more jittery and error-prone than *head gaze* [46]. However, *head gaze* has been found to suffer from neck fatigue [46, 50]. Blattgerste et al. [9] also found benefits for *eye gaze*, highlighting its advantages over *head gaze* regarding speed, task load, and user preference.

In the rest of this article, we use the term *gaze* to mean *head gaze*, rather than *eye gaze*, since *head gaze* is the dominant gaze method used in current AR and VR HMDs, and it provides a baseline comparative method for our investigation into freehand crossing.

2.2.4 Crossing in 3D Environments. There is little work on the exploration of crossing for selection in 3D environments. However, there is existing work on crossing in 3D using the Kinect [49] and in VR [55]. To our knowledge, only Tu et al. [55] have investigated crossing in VR using raycasting pointing. Tu et al. found that crossing performance was no worse than pointing (a selection technique in VR), which provided a baseline comparative method. The work by Tu et al. also verified that the target acquisition time in a VR environment can be modelled using Fitts' Law. However, crossing in VR, as described in Tu et al. [55] relies on selections using physical controllers. Apart from the downside of encumbering the user, such controller selections are not representative of similar selections using AR, OST, and HMDs. This is due to the fact that a physical handheld controller may not necessarily be used in the same way as in freehand selection. For instance, freehand selection in AR requires holding, and waving, the hand in mid-air (see Figure 2(A)). In contrast, a user can rotate a VR controller, using their wrist, to achieve a similar selection [5]. Furthermore, freehand selection can be noisy, jittery, and relatively inaccurate [5, 35, 49], compared with device-based selection. Hence, there is a need for further investigation into freehand crossing selections in AR, which VR-focused works (e.g., Tu et al. [55]) do not fully address. Nevertheless, there are similarities across both VR and AR, regarding pointing [7]. Therefore, we learn from the design recommendations outlined in prior work (e.g., Tu et al. [55]), which are applicable to our work in certain cases.

Crossing has also been explored in freehand AR HMD selection scenarios. For instance, Wolf et al. [62] investigated target acquisition, crossing and circular steering using a freehand *virtual hand* method compared to smartwatch interaction using the HoloLens. Similarly, Ren and O'Neill [49] investigated freehand crossing, using the Kinect, in a raycasting context, which formed

the foundation for their *Stroke* and *Reach* techniques. Their work highlights further opportunities to understand how techniques, such as crossing, are used in the AR HMD space.

In this work we compare freehand crossing in AR against an established baseline: The state-of-the-art HoloLens GC selection method, which uses head tracking and raycasting (i.e., *head gaze*) to acquire targets. We show that crossing can perform better than GC, paving the way for new possibilities for freehand crossing selection in AR HMD scenarios.

3 RESEARCH OBJECTIVES

(1) Investigate the feasibility and performance of freehand crossing in a variety of selection use cases in AR HMDs.
(2) Identify the benefits and drawbacks of freehand crossing in AR HMD selection scenarios.
(3) Investigate user preferences regarding the use of freehand crossing, compared with state-of-the-art selection techniques used in AR HMDs.
(4) Identify recommendations, based on our investigations, for the design of freehand crossing-based interactions in AR HMD selection scenarios.

4 APPARATUS

4.0.1 Hardware. We used the first generation Microsoft HoloLens [14], a commercial AR OST HMD. The technical specifications for the HoloLens, include four environment cameras (for spatial mapping), an inertial sensor, an HD video and depth camera, an ambient light sensor, and four microphones [13]. The HoloLens can also track general hand position in addition to tracking hand gestures—e.g., the air-tap. We use this as a foundation for our crossing methods in all of our studies—both *raycasting* and *virtual hand* crossing.

4.0.2 Software. We developed all of our software applications in Unity [54], one of the most popular game engines used today, which allows for the deployment of 2D and 3D applications across a variety platforms, including the HoloLens.

5 STUDY 1: FREEHAND CROSSING EFFICIENCY IN AR

The first study serves as a foundational study to investigate how selection efficiency in raycasting-based freehand crossing compares to *head gaze* selections in a variety of target depths and directions.

5.1 Selection Techniques

5.1.1 Gaze and Commit (GC). This is the default selection technique used in the HoloLens headset. GC employs head tracking to determine a cursor position using the raycasting method. Users gaze at the interface element of interest (e.g., a button), aided by a cursor, then they select the target using the *air-tap* gesture. Note that, in this article, we refer to gaze with regard to head tracking, rather than eye tracking, which is also commonly explored in the literature [9, 27, 46, 50, 52].

5.1.2 Raycasting Crossing (CR). Similar to prior work [49], we used freehand pointing as a foundation for our raycasting crossing method. A ray is cast from the user's hand, along a directional vector that originates from a point [31, 39] located 10 cm (to the right or left depending on handedness) from the estimated base of the neck (see Figure 2(A)).

5.2 Study Design

We employed a within-subjects design, exploring a single selection task with four independent variables: *technique* (GC and raycasting crossing), *target amplitude* (9 cm and 21 cm), *target depth* (100 cm, 175 cm, and 250 cm), and *target direction* (0, 45, 90, 135, 180, 225, 270, and 315 degrees—see
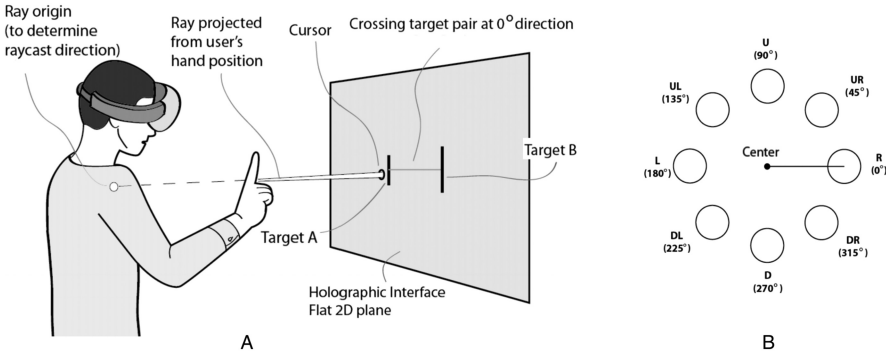
Fig. 2. (A) Raycasting in our crossing technique—in the GC technique, the ray is cast from the HoloLens, while the hand confirms selections with an air-tap gesture. (B) Target directions investigated in our work—adapted from Murata and Iwase [40] and used in prior work [10, 49].
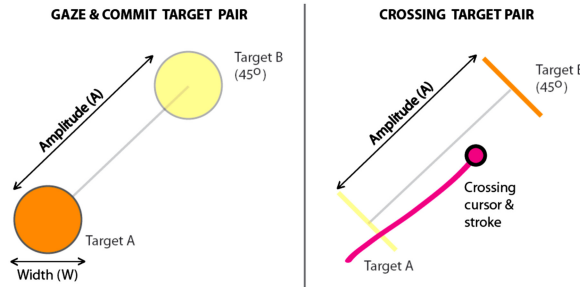

Fig. 3. Target pairs in both techniques—second target is highlighted only after selecting the first one (right).

Figure 2(B)). Note that we refer to target depth as the forward distance between the user's head and the target. Target amplitudes were arranged, such that if the first target (Target A) appeared at the centre, the second target (Target B) appeared at one of circles (see Figure 3), indicated by directions: R (right), UR (up right), U (up), UL (up left), L (left), DL (down left), D (down), and DR (down right). We chose the above amplitudes so that both targets are visible within the HoloLens field of view (FoV), which is 35 degrees [62]. This allows for: (a) potentially more accurate selections to be made [51, 62]; (b) the elimination of potential confounding variables, such as searching for targets that are out of the FoV [46]; and (c) a selection range that is representative of real-world scenarios, in other words, an arrangement such that the entire operational interface is visible to the user. Targets were presented in pairs (see Figure 3). GC targets were presented as circles, whereas crossing targets were presented as goal lines of the same length (3 cm). Crossing goal lines were presented in a direction-constrained configuration [2, 55]—facing each other. Flat targets were used, as they are more suitable in ray-casting scenarios [55].

When a new pair appeared, in either technique, the first target was highlighted and the second one was faded [6]. This was done to show the user the target with which to begin the selection (see Figure 3 left). After selecting the first target, the second was highlighted and the first faded. Figure 3 (right) shows a crossing pair with the first target already selected. After selecting both targets, a new pair appeared following a 3 second delay. Each target pair appeared with a random combination of depth, direction, and amplitude, according to the parameters described above. Participants were required to do crossing actions using continuous, rather than discrete, strokes [2, 55]. The experiment was conducted against a black curtain to enable clear viewing of the targets.

For both GC and CR techniques, each participant performed 5 blocks of 48 selections (1 width × 2 amplitudes × 3 depths × 8 directions). Therefore, there were 480 trials per participant (48 selections × 5 blocks × 2 techniques), for a total of 7,200 trials for all 15 participants. We used Block 1 as a training block and the resulting findings were not included in the analyses. The participants were required to rest for 2 minutes between blocks and 5 minutes between techniques. This was done to reduce the effect of arm fatigue on subsequent selections. The study lasted for about 1 hour. We did not investigate motion sickness, since the AR effects are negligible [60] compared with VR. None of the participants reported feeling sick during the study.

5.3 Recruitment

We recruited 15 right-handed participants (7 females, 8 males; mean age = 27, *sd* = 4 years) from the authors' university. Participant demographics consisted of post-doctoral researchers (n = 5), research students (n = 8) and other relatively young professionals. The participants rated their experience with general technology (e.g., computers, mobile devices) and VR/AR (e.g., HoloLens, Oculus Rift) on a Likert scale of 1 (no experience) to 5 (very experienced). The average distribution was 4.2/5 (*sd* = 0.75) for general and 2/5 (*sd* = 1.4) for VR/AR technologies.

Ethics approval, for this work, was obtained from the authors' university departmental ethics committee. The participants, in all of our studies, gave informed consent prior to participation and were rewarded with a £10 Amazon voucher.

5.4 Study Metrics

— Selection Time (ST): The total time from the selection of the first to the second target.
— Error Rate: We defined this as the rate of selections made outside the target area [55] to the total selections. Error selections occurred within two widths of the target. If an error occurred, the pair was repeated. Error selections were not included in the ST analysis, as practiced in previous work [2, 37, 40, 55].

5.5 Study Results

We analyzed data using Microsoft Excel 2015 (for simple analyses and descriptive statistics, e.g., *mean*, *sd*) and IBM SPSS Statistics 25. We used **repeated measures factorial analyses of variance (RMANOVA)** for analyzing *ST* and error rate. Post-hoc comparisons were Bonferroni corrected using an initial significance level of α = 0.05. In cases where Mauchly's test of sphericity was violated, we applied a Greenhouse–Geiser correction.

5.5.1 Selection Time. We found shorter *STs*, on average, in the crossing technique, across all target depths (Figure 4) and directions (Figure 5). **The difference, in *ST*, between both techniques was statistically significant ($F_{1,47}$ = 24.6, $p < 0.01$, η_p^2 = 0.34).** This result suggests that crossing encourages quicker selections than GC. We found no significant differences between the various target depths and directions. There were also no observed interaction effects on *ST* for [*technique* × *depth*], [*technique* × *direction*], or [*technique* × *direction* × *depth*].

5.5.2 Error Rate. We found higher error rates in the GC technique (*mean* = 6%, *sd* = 10) than in CR (*mean* = 4%, *sd* = 8.5). Results are highlighted in Figure 6 for target depth and Figure 7 for direction.

The difference, in error rate, between both techniques was also found to be statistically significant ($F_{1,48}$ = 85.4, $p < 0.01$, η_p^2 = 0.64). These findings further reinforce the results showing better performance in CR than GC. The increased error rate in GC can be attributed to the user holding their gaze steady (dwell time) while confirming their selection with the air-tap. Next, we investigated the effect of target depth and direction on error rate. The results can be
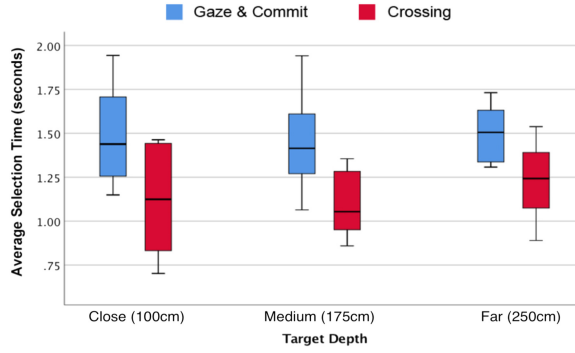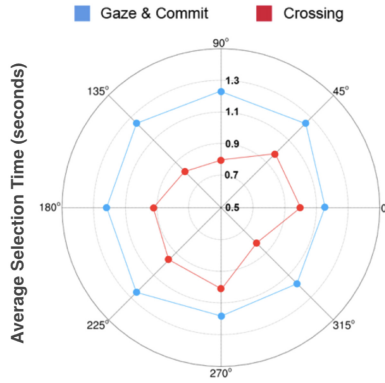
Fig. 4. Average *ST* at all target depths.

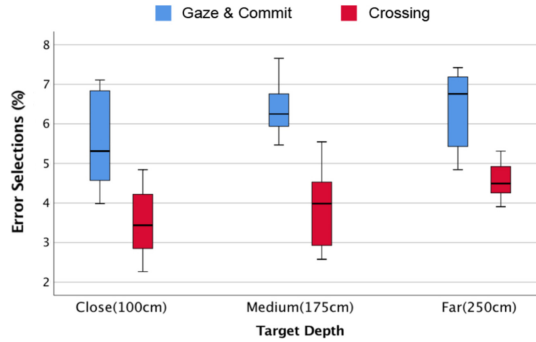Fig. 5. Polar plot of *ST* at different target directions in both techniques.

Fig. 6. Error rates at all target depths.

seen in Figure 6 for depth and in Figure 7 for direction. No significant effect was found for direction, [*direction* × *technique*], [*depth* × *technique*], [*direction* × *depth*], and [*direction* × *depth* × *technique*]. However, **a significant effect was found for depth ($F_{2,48} = 21.2$, $p < 0.05$, $\eta_p^2 = 0.17$)**. Post-hoc analyses showed that there was only a **significant difference between close (100 cm) and far (250 cm) depths**—$F_{2,48} = 4.9$, $p < 0.01$, $\eta_p^2 = 0.17$. Since, there was no interaction effect for [*depth* × *technique*], the findings suggest that significantly more errors arise when selecting far targets compared with close targets, irrespective of selection method.
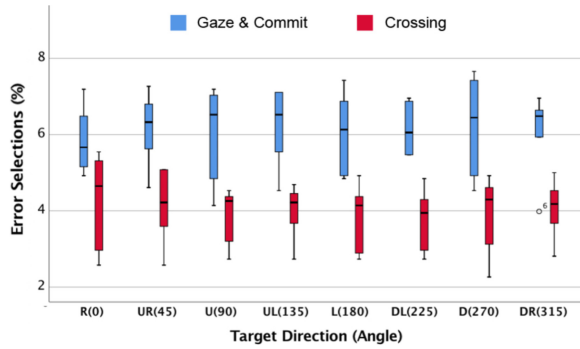
Fig. 7. Error rates at all target directions.

5.6 Study Discussion

CR resulted in significantly less ST, compared to the GC technique under the studied parameters. This result parallels similar work in VR [55], where raycasting crossing, with a controller, was found to encourage quicker STs than a comparable pointing technique. Our study findings did not reveal any significant effect of direction on ST in either the CR or GC technique. This result is in contrast to Ren and O'Neill [49] who found a significant directional effect on crossing ST, but agrees with Grossman and Balakrishnan [23], where direction was not found to impact ST in 3D *virtual hand* selection. Our target amplitudes (9 cm and 21 cm) were smaller than that used in Ren and O'Neill [49]. Therefore, it is possible that direction has more of an impact at larger target amplitudes.

Target depth was not found to have a significant effect on ST. This is a similar finding to Tu et al. [55], where there was no difference in ST between depths of 76 and 230 cm in VR raycasting pointing/ crossing. However, in Tu et al. [55] a depth of 345 cm was found to take significantly more time than the other two depths. Note that our far depth is 250 cm, which is similar to medium depth in Tu et al. [55]. Contrary to Tu et al.'s study, we found that a 250 cm depth incurs significantly more errors than close (100 cm) targets, but not medium (175 cm) ones. There are two possible factors that contribute to the difference between our work and Tu et al.'s work. First, we use freehand selection which is relatively inaccurate, due to subtle hand tremors, compared to VR controllers [49]. Therefore, it is perhaps not strange to observe more errors in less AR space than the equivalent space in VR using a controller. Second, it is noteworthy that target depth is inversely proportional to the angular size of the selection (width and amplitude). Therefore, given the relatively smaller angular width and amplitude at far depth, there is a potential for higher errors and more time for the selection as users attempt to compensate for precision due to the movement of the arm.

6 STUDY 2: FREEHAND CROSSING IN MULTIPLE SELECTION SCENARIOS

The goal of this study was to investigate the application of raycasting-based freehand crossing in multiple selection scenarios of varying complexity and dynamicity. This investigation is necessary because nested menus—for example, drop-down and radial menus, are used in several Windows, Icons, Menus, Pointer (WIMP) and non-WIMP interfaces [24]. Similarly, dynamic (moving) targets are common in many interaction scenarios, such as games, air traffic control systems, and interactive sports [25, 26]. In these cases, additional challenges arise for efficient selection and interaction [24, 26]. Hence, we explored the application of freehand crossing in these (static, radial menu, and dynamic) distinct selection scenarios to identify implications for design regarding efficiency and user preferences for in-air selection. While there are other interface layouts and sce-
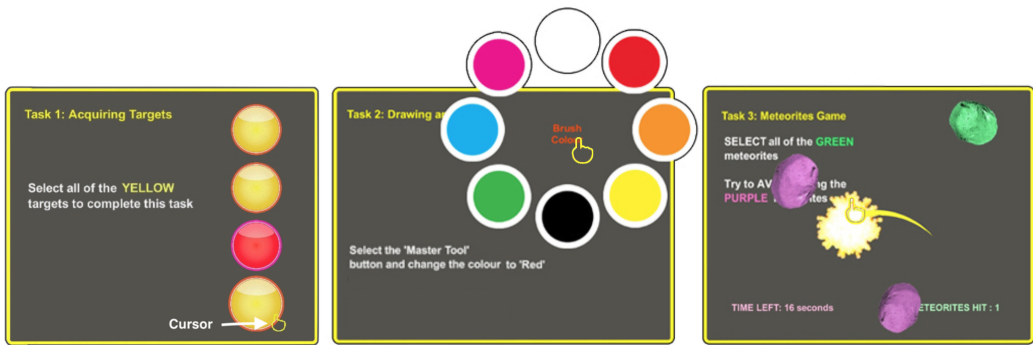
Fig. 8. Task interfaces highlighting the various selection scenarios and techniques (cursors). Static with GC action (left), Radial Menu with *Freehand Pointing* (middle), and Dynamic with *Raycasting-based Freehand Crossing* (right).

narios worthy of investigation, we chose these three selection scenarios since they are commonly used in user interfaces and they provide a suitable diverse range of scenarios to explore within the scope of the current work.

6.1 Selection Techniques

6.1.1 Gaze and Commit (GC). This is the default selection technique used in the HoloLens headset (also studied in Study 1—see Section 5.1).

6.1.2 Freehand Pointing (FP). This technique primarily uses hand pointing rather than head tracking. The cursor position is determined using a ray projected forward from the user's hand (see Figure 2) as described in the freehand crossing method—see Section 5.1. In the work by Tu et al. [55], the pointing ray was projected from a VR controller, whilst a button was used to confirm selections. Our freehand pointing technique uses an *air-tap* to confirm selections, as in GC, since it is freehand-based.

6.1.3 Raycasting-Based Freehand Crossing (CR). This is the same crossing technique explored in Study 1—see Section 5.1).

6.2 Selection Scenarios

In this section, we describe the selection scenarios explored in Study 2.

6.2.1 Static Selection. The static selection interfaces contained buttons with static positions, that is to say, they do not move. A majority of interfaces contain this type of button. We included three targets for the user to select and one target to avoid—to know whether, or not, selections were intended (see Figure 8 left).

6.2.2 Menu Selection. We drew inspiration from Guimbretiére and Winograd [24] and Ren and O'Neill [49], by using a radial-style menu, which gives rise to potential selection challenges in a relatively complex scenario. The menu was nested on three levels—the red target (see Figure 8 center), could be selected using the path shown in Figure 9: A (menu) > B (color) > C (red color). One option was presented at the first level (A), four were presented at the second level (including B), and eight options were presented at the third level (including C).

6.2.3 Dynamic Selection. Similar to the radial menu (we will refer to this simply as "menu") scenario, we included the dynamic selection scenario to investigate efficiency and user experience when dealing with moving targets. For this scenario, we developed a simple game, where users
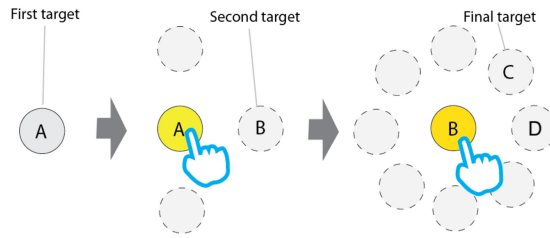
Fig. 9. *Radial Menu* selection scenario task—the user selects A, then B, and then C to complete the task. Targets A and B open the additional expanding radial menus highlighted with dashed lines at each level.

had to destroy falling meteorites by selecting them (Figure 8, right). Meteorite origins were spaced 20 cm apart and appeared every second. They fell at a constant speed of 10 cm per second, and the order of appearance was the same in all conditions.

6.3 Interface Design

We used 2D image plane interfaces and buttons that enabled users to perform various tasks, depending on the selection scenario. The interfaces were developed in accordance with the following guidelines from the literature and the HoloLens documentation.

— `Interface Distance and Size`: All interfaces were facing the user, as a flat 2D plane in a 3D environment [48, 51, 66], and located at a distance of 1.8 m from the user (optimal effective target range of 1.25–5 m, with a recommended range of 2 m [13]). Interfaces were approximately 1 m in diameter, enough to fill the HoloLens field of view (35 degrees) at a distance of 1.8 m. Note that at the specified distance from the user, a 1-meter diameter allows for a 1–1.5 ft range of movement for the user's arm from one end to the other.

— `Target Width and Amplitude`: Targets (e.g., buttons) were 8 cm in diameter—large enough to accommodate for reasonable selections at the chosen interface distance (1.8 m). Targets were flat and circular to encourage effective selection using raycasting methods [51, 55], including *in-air* crossing. In the crossing scenario, selection was achieved when the user crossed the centre line of the target (perpendicular to the direction of arm movement). Target amplitude (distance between center points) varied between targets: 15 cm for static (Fitts' **index of difficulty** (**ID**), or *ID*, = 1.8 bits), 15–30 cm (maximum) for menu (*ID* range: 1.8–3 bits), and 30 cm for dynamic targets (*ID* = 3 bits). These parameters were chosen to allow for effective selections, using a relatively small *ID* range and within the ideal interaction space for the HoloLens.

— `Interface Color`: Bright colors were used, on both the interfaces and buttons, to allow for good contrast with the environment. The primary button color was "deep yellow".

— `Cursors`: Cursors were designed for good contrast with other interface elements (the default color for cursors was "bright yellow") to enhance the cursor visibility.

6.4 Study Design

We employed a within-subjects design with *technique* as the independent variable. The participants were asked to perform one task for each technique, in each scenario (see Table 1). Therefore, there were nine performed tasks in total. The order of presentation of the various techniques and scenarios was counterbalanced, such that an equal number of participants were exposed to one order. This was done to reduce potential confounding effects and bias, for example, due to learning.

Table 1. Specific Tasks in Each of the Selection Scenarios
in Study 2—Image Reference in Figure 8

Scenario	Task
Static	Select all of the "Yellow" targets.
Menu	Select the "Red" target.
Dynamic	Select ten "Green" meteorites.

Participants all sat in a chair facing a black curtain, allowing them to see the interface elements clearly. Between each technique the participants were asked to rest for 2 minutes, to reduce the effect of fatigue on the next set of tasks. Given the simplicity and duration of the tasks we did not anticipate high levels of fatigue, hence the 2-minute rest time. All of the participants were right handed. The study lasted for 45 minutes. None of the participants reported that they felt sick during the study.

6.5 Recruitment

We recruited 24 participants (9 females, 15 males; mean age = 30, sd = 5.04 years) from the authors' university. Participants consisted of post-doctoral researchers (n = 8), research students (n = 10), and other relatively young professionals. The participants rated their experience with general technology (e.g., computers, mobile devices) and VR/AR (e.g., HoloLens, Oculus Rift) on a Likert scale of 1 (no experience) to 5 (very experienced). The average distribution was 4.25 (sd = 0.74) for general and 2.38 (sd = 1.28) for VR/AR technologies.

6.6 Study Metrics

— Time to Complete (TTC) Task: This is the time the user completed each task, from the selection of a start button to the last button in the task.
— Selection Accuracy: We defined this as a percentage of the number of correct selections to the total number [51] of required selections necessary to accomplish each task.
— User Experience: Using an exit questionnaire and a debrief session, the participants provided further insights into their preferred selection technique and user experience regarding using the various techniques for the described tasks.

6.7 Study Results and Discussion

We used **analyses of variance** (**ANOVAs**) for analyzing TTC and selection accuracy. Post-hoc comparisons were Bonferroni corrected using an initial significance level of α = 0.05. In cases where Mauchly's test of sphericity was violated, we applied a Greenhouse–Geiser correction. Direct participant quotes are used to provide further insights into the results.

6.7.1 Time to Complete Task (TTC). **Significant differences were found, between the selection techniques, in the *Static* [$F_{1.5,34.9}$ = 7.25, p < 0.01] and *Menu* [$F_{1.8,41.3}$ = 6.97, p < 0.01] tasks**, but not in the *Dynamic* task—see Figure 10.

Static task: According to pairwise comparisons, **GC resulted in significantly quicker times compared to FP (p < 0.01)**. No other significant findings were found in this case. This suggests that GC task times are quicker than FP, whilst CR performs just as well as GC, when targets are static. Given these findings, we cannot conclude that crossing is quicker, for multiple static selections, than freehand pointing, and further exploration is needed in this regard.

Menu task: **GC resulted in significantly quicker times than both FP (p < 0.05) and CR (p < 0.01)**. We observed that users often made relatively long strokes with the hand in the FP and
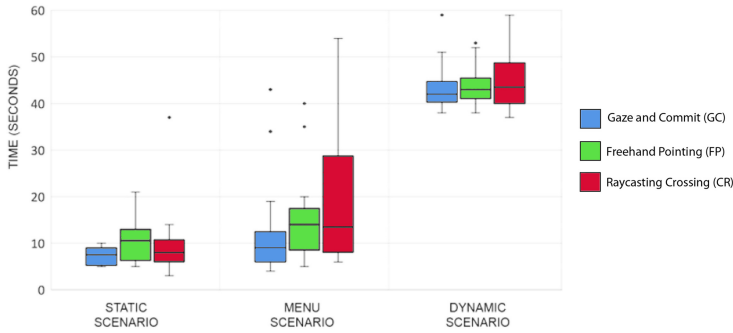
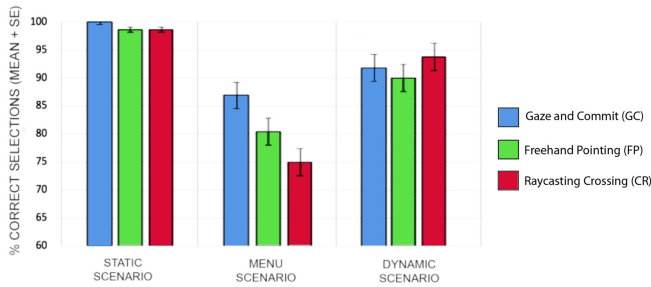Fig. 10. Time to complete task in each scenario.

Fig. 11. Selection accuracy in each scenario (percentage of correct selections and overall selections), with error bars highlighting standard error.

CR techniques. Since both of these techniques use raycasting pointing, strokes made at the hand position translated to longer strokes at the cursor position. Therefore, there was a tendency to overshoot targets, resulting in longer times and incorrect target selections (in the CR technique).

No significant differences were found between the hand-based techniques in the *Menu* task. This suggests that either hand-based selection technique could be used to a similar effect. Dixon et al. [17] found a comparable result in touch interfaces, showing that crossing resulted in identical task times to *point and click* selection. In this regard, our work highlights parallels between 2D touch and 3D AR interfaces, although both interfaces cannot be directly compared due to the differences in the input effort required. Our work also agrees with prior work that found raycasting crossing to be comparable to pointing in certain VR tasks [55].

6.7.2 Selection Accuracy. No significant difference was found between techniques in the *Static* task. Indeed, results in this task were similar and generally low for all techniques (see Figure 11). Conversely, **significant differences were found in the *Menu* [$F_{2,46}$ = 7.61, $p < 0.01$] and *Dynamic* [$F_{1.3,30}$ = 12.4, $p < 0.01$] tasks.**

According to pairwise comparisons, **statistically significant differences exist only between GC and CR ($p = 0.02$) in the *Menu* task**. The apparent disadvantage for CR in this case is due to the previously mentioned issue with incorrect target selections. The "Midas touch" effect [57] suitably explains this result. This problem can be more prevalent in relatively complex freehand configurations, for instance, radial menus, where buttons are closely grouped together—especially when no deliberate action (e.g., *air-tap* in GC and FP) is used.

In the *Dynamic* task, the results show that **CR encourages less errors than both GC ($p = 0.01$) and FP ($p < 0.01$)**. There is a reinforcement of the idea that in more widely spaced targets

(e.g., the meteorites in the game), the *Midas touch* effect is less of an issue for CR. This supports the work by You et al. [65] where low error rates were found when using crossing for moving targets. Since crossing relies on selection by intersecting a target, rather than *point and air-tap* (as in GC and FP), CR appears to lend itself well to selecting dynamic or moving targets. Conversely, in the other techniques that depend on an explicit *air tap*, for selection, there seems to be a chance that the target will be missed after acquiring it (i.e., during the *air tap* action) since the target is in motion.

6.7.3   User Preferences. Preferences for technique were varied. Approximately 50% of the participants said that they would primarily use the GC technique. User preference for *head gaze* was mostly due to reduced fatigue, experienced by the participants, since GC involved little hand/arm movement, compared with the hand-based ones. To this end, P8 commented: "[GC] has the least jitter and fatigue," while P22 commented: "[GC] has the best accuracy and least fatigue." It is possible that the participants experienced much more fatigue in the hand-based techniques due to the fact that multiple selections were involved in each scenario. Had the selection tasks been short (e.g., in Study 1), these results might have been different, although further work is warranted in this regard.

There were certain cases where GC was not favored over the hand-based techniques. For instance, participants P1, P13, and P24 felt that pointing was the easier and most natural technique (P1: "(...) pointing felt more natural to me;" P24: "[Freehand] pointing was the easiest." Crossing was favored by certain participants, as it allowed them to perform actions quickly (P8: "(...)I think [CR] could be very fast for tasks that do not require precision") and had an appealing characteristic (P19: "[CR] felt more like using a pen;" P16: "I like the swipe action;" P22: "[CR] is fun and is best for games"). This further supports the literature on crossing in touch-based interfaces [24, 37] and our findings on TTC in the *static* and *dynamic* selection scenarios.

Finally, P10, who reported no favorite selection technique, acknowledged that there are different scenarios in which all selection techniques could be useful: "They can all be used for different purposes, [GC] for sensitivity, FP and CR for quickly selecting."

7   STUDY 3: FREEHAND CROSSING IN A DRAWING APPLICATION

The purpose of Study 3 was to formally investigate freehand crossing selections, using a number of gestures, within the context of a practical drawing application. Further, in keeping with the theme of this work, we intended to compare freehand crossing performance to *head gaze*, as the current state-of-the-art in AR HMDs. Through this investigation, we can identify the benefits and drawbacks of different types of freehand crossing techniques in practice. Consequently, we investigated two types of 3D crossing methods previously studied in the literature—raycasting crossing (the same CR technique in Studies 1 and 2) and *virtual hand* crossing (using a point cursor placed at the user's hand position) [62]. We refer to the *virtual hand* crossing technique as virtual hand crossing (CV).

7.1   Study 3 Procedure

We employed a within-subjects design, with *technique* as the main independent variable. Users performed five different tasks (see Figures 12 and 13 and Table 2). We limited the study to five tasks to avoid the effects of fatigue due to the extensive number of hand gestures required.

In order to understand how freehand crossing compares to *head gaze* in this context, we implemented a similar application in which users could perform the same tasks using a traditional gaze-based AR interface (see Figure 13, GC column). Note that this experiment was not intended

Table 2. Selection Tasks Explored in Study 3

Task	Description
Single selection	Select the "Pen" tool. In the crossing techniques, this is achieved with a crossing stroke through the "Tool" button and ending with a stroke through the *Pen* tool button. In GC, the selection sequence: [Tool button -> Pen tool] is used.
Multiple selection	Select the following brush attributes: Thickness (Level 1) = 0.3, Opacity (Level 2) = 0.8, and Colour (Level 3) = Red. In CR and CV, this is achieved with a crossing stroke through the *Tool* button and moving the stroke through the various attribute levels, ending with a stroke through the colour scale. In GC, the same goal is achieved through multiple actions: [Thickness button -> Thickness value], [Opacity button -> Opacity value], and [Colour button -> Colour value].
Scale transform	Scale an image to match the scale of another image (0.5 scale). Note: in CV, a 3D model was used instead of a 2D image for transform operations. In CR and CV, the operation is achieved with an initial sequence of a crossing stroke through the "Transform" button, moving the stroke through the "Scale" button, then scaling—moving the stroke toward or away from the image to scale it. In GC, the initial selection sequence is [Transform button -> Scale button], then scaling.
Rotate transform	Rotate an image to match the rotation another image (rotated at 75 degrees clockwise). In CR and CV, the operation is achieved by beginning with a crossing stroke through the *Transform* button, moving the stroke through the *Rotate* button, then rotating—moving the stroke round the image to rotate it in the required direction. In GC, the initial selection sequence is [Transform button -> Rotate button], then rotating.
Translate transform	Move an image to match the position of another image (20 cm to the right). In CR and CV, the operation is achieved by beginning with a crossing stroke through the *Transform* button, moving the stroke through the *Translate* button then translating—moving the image with a stroke toward the right. In GC, the initial sequence is [Transform button -> Translate button], then translating.

Refer to Figure 12 for the crossing gestures in each of the tasks for the crossing techniques. Note: In the CV condition, crossing lines had depth (as can be seen in Figure 13).
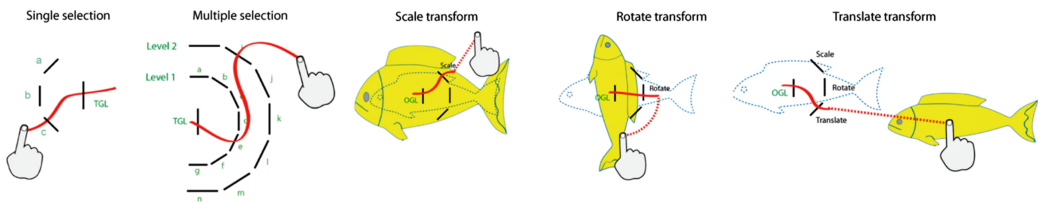

Fig. 12. Crossing actions and gestures required for each of the Study 3 tasks, highlighting crossing actions starting from either a **tool crossing goal line** (**TGL**) or an **object crossing goal line** (**OGL**). Note: dotted lines in transformations indicate initial object position, whereas the solid objects show the final transformation.

as a one-to-one comparison between *head gaze* and freehand crossing given that the interfaces for each technique are expected to be different, in order to maximize their strengths. However, the layout of the buttons, with respect to the tasks (see Figure 13) was similar in the sense that the distance between targets (initial and final) was the same. The target directions (angle) and sizes were also similar, i.e., the (ID was the same, providing a foundation for comparison.
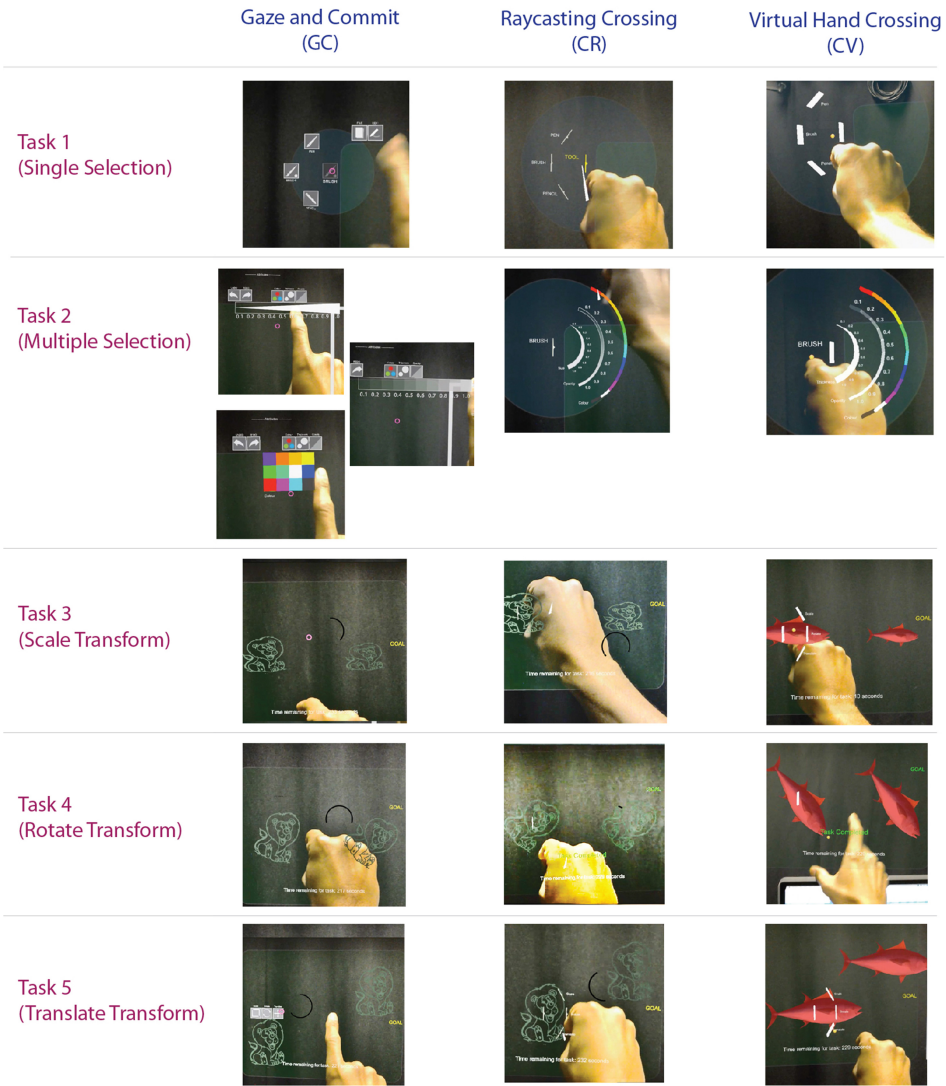
Fig. 13. Tasks as completed with the various selection techniques. Note that Task 2 is accomplished in three steps in GC, compared with one step in both CR and CV.

7.1.1 Experiment Setup. Participants sat facing a black curtain (2 m away) to ensure that the interface elements were clearly visible. In GC and CR, participants were always in their original seated position. In CV, they could lean forward to "touch" the 3D buttons in space. The interface was situated at a distance of 1 m (from the user) to allow for the interface to be visible within the viewport. Technique order was randomized according to a Latin square design. Participants performed two similar training tasks before each main task to familiarize themselves with the selections in each technique. In the first training task, participants could perform the task as many times as they wanted within 1–2 minutes, depending on the length of the task. The second training task was similar to the main task, with a clear objective—e.g., select "Pencil".
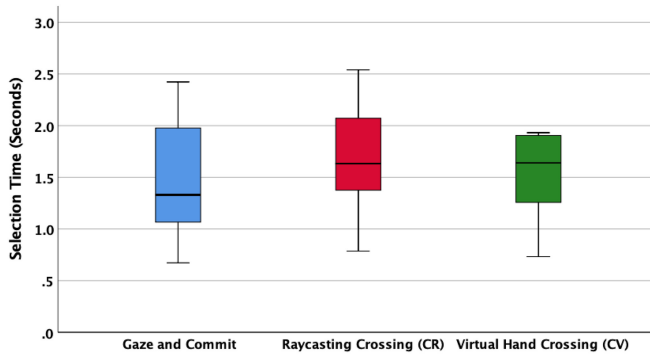
Fig. 14. Results on Task 1, highlighting similar STs across the various techniques.

7.2 Study Metrics

— Selection Time: Time the user completed a task, from initial to final target selection.
— Fatigue: This is the amount of fatigue felt by the participants (neck fatigue in GC [46] and arm fatigue in CR [49]). There is no standard quantitative method for measuring fatigue; although this has been attempted, to some degree, by Hincapié-Ramos et al. [28]. We measured fatigue using a scale ranging from 1 (no fatigue) to 10 (maximum fatigue) after each technique.
— Workload: We used the NASA TLX to explore the workload after each technique.
— User Experience: Using an exit interview, the participants provided further insights into their preferences and experiences regarding using the techniques for the described tasks.

7.3 Recruitment

We recruited 16 participants (9 female, 7 male) from the authors' institution using flyers. Participants included post-doctoral researchers (n = 4), research students (n = 8), and other relatively young professionals. Similar to the previous studies, the participants were asked to rate their experience with general technology and VR/AR (e.g., HoloLens, HTC Vive) on a Likert scale (1 = no experience to 5 = very experienced). The average distribution was 4.5 (sd = 0.74) for general and 2.6 (sd = 1.24) for VR/AR technologies.

7.4 Study 3 Findings and Discussion

For *ST* we used both a RMANOVA and a univariate ANOVA to identify potential differences between techniques, interaction effects, and differences between levels (i.e., Levels 1, 2, and 3 in Task 2) and transforms (i.e., scale, rotate, and translate). Post-hoc comparisons were Bonferroni-corrected using an initial significance level of α = 0.05. In cases where Mauchly's test of sphericity was violated, we applied a Greenhouse–Geiser correction. For count data (e.g., fatigue and workload), a combination of Kruskal–Wallis H tests and Mann–Whitney U tests were used to compare values between the techniques. Regarding user preferences, direct participant quotes are used to provide further insights into the results.

7.4.1 Task1: Single Selection. The study findings revealed similar STs across all techniques (see Figure 14), and we did not find any statistically significant differences between them; this suggests that both raycasting and *virtual hand* crossing are no less efficient than a *head gaze* technique for a comparable single selection operation.
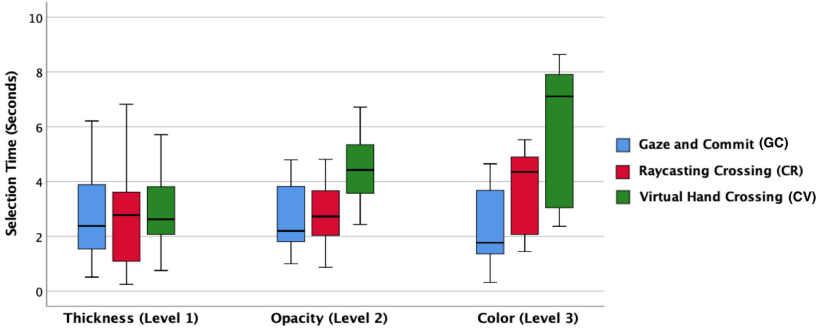
Fig. 15. Results on Task 2—multiple selection of attributes.

7.4.2 Task 2: Multiple Selections. As Figure 15 highlights, the findings are, somewhat, similar for *ST* across all techniques at Level 1 (thickness) and Level 2 (opacity), and we did not find any statistically significant differences between the techniques at these levels. However, **at Level 3 (colour), the findings reveal the contrary** ($F_{2,43}$ = **7.01, $p < 0.01$)**. A post-hoc analysis showed that **the task takes significantly longer to complete with CV, compared to GC ($p < 0.01$)**. We did not find any statistically-significant differences between GC and CR or between CR and CV. Informed by the previous studies, the apparent advantage for GC at Level 3 selection can be attributed to the *Midas touch* problem inherent in the crossing technique, since it does not require an explicit selection action. Users are likely to be more careful when selecting these complex menu items, leading to longer STs.

This explanation can be justified by the fact that the *head gaze* advantage appears to be directly proportional to the degree of the level, that is, the distance travelled by the user's hand. The apparent disadvantage for ST in CV, compared to CR, likely stems from the added effort required to move the hand in the depth (z)-axis, compared to a 2D raycasting scenario.

Further analyses using a RMANOVA **revealed a statistically-significant difference, in *ST*, between the various levels** ($F_{2,86}$ = **9.6, $p < 0.01$, η_p^2 = 0.18)**. Post-hoc analyses confirmed that the **Level 3 selection task takes significantly longer to complete than Level 1 ($p < 0.01$)**. This suggests, as expected, that STs increase the further the user's arm has to travel from the initial selection point. **The results also highlight an interaction effect for [*technique × level*]** ($F_{4,86}$ = **3.8, $p < 0.01$, η_p^2 = 0.15)**, suggesting that STs significantly increase at higher levels depending on the technique used. In this case, the crossing techniques would result in higher STs at longer arm distances than a typical *head gaze*-based selection method.

7.4.3 Tasks 3, 4, and 5: Transformations. In Tasks 3, 4, and 5, we explored *Scale, Rotate,* and *Translate* transformations, respectively—the results are highlighted in Figure 16.

Using univariate ANOVAs, we did not find any statistically-significant differences between the techniques in the *Scale* operation. However, **we found a difference in the *Rotate*** ($F_{2,45}$ = **3.8, $p < 0.05$, η_p^2 = 0.15) and *Translate* transform operations** ($F_{2,45}$ = **6.3, $p < 0.01$, η_p^2 = 0.22)**. According to the post-hoc analyses, **the task took significantly longer to complete in GC than in CR ($p < 0.05$) for the *Rotate* operation. In the *Translate* operation, GC resulted in significantly longer times than both CR ($p < 0.01$) and CV ($p < 0.05$)**. We found no statistically-significant differences between the crossing techniques in either case.

The results on Tasks 4 and 5 are particularly interesting, due to the fact that GC differs significantly from CR (and CV, albeit only in the *Translate* operation), even though all transformations are achieved in a similar manner across techniques. We expect the advantage of the crossing
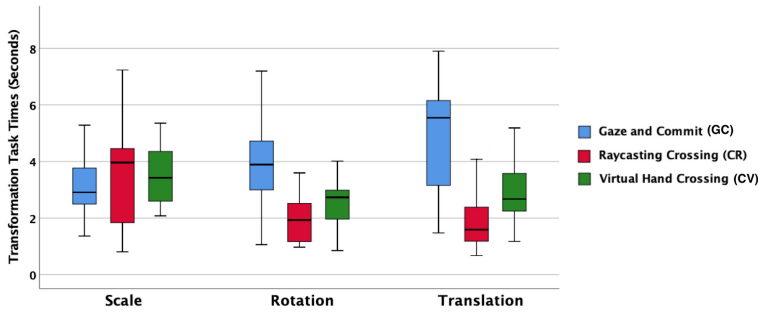
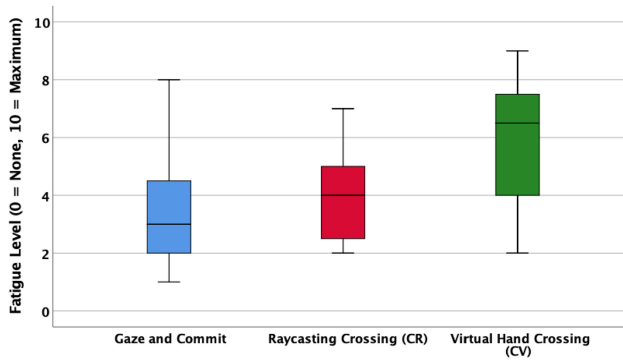Fig. 16. Results on Transformations in Study 3 (Tasks 3, 4, and 5).

Fig. 17. Fatigue as expressed through a 10-point scale (0 = no fatigue at all; 10 = maximum possible).

techniques, in this regard, to be due to the fact that the direction to the target object, with respect to the moving object's initial position, is similar to the crossing direction necessary to activate the operation. Therefore, after activating the operation, the moving object is already in motion toward the target. In contrast, there would likely be a short pause after selecting the moving object (with an *air-tap*) in GC, which potentially extends the task time, compared with the crossing techniques. We expect that this apparent advantage in the crossing techniques could be diminished if, say, the target object were initially to the left, above or below the moving object. This is somewhat evident in the *Scale* operation, which requires different directional strokes following the initial activation strokes—hand movement that is opposite (down and left) to the initial activation stroke direction to reduce the object's size. In Task 4, rotational hand movement is in a clockwise direction toward the goal—this flows naturally from a stroke from left to right.

7.4.4   Fatigue. The results on fatigue are highlighted in Figure 17. A Kruskal–Wallis H test **revealed a statistically significant difference between the different techniques**—$\tilde{\chi}^2$ = **10.3**, p **< 0.01**.

Individual Mann–Whitney U-tests only showed **significantly higher fatigue in CV compared to GC** (p **< 0.01**) **and CR** (p **< 0.05**). It is not surprising to observe higher fatigue in the freehand crossing techniques considering the additional hand movement required in this regard. However, we note that freehand crossing can induce significantly higher fatigue in a *virtual hand* context compared with a raycasting one, according to the participants' subjective perceptions. This result has implications for the design of *in-air* crossing interactions, in the sense that raycasting seems more appropriate in this regard.
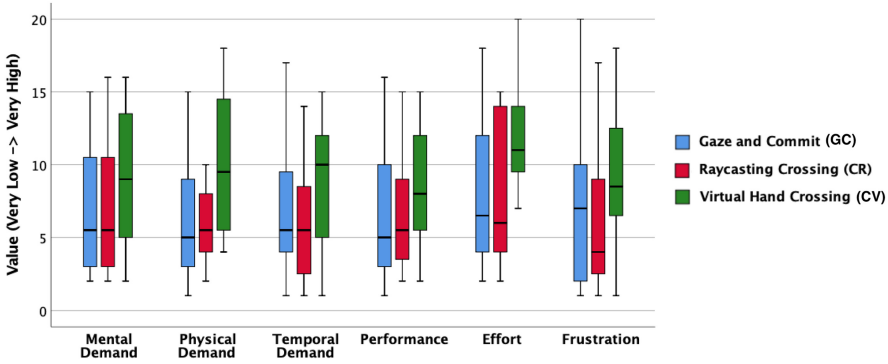
Fig. 18. Subjective workload from the NASA TLX scale.

7.4.5   Workload. The results on subjective workload, in Study 3, are highlighted in Figure 18. We only **found a statistically-significant difference between techniques in the *Physical* component, according to a Kruskal–Wallis H test—$\tilde{\chi}^2$ = 7.63, p < 0.05**.

Multiple Mann–Whitney U tests highlighted that **CV is perceived to be associated with significantly higher workload in the *Physical* (p < 0.05) and *Effort* (p < 0.05) components**. We did not find a significant difference between GC and CR in either case or between CR and CV in both cases. These findings suggest that *raycasting crossing* on a 2D plane is comparable to *head gaze*, from a workload standpoint; however, users perceive *virtual hand crossing* to require more physical effort than both *head gaze* and potentially raycasting crossing, as is evident in Figure 18.

7.4.6   User Experience. Similar to Study 2, user preferences were varied. However, whereas half of the participants preferred GC to the hand-based techniques explored in Study 2, 30% of the participants in Study 3 responded that they would primarily use GC. The most preferred technique in Study 3 was CR (8/15 participants). In this case, these participants appreciated the free-flowing movement in crossing, especially in a drawing application. To this effect, P8 commented: "I feel that [CR] was the best [technique] for this kind of application".

Two main challenges emerged from the participants' feedback regarding the crossing techniques—especially CV. Firstly, 7/16 participants experienced difficulties in judging depth, using CV, as justified by the comments by P8 ("*I struggled most with [CV...]; I could not judge where the buttons were in space, so it required a lot of effort*") and P11 ("*[CR] is much easier than [CV]; I had trouble detecting depth initially*"). Secondly, Task 2 (involving multiple selection) was considered difficult by certain participants, partly due to the size of the radial menu, as noted by P1: "*Multiple selection was very difficult with CV, perhaps if the gap between attributes was larger [...]*". The steering path also seems to have contributed to the task difficulty in both CR and CV. On this point, P7 stated: ("*It is much more energy consuming to be precise [when moving through the steering path]*").

Finally, as was the case in Study 2, preferences for various techniques were contextual, based on the active task, as evidenced by P9's comment: "*For selecting single things, I would prefer [GC], but for multiple objects, I would prefer [CR and CV because they are] pretty cool*".

8   GENERAL DISCUSSION

In this section, we discuss our work in light of the research objectives highlighted in Section 3.

8.1 Feasibility and Efficiency of Freehand Crossing (Objective 1)

We found that raycasting-based freehand crossing (CR) performs as well as (and, in certain cases, outperforms) the state-of-the-art *head gaze*-based method used in current AR HMDs. In Study 1, we found that CR outperforms *head gaze* (GC) at different target directions and depths. The findings from Studies 2 and 3 support those from Study 1 in the sense that CR performs just as well as GC in selection scenarios of varying complexity and dynamicity. We also found that CR is comparable, performance-wise, to a similar *point-and-tap* technique; this supports the findings in the study by Tu et al. [2019], where VR controller-based crossing was found to be comparable to a similar *point-and-click* technique. Although our work differs from Tu et al.'s work in the sense that VR controller-based selection is not identical to freehand selection [5, 35, 49], our work shows that the comparison between CR and *point-and-tap* extends beyond tapping-based tasks (explored in Tu et al. [55]) to radial (expanding) menus and dynamic (moving target) selections as well.

Based on our investigations, in Study 3, freehand crossing can be more efficient than *head gaze* in a raycasting context, than in a *virtual hand*-based (CV) one. We note that additional effort is required in a 3D selection scenario, compared with a 2D raycasting one, especially since certain users may find it challenging to perceive depth in virtual environments. Consequently, selection precision could be diminished (in CV) as a result of factors such as selection distance and path.

8.2 Benefits and Drawbacks of Freehand Crossing (Objective 2)

The advantages of crossing, according to the literature, include the elimination of the need for an explicit selection confirmation action [55] and the ability to select multiple operations in one action [4]. Our work finds that these advantages persist in the AR HMD freehand space, thereby both supporting the literature on crossing in other interaction spaces and providing new insights into the use of different types of crossing-based selections in a freehand AR HMD context.

Unsurprisingly the drawbacks of freehand crossing, as noted in the literature, can also be found in the freehand AR space. For instance, freehand crossing can result in unintentional selections, thereby increasing the potential for error—such issues are exacerbated by arm fatigue [62]. Based on our work, we expect that this potential for error also increase when multiple targets are grouped close together, when additional targets are not anticipated and when the *steering path* to further crossing selections is relatively difficult (e.g., circular rather than linear). Such a result would be difficult to identify in typical two-target selection tasks, such as in Study 1 or in the general literature on selection techniques (e.g., Tu et al. [55]); this is one of the unique contributions of our work.

8.3 User Experience and Preferences for Freehand Crossing (Objective 3)

Most of the studies on selection techniques, in 1D, 2D, and 3D interfaces, focus on selection efficiency and performance. In the crossing domain, researchers have also explored user experience and preferences, albeit mostly in informal circumstances (e.g., Apitz and Guimbretiére [4]). Although our work agrees with the literature that crossing can be an effective and useful selection technique in 3D space [51, 55], we present user preferences, in formal studies (Studies 2 and 3), regarding crossing in various selection scenarios. We found user preferences to vary, depending on the technique and interaction context. In cases where prolonged selections are required in freehand crossing, potentially increasing arm fatigue, users preferred GC. However, for selection operations with a short duration, freehand crossing offers several advantages, as it can be considered natural and intuitive to select objects with the hand. Crossing also offers benefits for quick selection, which users consider important for completing real-world selection tasks. Further, crossing can encourage an enjoyable experience, especially in games and drawing, although there are

challenges for freehand crossing, regarding steering, and in the case of *virtual hand* crossing, perception of depth in 3D scenarios. Finally, the participants expressed the need for the contextual use of different techniques, depending on the interaction scenario. AR HMDs, such as the HoloLens 2, employ eye gaze in addition to head gaze and hand tracking to provide a wealth of opportunities for designers of AR interfaces.

8.4 Design Implications (Objective 4)

Based on the findings from our studies, we highlight the following implications for design regarding the use of freehand crossing in HMD AR selection.

(1) *Leverage the benefits of different techniques*: Combine different selection techniques as necessary to benefit selection performance. For instance, *head gaze* can be used in cases where target amplitudes are large, whereas crossing can be used for local selection operations—for example, marking or drop-down menus, and so on.

(2) *Optimize 3D crossing user interfaces (UIs) for different users*: We recommend that crossing directions be designed in favor of the user's natural abilities and dexterity, for instance, dominant hand—e.g., crossing strokes from left to right for right-handed users.

(3) *Use close and medium range target depths*: Targets within close (100 cm) and medium (175 cm) depths (from the user) result in fewer selection errors, in both *head gaze* and crossing selections, than far (250 cm) targets. Hence, targets should be situated within short depths (i.e., keeping angular size relatively large) to improve selection performance in AR HMDs.

(4) *Reducing fatigue*: Fatigue can be reduced in freehand *in-air* selection scenarios by considering the following: (a) enable relatively short freehand selection (and task) duration, as suggested by Wolf et al. [62], (b) depending on the scenario, suitable rest periods (lasting a few seconds) can be incorporated between crossing interactions, and (c) favor smaller target distances and larger widths in freehand crossing selections.

(5) *Complex target selection*: In accordance with guidelines for designing touch interfaces [63], ensure reasonable spacing between *in-air* interface elements to reduce the probability for accidental selections in multiple selection use cases with freehand crossing.

(6) *Moving target selection*: Crossing is ideal for selecting moving targets since it has no apparent disadvantage, with respect to task time, compared to *head gaze*. Furthermore, crossing results in less errors when selecting moving targets and can be perceived as enjoyable by participants.

(7) *Optimize selection paths*: Although radial menus ensure relatively short selection paths, our Studies (2 and 3) highlight potential drawbacks regarding crossing selection efficiency in radial menus—both static (Study 3) and expanding (Study 2) styles. Therefore, for multiple crossing selections (e.g., attribute selection in Study 3), consider linear scales to reduce difficulties due to prolonged steering.

9 LIMITATIONS AND FUTURE WORK

First, we explored fixed target sizes and layouts in each of the selection scenarios; this was necessary, within the scope of this work, for a valid comparison between the selection techniques. Therefore, our findings should be accepted within the context of the parameters explored in this work.

Second, Fitts' Law shows that the ID is inversely proportional to target width; therefore, it remains unknown how each of the chosen selection techniques perform under scenarios of varying target sizes. Tu et al. [55] provide some foundational insights into controller-based pointing and crossing in VR, and we recommend their work to the reader. However, there are differences in

modality and tracking fidelity between freehand in AR and VR controllers [49]. Therefore, further work is necessary to determine how pointing and crossing compare in the freehand AR space, with varying target parameters.

Third, in Study 1, we investigated freehand crossing only according to a continuous direction-constrained configuration. Accot and Zhai [2] and Tu et al. [55] found subtle differences in crossing selections between amplitude/direction-constrained and discrete/continuous crossing selections in touch and VR controllers, respectively. Future work can explore other types of crossing methods, such as in previous work—e.g., Accot and Zhai [2], Tu et al. [55], using freehand selection in AR HMDs.

Finally, we included only right-handed participants in our studies. Furthermore, our participants were relatively young and had less-than-average experience with VR and AR devices. Therefore, opportunities exist for future work to investigate freehand crossing and *head gaze* performance with a wider variety of users (e.g., left handed, older, and expert VR/AR users).

10   CONCLUSION

Gaze-based methods, for example head and eye gaze, are the norm in HMDs for acquiring targets in mixed reality environments. In most freehand scenarios (e.g., in AR HMDs), hand involvement is normally limited to *air-taps* to confirm selections. However, with the capability of hand tracking in devices, such as the HoloLens, freehand selection techniques are relatively under-explored in such 3D environments. Unencumbered selection without reliance on a handheld physical controller is a critical feature in commercial truly mobile AR OST HMDs, making it worth investigating. Our work (and the wider literature [5, 11, 16, 51, 55]) shows that, despite the disadvantage of increased arm fatigue, it is useful to take advantage of hand-based selection techniques, such as freehand crossing, in 3D UIs. These hand-based techniques are just as efficient as *head gaze* in certain (e.g., static and dynamic) scenarios, as is evident in our work. Further, freehand crossing offers opportunities for enhanced and enjoyable user experiences, for instance, in games and more fluid interactions, such as manipulation of interactive visualizations. We learn that in less-fatiguing interaction situations, for example, employing relatively shorter and sparser hand-based interactions [62], freehand crossing may be preferred by users. The use of hand-based techniques in such contextual use cases could be beneficial to the adoption of these AR technologies.

REFERENCES

[1] Johnny Accot and Shumin Zhai. 1997. Beyond Fitts' law: Models for trajectory-based HCI tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'97)*. ACM Press, 295–302. DOI:https://doi.org/10.1145/258549.258760

[2] Johnny Accot and Shumin Zhai. 2002. More than dotting the i's — foundations for crossing-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Changing Our World, Changing Ourselves (CHI'02)*. ACM Press, 73–80. DOI:https://doi.org/10.1145/503376.503390

[3] Bashar I. Ahmad, James K. Murphy, Patrick M. Langdon, Simon J. Godsill, Robert Hardy, and Lee Skrypchuk. 2016. Intent inference for hand pointing gesture-based interactions in vehicles. *IEEE Transactions on Cybernetics* 46, 4 (April 2016), 878–889. DOI:https://doi.org/10.1109/TCYB.2015.2417053

[4] Georg Apitz and Francois Guimbretiére. 2005. CrossY: A crossing-based drawing application. *ACM Transactions on Graphics* 24, 3 (July 2005), 930–930. DOI:https://doi.org/10.1145/1073204.1073286

[5] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (May 2013), 121–136. DOI:https://doi.org/10.1016/j.cag.2012.12.003

[6] Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2019. The effect of stereo display deficiencies on virtual hand pointing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. ACM Press, 1–14. DOI: https://doi.org/10.1145/3290605.3300437

[7] Anil Ufuk Batmaz, Mayra Donaji Barrera Machuca, Duc Minh Pham, and Wolfgang Stuerzlinger. 2019. Do head-mounted display stereo deficiencies affect 3D pointing tasks in AR and VR?. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR'19)*. IEEE, 585–592. DOI:https://doi.org/10.1109/VR.2019.8797975

[8] Geoffrey P. Bingham, Arthur Bradley, Michael Bailey, and Roy Vinner. 2001. Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments. *Journal of Experimental Psychology: Human Perception and Performance* 27, 6 (2001), 1314–1334. DOI: https://doi.org/10.1037/0096-1523.27.6.1314

[9] 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 (COGAIN'18)*. ACM Press, 1–9. DOI: https://doi.org/10.1145/3206343.3206349

[10] Jennifer Bützler and Christopher Schlick. 2015. Investigation of the angle and age effect using Fitts' Law for mouse input. In *Proceedings of the 19th Triennial Congress of the IEA*. 9–14.

[11] Yeonjoo Cha and Rohae Myung. 2013. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics* 43, 4 (July 2013), 350–355. DOI: https://doi.org/10.1016/j.ergon.2013.05.005

[12] HTC Corporation. 2020. HTC Vive. Retrieved on 16th July, 2021 from https://www.vive.com/uk/.

[13] Microsoft Corporation. 2020. Hololens Technical Specification. Retrieved on 16th July, 2021 from https://docs.microsoft.com/en-us/windows/mixed-reality/hololens-hardware-details.

[14] Microsoft Corporation. 2020. Hololens Website. Retrieved on 16th July, 2021 from https://www.microsoft.com/en-us/hololens.

[15] Nathan Cournia, John D. Smith, and Andrew T. Duchowski. 2003. Gaze- vs. hand-based pointing in virtual environments. In *Proceedings of the CHI'03 Extended Abstracts on Human Factors in Computing Systems (CHI'03)*. ACM Press, 772–773. DOI: https://doi.org/10.1145/765891.765982

[16] Kaushik Das and Christoph Borst. 2010. An evaluation of menu properties and pointing techniques in a projection-based vr environment. In *Proceedings of the 2010 IEEE Symposium on 3D User Interfaces (3DUI'10)*. 47–50. DOI: https://doi.org/10.1109/3DUI.2010.5444721

[17] Morgan Dixon, Francois Guimbretiére, and Nicholas Chen. 2008. Optimal parameters for efficient crossing-based dialog boxes. In *Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. ACM Press, 1623–1632. DOI: https://doi.org/10.1145/1357054.1357307

[18] Pierre Dragicevic. 2004. Combining crossing-based and paper-based interaction paradigms for dragging and dropping between overlapping windows. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST'04)*. ACM Press, 193–196. DOI: https://doi.org/10.1145/1029632.1029667

[19] John J. Dudley, Hendrik Schuff, and Per Ola Kristensson. 2018. Bare-handed 3D drawing in augmented reality. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 (DIS'18)*. ACM Press, 241–252. DOI: https://doi.org/10.1145/3196709.3196737

[20] LLC Facebook Technologies. 2020. Oculus Website. Retrieved on 16th July, 2021 from https://www.oculus.com/.

[21] Paul M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 6 (1954), 381–391. DOI: https://doi.org/10.1037/h0055392

[22] Tovi Grossman and Ravin Balakrishnan. 2004. Pointing at trivariate targets in 3D environments. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI'04)*. ACM Press, 447–454. DOI: https://doi.org/10.1145/985692.985749

[23] Tovi Grossman and Ravin Balakrishnan. 2006. The design and evaluation of selection techniques for 3D volumetric displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST'06)*. ACM Press, 3–12. DOI: https://doi.org/10.1145/1166253.1166257

[24] Francois Guimbretiére and Terry Winograd. 2000. FlowMenu: Combining command, text, and data entry. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST'00)*. ACM Press, 213–216. DOI: https://doi.org/10.1145/354401.354778

[25] Tyler J. Gunn, Pourang Irani, and John Anderson. 2009. An evaluation of techniques for selecting moving targets. In *Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA'09)*. ACM Press, 3329–3334. DOI: https://doi.org/10.1145/1520340.1520481

[26] Abir Al Hajri, Sidney Fels, Gregor Miller, and Michael Ilich. 2011. Moving target selection in 2D graphical user interfaces. In *Proceedings of the Human–Computer Interaction (INTERACT'11)*. David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler (Eds.), Vol. 6947. Springer, Berlin Heidelberg, 141–161. DOI: https://doi.org/10.1007/978-3-642-23771-3_12

[27] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Baekgaard. 2018. A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction (COGAIN'18)*. ACM Press, 1–5. DOI: https://doi.org/10.1145/3206343.3206344

[28] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI'14)*. ACM Press, 1063–1072. DOI: https://doi.org/10.1145/2556288.2557130

[29] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A survey of design issues in spatial input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology (UIST'94)*. ACM Press, 213–222. DOI: https://doi.org/10.1145/192426.192501

[30] Gabe Johnson, Mark D. Gross, and Ellen Yi-Luen Do. 2006. Flow selection: A time-based selection and operation technique for sketching tools. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI'06)*. ACM Press, 83–86. DOI: https://doi.org/10.1145/1133265.1133281

[31] Hansol Kim, Yoonkyoung Kim, and Eui Chul Lee. 2014. Method for User Interface of Large Displays Using Arm Pointing and Finger Counting Gesture Recognition. *The Scientific World Journal* 2014, (2014), 1–9. DOI: https://doi.org/10.1155/2014/683045

[32] Arun Kulshreshth and Joseph J. LaViola. 2014. Exploring the usefulness of finger-based 3D gesture menu selection. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI'14)*. ACM Press, 1093–1102. DOI: https://doi.org/10.1145/2556288.2557122

[33] Kiron Lebeck, Tadayoshi Kohno, and Franziska Roesner. 2016. How to safely augment reality: Challenges and directions. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications (HotMobile'16)*. ACM Press, 45–50. DOI: https://doi.org/10.1145/2873587.2873595

[34] Juha Lehikoinen and Mika Röykkee. 2003. 1D selection of 2D objects in head-worn displays. *Personal and Ubiquitous Computing* 7, 1 (May 2003), 44–52. DOI: https://doi.org/10.1007/s00779-002-0212-9

[35] Jialei Li, Isaac Cho, and Zachary Wartell. 2018. Evaluation of cursor offset on 3D selection in VR. In *Proceedings of the Symposium on Spatial User Interaction (SUI'18)*. ACM Press, 120–129. DOI: https://doi.org/10.1145/3267782.3267797

[36] Paul Lubos and Frank Steinicke. 2014. Analysis of direct selection in head-mounted display environments. In *Proceedings of the 2014 IEEE Symposium on 3D User Interfaces (3DUI'14)*. 11–18. DOI: https://doi.org/10.1109/3DUI.2014.6798834

[37] Yuexing Luo and Daniel Vogel. 2014. Crossing-based selection with direct touch input. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI'14)*. ACM Press, 2627–2636. DOI: https://doi.org/10.1145/2556288.2557397

[38] Yuexing Luo and Daniel Vogel. 2015. Pin-and-cross: A unimanual multitouch technique combining static touches with crossing selection. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST'15)*. ACM Press, 323–332. DOI: https://doi.org/10.1145/2807442.2807444

[39] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The effect of offset correction and cursor on mid-air pointing in real and virtual environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM,1–13. DOI: https://doi.org/10.1145/3173574.3174227

[40] Atsuo Murata and Hirokazu Iwase. 2001. Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science* 20, 6 (Dec. 2001), 791–805. DOI: https://doi.org/10.1016/S0167-9457(01)00058-6

[41] Takashi Nakamura, Shin Takahashi, and Jiro Tanaka. 2008. Double-crossing: a new interaction technique for hand gesture interfaces. In *Proceedings of the 8th Asia-Pacific Conference on Computer-Human Interaction*. Seongil Lee, Hyunseung Choo, Sungdo Ha, and In Chul Shin (Eds.), Vol. 5068, Springer, Berlin Heidelberg, 292–300. DOI: https://doi.org/10.1007/978-3-540-70585-7_33

[42] Alex Olwal and Steven Feiner. 2003. The flexible pointer: an interaction technique for selection in augmented and virtual reality. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. 81–82.

[43] Robert Earl Patterson. 2012. Human factors of 3D displays. In *Proceedings of the Handbook of Visual Display Technology*. Janglin Chen, Wayne Cranton, and Mark Fihn (Eds.), Springer, Berlin Heidelberg,1815–1822. DOI: https://doi.org/10.1007/978-3-540-79567-4_109

[44] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology (UIST'96)*. ACM Press, 79–80. DOI: https://doi.org/10.1145/237091.237102

[45] Timothy G. Purdy and Young Mi Choi. 2014. Enhancing augmented reality for use in product design. In *Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI EA'14)*. ACM Press, 1303–1308. DOI: https://doi.org/10.1145/2559206.2581251

[46] 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 (SUI'17)*. ACM Press, 91–98. DOI: https://doi.org/10.1145/3131277.3132182

[47] Robert G. Radwin, Gregg C. Vanderheiden, and Mei-Li Lin. 1990. A method for evaluating head-controlled computer input devices using fitts' law. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 32, 4 (Aug. 1990), 423–438. DOI: https://doi.org/10.1177/001872089003200405

[48] Adrian Ramcharitar and Robert Teather. 2018. Ezcursorvr: 2D selection with virtual reality head-mounted displays. In *Proceedings of the ACM Graphics Interface Conference 2018 (GI'18)*.

[49] Gang Ren and Eamonn O'Neill. 2013. 3D selection with freehand gesture. *Computers & Graphics* 37, 3 (May 2013), 101–120. DOI: https://doi.org/10.1016/j.cag.2012.12.006

[50] Immo Schuetz, T. Scott Murdison, Kevin J. MacKenzie, and Marina Zannoli. 2019. An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. ACM Press, 1–13. DOI : https://doi.org/10.1145/3290605.3300765

[51] Huang Susu, Qi Daqing, Yuan Jiabin, and Tu Huawei. 2019. Review of studies on target acquisition in virtual reality based on the crossing paradigm. *Virtual Reality & Intelligent Hardware* 1, 3 (2019), 251–264. DOI : https://doi.org/10.3724/SP.J.2096-5796.2019.0006

[52] 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 (CHI'00)*. ACM Press, 265–272. DOI : https://doi.org/10.1145/332040.332443

[53] Robert Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3d targets in a stereo head-tracked virtual environment. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces (3DUI'11)*. 87–94. DOI : https://doi.org/10.1109/3DUI.2011.5759222

[54] Unity Technologies. 2020. Unity Technologies. Retrieved from https://unity.com/.

[55] Huawei Tu, Susu Huang, Jiabin Yuan, Xiangshi Ren, and Feng Tian. 2019. Crossing-based selection with virtual reality head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*. ACM Press, 1–14. DOI : https://doi.org/10.1145/3290605.3300848

[56] D. W. F. Van Krevelen and R. Poelman. 2010. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality* 9, 2 (Jan. 2010), 1–20. DOI : https://doi.org/10.20870/IJVR.2010.9.2.2767

[57] Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. 1997. Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In *Proceedings of the Human–Computer Interaction (INTERACT'97)*. Steve Howard, Judy Hammond, and Gitte Lindgaard (Eds.), Springer US, 509–516. DOI : https://doi.org/10.1007/978-0-387-35175-9_77

[58] Daniel Vogel and Ravin Balakrishnan. 2005. Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST'05)*. ACM Press, 33–42. DOI : https://doi.org/10.1145/1095034.1095041

[59] Alexandra Voinea, Alin Moldoveanu, and Florica Moldoveanu. 2016. Bringing the augmented reality benefits to biomechanics study. In *Proceedings of the 2016 Workshop on Multimodal Virtual and Augmented Reality (MVAR'16)*. ACM Press, 1–6. DOI : https://doi.org/10.1145/3001959.3001969

[60] Alla Vovk, Fridolin Wild, Will Guest, and Timo Kuula. 2018. Simulator sickness in augmented reality training using the Microsoft HoloLens. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*. ACM Press, 1–9. DOI : https://doi.org/10.1145/3173574.3173783

[61] Jacob O. Wobbrock and Krzysztof Z. Gajos. 2008. Goal crossing with mice and trackballs for people with motor impairments: performance, submovements, and design directions. *ACM Transactions on Accessible Computing* 1, 1 (May 2008), 1–37. DOI : https://doi.org/10.1145/1361203.1361207

[62] Dennis Wolf, John Dudley, and Per Ola Kristensson. 2018. Performance envelopes of in-air direct and smartwatch indirect control for head-mounted augmented reality. In *Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces*. DOI : https://doi.org/10.17863/CAM.25799

[63] Shota Yamanaka. 2018. Effect of gaps with penal distractors imposing time penalty in touch-pointing tasks. In *Proceedings of the 20th International Conference on Human–Computer Interaction with Mobile Devices and Services*. ACM,1–11. DOI : https://doi.org/10.1145/3229434.3229435

[64] Dianna Yim, Garance Nicole Loison, Fatemeh Hendijani Fard, Edwin Chan, Alec McAllister, and Frank Maurer. 2016. Gesture-driven interactions on a virtual hologram in mixed reality. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces (ISS Companion'16)*. ACM Press, 55–61. DOI : https://doi.org/10.1145/3009939.3009948

[65] Chuang-Wen You, Yung-Huan Hsieh, Wen-Huang Cheng, and Yi-Hsuan Hsieh. 2014. AttachedShock: Design of a crossing-based target selection technique on augmented reality devices and its implications. *International Journal of Human–Computer Studies* 72, 7 (July 2014), 606–626. DOI : https://doi.org/10.1016/j.ijhcs.2014.03.001

[66] Xiaolu Zeng, Alan Hedge, and Francois Guimbretiére. 2012. Fitts' law in 3D space with coordinated hand movements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (Sept. 2012), 990–994. DOI : https://doi.org/10.1177/1071181312561207

[67] Shumin Zhai, William Buxton, and Paul Milgram. 1994. The "Silk Cursor": Investigating transparency for 3D target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence (CHI'94)*. ACM Press, 459–464. DOI : https://doi.org/10.1145/191666.191822