2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3315-1647-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISMAR62088.2024.00046

Whirling Interface: Hand-based Motion Matching Selection for Small Target on XR Displays

Juyoung Lee* KAIST UVR Lab.
Seo Young Oh† KAIST UVR Lab.
Minju Baeck‡ KAIST UVR Lab.
Hui Shyong Yeo§ Huawei
Hyung-il Kim¶ KI-ITC ARRC
Thad Starner‖ Georgia Institute of Technology
Woontack Woo** KAIST UVR Lab. KI-ITC ARRC

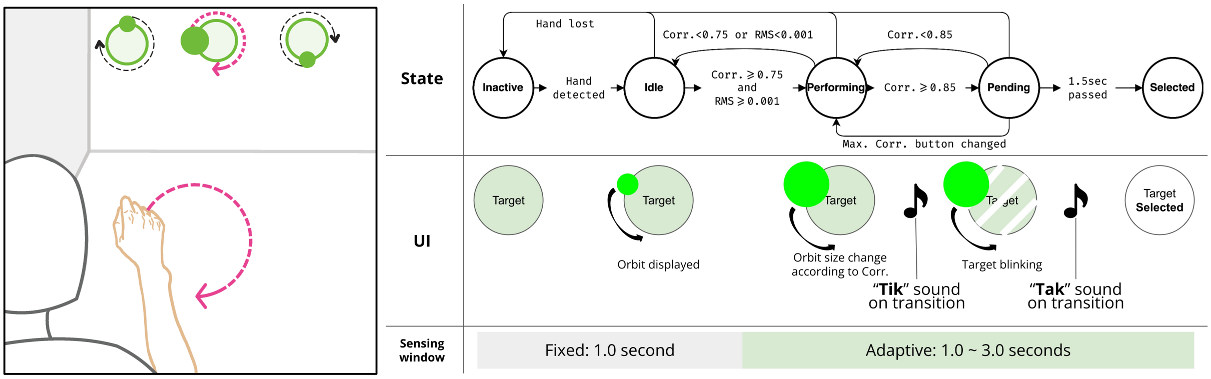


Figure 1: (left) The Whirling Innterface enables users to make selections by synchronizing their hand motion with the orbiting movement of targets. (right) The system consists of five distinct states, each transitioning to the next based on specific conditions and accompanied by relevant visual and audio feedback.

Abstract

We introduce "Whirling Interface," a selection method for XR displays using bare-hand motion matching gestures as an input technique. We extend the motion matching input method, by introducing different input states to provide visual feedback and guidance to the users. Using the wrist joint as the primary input modality, our technique reduces user fatigue and improves performance while selecting small and distant targets. In a study with 16 participants, we compared the whirling interface with a standard ray casting method using hand gestures. The results demonstrate that the Whirling Interface consistently achieves high success rates, especially for distant targets, averaging 95.58% with a completion time of 5.58 seconds. Notably, it requires a smaller camera sensing field of view of only 21.45° horizontally and 24.7° vertically. Participants reported lower workloads on distant conditions and expressed a higher preference for the Whirling Interface in general. These findings suggest that the Whirling Interface could be a useful alternative input method for XR displays with a small camera sensing FOV or when interacting with small targets.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction techniques—Gestural input; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality;

1 Introduction

The advancement of smartglasses, exemplified by product lines such as Google Glass, Vuzix Z100, and more immersive extended reality (XR) headsets like MS HoloLens 2, Meta Quest Pro, and Apple Vision Pro, has made the technology more compact and suitable for all-day wear. However, a significant challenge remains in developing effective input methods tailored to the everyday use of XR devices. Current interaction techniques for XR, such as touchpads on the temple, voice commands, or external controllers, all have their own limitations. While touchpads are well-suited for on-screen navigation, they are less effective for selecting objects distributed in 3D space. Voice commands can be a suitable choice for launching applications, and recent research suggests the possibility of silent voicing for control [28, 32]. However, voice may not be the most appropriate modality for selecting specific objects. External controllers can be powerful, and numerous techniques have been proposed to select small or distant objects [6, 7]; nevertheless, carrying external controller on a daily basis is often inconvenient.

Given these challenges, bare-hand input is widely considered the most natural and convenient interaction method for 3D object selection [3]. Casting a ray in the direction indicated by the hand has become a common approach for object selection. However, according to Fitts' Law [12], the difficulty of selecting small or distant objects increases significantly. The quality of hand tracking also plays a crucial role [4, 8]. In addition, it requires an additional gesture, such as a tap, pinch, or dwell, to confirm the selection [17], which does not guarantee success as it can be affected by tracking loss or self-occlusion. Eye-based pointing and confirming with hand gestures could be a possible solution, as a user can keep their hand in a stable position to track. However, gaze-based pointing has not shown acceptable accuracy for selecting small and distant targets [47, 50]. There were also trials to use multi-modal gaze-hand input to address this issue [39, 61], but it increases the cognitive load. In addition, hardware limitations related to form factor, battery

*e-mail: ejuyoung@kaist.ac.kr
†e-mail: seoyoung.oh@kaist.ac.kr
‡e-mail: minjnubaeck@kaist.ac.kr
§e-mail: yeo.hui.shyong@huawei.com
¶e-mail: hyungil@kaist.ac.kr
‖e-mail: thad@gatech.edu
**e-mail: wwoo@kaist.ac.kr

979-8-3315-1647-5/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR62088.2024.00046

consumption, and price remain a concern. Multi-step selection is another possible solution [31,53,54], but requires additional attention or steps, which, although beneficial for precise selection, can be disadvantageous for quick interactions in everyday life.

In response to these issues, and inspired by the potential of motion matching interaction methods [19, 29, 37, 40, 59], this paper introduces the Whirling Interface, a motion matching-based selection interface for distant object selection. We implemented the Whirling Interface with user feedback, addressing Norman's concerns with gesture interfaces [44]. The proposed system utilizes motion matching with user's hand, providing a simple method for selecting distant targets by mirroring an on-screen orbital motion with hand gestures.

To evaluate the effectiveness of the Whirling Interface, we conducted a comprehensive experiment under varying locations of small targets. Our result provides a blueprint for assessing the efficacy and suitability of the Whirling Interface for typical XR usage. Through our investigation, Whirling demonstrated a high success rate of over 95% in both near (95.63%) and far (95.53%) conditions, with consistent completion times of 5.58 seconds on average, respectively. At the same time, participants reported that Whirling required less workload than ray casting for far targets and was competitive for near-distance targets. Additionally, Whirling required a relatively smaller and more consistent range of camera sensing field of view (FOV), averaging 21.45° and 24.7° on the X-axis and Y-axis, respectively. In contrast, ray casting required a larger sensing FOV, averaging 29.15° and 43.2°. The robustness of the Whirling Interface in handling self-occlusion and minor tracking errors further contributes to its potential as a reliable input method for XR.

In our discussion, we underscore the potential of the Whirling Interface as a viable alternative for selecting small or distant objects, a task of significant importance in the everyday use of XR. Our findings show that the Whirling Interface offers consistent performance across different display sizes and distances, with users demonstrating the ability to keep their gesture trajectories steady despite these variations. Notably, Whirling shows comparable performance to ray casting for near targets while outperforming it for far targets. Given these advantages, we envision Whirling serving as an effective method for quick selections or as a complementary technique alongside traditional selection methods.

We can summarize the contributions in three parts. First, we designed the Whirling Interface, which extends the motion matching-based selection technique to interface with feedback and confirmation features. Second, we conduct a comparative performance analysis of the Whirling Interface with the ray casting technique, providing valuable insights into the effectiveness of motion matching-based selection. Third, we investigate the advantages of the Whirling Interface in terms of sensing capabilities, highlighting its robustness in handling self-occlusion and minor tracking errors. Our findings on Whirling may inform future research on XR experiences, particularly with smaller targets or XR devices with sensing constraints.

2 Related Work

2.1 Input Methods for Smartglasses/XR devices

Smartglasses often use a touchpad on the temple for interaction, as seen with commodity devices such as Google Glass and Vuzix Blade, but this method has limitations beyond basic screen navigation. To address this issue, attempts have been made to extend the touchpad functionality through the use of swipe gestures [24, 67] or by adding other modalities such as voice or gaze [1].

Concurrently, research into the use of gestures employing inertial measurement units (IMUs) integrated within smartglasses has also been undertaken. This exploration includes head gestures [65, 66], or combining head gestures with movements from other body parts to mitigate false positive errors [36].

Hand tracking, as proposed by early research in augmented reality [57], is arguably the most natural form of input. Recent commodity devices have incorporated hand tracking features and continue to improve on them. Studies have explored implementing hand tracking using monochrome cameras [25] or multiple views to extend the field of view (FOV) and address self-occlusion [26]. Yet, interacting with virtual objects requires further steps, such as grasping gesture detection [2] or adding an appendage to prevent occlusion by the fingertip [58]. Therefore, hand tracking to interact with small and distant objects necessitates precise tracking and detection to provide a usable experience.

2.2 Object Selection at a Distance

Various techniques have been extensively explored in the literature for the selection of distant objects [49]. Among these, ray casting emerges as a straightforward approach akin to using a laser pointer in the physical world [23, 60]. However, ray casting suffers from certain limitations, including issues with precision and the Heisenberg effect [63], which make it challenging to select small objects located at a distance and when selecting between multiple closely positioned objects. IntenSelect+ combines score based selection with intention to overcome these limitations [33]. However, the presence of intervening objects can obstruct the path of the ray, which requires additional methods to select objects located behind others [7, 68]. Other hand-based object selection techniques such as Go-Go [48], HOMER [11], scaled-world grab [43] are also common in the HCI and VR literature. However, Go-Go may have limitations due to the limited range of hand movement, while HOMER may face challenges related to occlusion when multiple objects are present. Scaled-world grab may encounter difficulties when manipulating virtual objects that are significantly larger or smaller in scale compared to the user's hand. Researchers also explored multimodal techniques such as combining gaze with hand [38, 61], gaze with head [35, 50], or gaze with voice. However, challenges arise in coordinating and integrating multiple modalities, managing user fatigue and cognitive load, and addressing environmental factors that can affect performance. Additionally, the learning curve and complexity simultaneously pose considerations for user training and guidance.

2.3 Display-Guided Interaction

The concept of identifying a user's intended target without directly pointing has been explored repeatedly. An early exploration of this concept involved studies of continuous user input to detect the intended object [18] and motion pointing with a mouse [45].

As wearable computing devices became popular, research has increasingly focused on gaze-based interactions. The pioneering work of Vidal et al. [42] introduced the concept of selecting items by tracking eye gaze. Later, Esteves et al. [21] demonstrated the effectiveness of gaze interaction with smartwatches, establishing the robustness of motion matching with false positive errors. Further advances in gaze-based smooth-pursuit interaction have allowed larger targets [46], combined with eye-typing [5]. In addition, there were studies to expand the understanding of smooth-pursuit interactions, Gafna et al. explored performance with various target speeds and trajectories [16], and Esteves et al. compared with other methods such as clickers and dwell in studies [20].

Numerous studies have explored the adaptability of different devices to varying contexts, such as virtual reality inputs for gaming or selecting occluded objects [30, 56]. Research has also been extended to the use of electrooculography (EOG) sensors, a cost-effective alternative to vision-based eye-tracking modules [15, 55]. Esteves et al. demonstrated the viability of head movement-based interactions in augmented reality headsets [22]. MatchPoint presented an approach that allows spontaneous spatial interaction across various devices [14]. This approach has been further customized for smartwatches, employing gestures such as rotation [34], swing [64], and magnet-detected thumb movements [51] to follow targets. The domain of mobile devices has also witnessed the implementation of display-guided interactions, facilitated by tapping [10] or touch gestures on the screen [19]. When it comes to hand-based motion matching interaction, PathSync [29] and TraceMatch [13, 37] have shown potential using a computer vision-based tracking system, while WaveTrace [40] demonstrated applicability using a smartwatch.

However, the emphasis in these works was on exploring the new input modality where feedback was not a priority. This lack may result in potential problems as users had to perform gestures without feedback for approximately three seconds (lack of feedback while performing the gesture in gesture interfaces is highlighted as a problem by Norman [44]). For this reason, we plan to investigate hand-based motion-matching interaction feedback as part of the interface, which aligns with Norman's principles for gesture interfaces and only requires a single head-worn XR device.

3 Interaction methods

As hand tracking technology has matured, hand-based midair interactions have become increasingly prevalent. Many companies are now introducing XR headsets that incorporate hand interactions, as one of most intuitive interaction techniques that do not require any external devices apart from the headset itself. However, when it comes to small or distant targets, challenges arise. Several approaches have been proposed to address these challenges, primarily by introducing additional steps [54] or modalities [39, 61]. Although these modifications can improve precision, they can potentially conflict with Norman's principles on gesture interaction [44]. Norman emphasized that gestural systems should provide feedback, clear cues for possible actions, and guidance. Taking this into consideration, we decided to use a ray casting system that addresses these principles and create a motion-matching interface that adheres to the principles.

3.1 Ray casting

First, as a guide, we enable a hand-initiated line ray whenever the device detects the right hand. We opt for a straight line pointer due to its directness and simplicity compared to alternative options like the sticky or curved ray with scoring the target object [33]. To further enhance user interactivity, we implemented a visual highlight on a target when the ray intersects with it, clearly indicating potential interaction points. As users start to finalize the input with the gesture, the ray changes its texture from a dotted line to a solid one, while simultaneously reducing the size of the intersecting point circle, serving as real-time feedback. For the concluding gesture, we decided on the "pinch" gesture. It is widely preferred by numerous companies because of its self-haptic feedback, rendering it more instinctive compared to gestures that lack this feedback, like air tap or push. Furthermore, research has indicated that it boasts the highest bits-per-second (bps) [17]. To implement an undo function, we confirm the input once the pinch gesture concludes when the fingers separate. This technique allows users to negate unintended activation by simply moving the ray out of the target before ending the pinch.

As for the toolkit to detect hand pointing and pinch gestures, we used the Mixed Reality Toolkit (MRTK) with Microsoft HoloLens 2. All our design elements were compatible with the features available in MRTK version 2.7.3[1]. Our pilot tests confirmed its accuracy in both pointing and pinching.

3.2 Motion Matching: Whirling Interface

Motion matching input is a selection technique activated when the user matches the movement of their input modality with moving targets. It originates from intuitive selection with smooth eye pursuit movements [21] and now extends to hand motions [29], or touch gestures [19].

We aimed to build an interface for the motion matching method and focused on improving the interactivity. We introduced five interaction states, ensuring flexibility and user-friendliness, with an included cancel or confirmation step - mirroring the same number of steps in the ray casting method we employed. Furthermore, we leveraged an adaptive time window mechanism that activates when user intent is detected, providing a responsive, real-time interaction experience. The subsequent sections will provide a more in-depth look at the design considerations and technical details of our system, termed the "Whirling Interface".

3.2.1 Design

Input modality. We strived to design an input method that does not require an external controller, as more and more XR headsets rely on vision-based hand tracking. Our design does not include eye-gaze or head tracking, and several considerations led us to this decision. Primarily, recognizing interactive elements without visual cues poses a challenge to the user. In cases where cues like cursor guidance or target illumination are provided, users still need to execute extra gestures to initiate input. Otherwise, these interactive indicators may pop up even when interaction is not intended, potentially intruding on the user's experience. Secondly, including the hand-based, all ray casting selection methods relies on the tracking quality. Commodity devices typically employ infrared cameras for tracking, which can be problematic in outdoor environments due to interference from sunlight [41]. Moreover, when dealing with distant targets, any tracking errors or hand jitter are amplified by the ray casting method, making it difficult for users to accurately select the intended target. This can lead to increased user frustration and reduced performance, hindering the overall usability of the system. In addition, requiring users to repeat head or eye movements for target selection can be demanding loads, impractical, and does not promise accurate selection for small targets. To address this issue, hand gestures are widely used to finalize the selection process. However, this approach still relies on dual modalities, such as combining head or eye movements with hand gestures, which can increase interaction costs. The use of multiple modalities may result in a higher workload for users, potentially affecting the overall efficiency and usability of the system. Therefore, it is essential to explore alternative input methods that minimize the reliance on dual modalities while maintaining or enhancing user performance.

Hand-based interactions offer inherent advantages for repeated use, as they are familiar, intuitive, and natural for users. We specifically chose to utilize the wrist joint instead of finger joints due to its greater stability during tracking losses, which are often caused by self-occlusions from the egocentric viewpoint of the camera. The wrist joint provides a more reliable and consistent input compared to finger joints, as it is less likely to be occluded by other parts of the hand or the user's body. This stability is crucial for maintaining accurate and precise input, especially in scenarios where the user's hands may obstruct the camera's view.

Feedback. We placed a single orbital target on the same orbit to avoid confusion during interaction. To ensure that users have a clear understanding of their performance, we have implemented a system with three different feedback states. These states indicate whether the user is able to interact with the target, is performing well, and can finalize or cancel the selection as illustrated in Fig. 1.

The first feedback appears when the user's hand is detected on the device. The system starts showing the moving orbit around the target. Then, the second feedback is to give guidance that they are performing well by changing the orbit's size. When the maximum correlation coefficient of targets (C_{max}) exceeds the minimum threshold, the orbit that shows the maximum coefficient changes its size according to the coefficient value. Finally, when C_{max} exceeds the

[1] Buttons — MRTK2 ,https://learn.microsoft.com/windows/mixed-reality/mrtk-unity/mrtk2/features/ux-building-blocks/button

higher threshold, the target starts blinking to inform the user that it will be selected in a few seconds. If the blinking target does not match the user's intent, the user has the opportunity to cancel the input. At the same time of the blinking start and end, we added auditory feedback to ensure the user, even if the blinking selection is not in their view. Our design addresses one of Norman's gesture interface critiques by meticulously curating feedback mechanisms. It ensures users consistently understand their interaction progress and can effortlessly correct their actions if necessary.

3.2.2 Techniques

As the orbital movement of the target varies on a 2D plane, we chose to use only the X and Y axes of the tracked hand. The idea behind this decision was to create a system that could be easily used on various devices, even those without built-in hand-tracking capabilities. By focusing on 2D hand position tracking, our approach can be implemented using open-source software like Google MediaPipe using a single front-facing camera.

The system collects the position of the right wrist in relation to the user's head. This strategy was implemented to minimize tracking complications arising from user movement and to cater to situations where the user may be in motion, such as walking. Similarly, we decided to use a displayed position coordinate anchored to the device instead of a position based on real-world coordinates for the orbital target. This choice allows for a consistent target position relative to the display, regardless of the user's location or orientation in the physical environment. To synchronize user input and target movement, we collected user input by matching the display frame timestamps, ensuring that the input and target positions correspond to the same moment. Then, we calculated the Pearson correlation coefficient for each axis and used the average value for detection.

To enhance the system's responsiveness while reducing false activations for motion matching input, we made several technical modifications. Firstly, to ensure immediate feedback about performance, we chose to use a relatively short sliding time window ($Timewindow_{min}$) of 1 second. Using a shorter window can increase responsiveness, but it also increases the risk of false positive errors, particularly when the system displays multiple orbits simultaneously. It is possible to have highly similar movement patterns when focusing on a small time window. Fig. 2 shows an example of the correlation coefficient calculated using different sliding time windows. The bold blue line represents the target, while the others represent the 11 targets displayed simultaneously. As depicted, the middle plot, with a one-second window length, shows that the correlation coefficient of a non-intended target exceeds that of the intended target at around 1.5 to 2.0 seconds. Using a longer window of 3 seconds shows a stable correlation tendency but requires the user to wait for initial feedback. This limitation can potentially lower the system's responsiveness and negatively affect the user experience.

To counteract false positives, we implemented a dual-threshold system with an adaptive sliding time window. When the correlation coefficient surpasses a lower threshold ($THRE_{min}$), it is counted as an intended trial, and the time window for calculation expands up to the upper limit of the time window ($Timewindow_{max}$). This adaptive approach helps the user reach the higher threshold ($THRE_{max}$) by extending the time window as more time passes. Nevertheless, the lower threshold can still result in false intentions during actions such as raising a hand or gesturing for content searching. To mitigate this issue, we introduced a root mean square (RMS_{xyz}) threshold to filter out these unintended actions. Additionally, certain user gestures could resemble multiple orbital movements when performed with similar periods and phases. To address this, we assess the differences between the maximum correlation coefficient (C_{max}) and the second-highest coefficient (C_{2nd}) at the final phase of the gesture. If the difference is below a certain threshold, the gesture is considered ambiguous and is not recognized as a valid input. Finally, we provided a 1.5-second confirmation time (Δt) from the last threshold check to the final confirmation. This delay allows users to undo their input if needed, enhancing the interface's flexibility and user control. These procedures, along with the corresponding feedback changes, are outlined in Fig. 1.

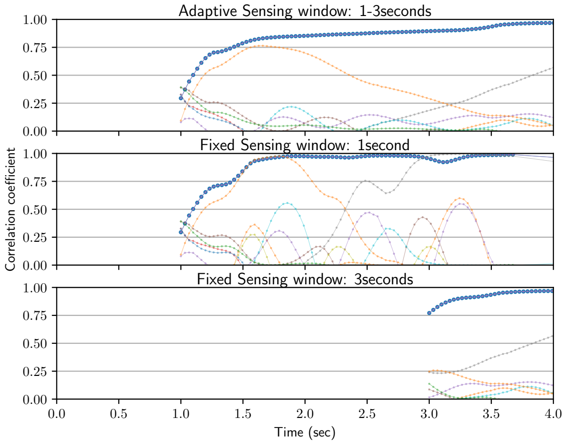


Figure 2: Each graph shows the correlation coefficient for the main target as a bold blue line, while the other 11 targets are displayed in various colors. The top graph features an adaptive time window ranging from 1 to 3 seconds, while the bottom two graphs show fixed sliding time windows of 1 and 3 seconds. The timeline indicates when the device starts to track hands.

4 Experiment

Through the experiment, we aimed to compare the performance of two input methods on the same device and hand-tracking system, observe user movements, and identify potential areas for improvement. The experiment was conducted in a controlled lab environment using a within-subject design. We recruited 16 right-handed participants (9 male, 7 female) through a university online noticeboard. Participant ages ranged from 18 to 28 years ($M = 23.4, SD = 3.2$). Prior experience with smartglasses, including XR headsets, varied: five participants had no previous experience, ten had used them less than three times, and only one more than ten times. Regarding bare-hand input experience, 12 participants had no prior experience, three had tried it less than three times, and one used it regularly.

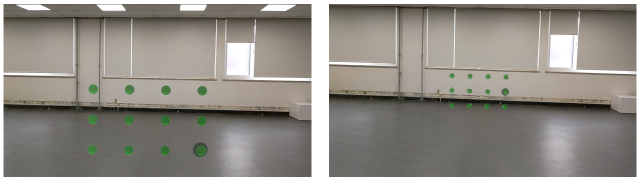
4.1 Apparatus

We used the Holographic Remoting Player[2] to connect the HoloLens device to a PC via an external network router and collected data on the PC. For both methods, we adapted the color of the MRTK's circular button to green for clarity which lowers visual stuttering from the color-sequential display of Microsoft HoloLens 2[3].

We extracted the wrist joint data from the MRTK and established the lower and higher thresholds at 0.75 and 0.85. We did not run personal fitting for neither hand pinch gesture nor Whirling, so the thresholds are set by the most used threshold in previous research, 0.80. We provided two different gesture speeds, 180 and 240°/sec, which require 2 and 1.5 seconds per cycle. For the variation, we used three different phases (0°, 120°, and 240°) in two directions (clockwise and counter-clockwise). This strategy generated a set of 12 potential gesture settings for the experiment. For other parameters used in Whirling, we did not tune

[2] Holographic Remoting Player: https://apps.microsoft.com/detail/9nblggh4sv40

[3] Color, light, and materials: https://learn.microsoft.com/windows/mixed-reality/design/color-light-and-materials



(a) **Near**: 1 meter

(b) **Far**: 2 meters

Figure 3: The experimental setup for both the 'Near'(a) and 'Far'(b) settings are depicted in these images, which were captured from the device. These images are representative of the device's specified field of view (FOV) of 64.69°, as detailed in its technical documentation.

per participant and used values selected by the pilot study as follows: $RMS_{xyz} = 0.001, Timewindow_{min} = 1.0, Timewindow_{max} = 3.0, \Delta t = 1.5, THRE_{min} = 0.75, THRE_{max} = 0.85$. In the experiment, as the target rotates, we chose the target from the combination of speed and directions, which resulted in four possible targets.

In each trial, we displayed and tested 12 targets, each 48mm in size, positioned 1 and 2 meters away from the user. These distances corresponded to target display sizes of approximately 2.75° and 1.38°, respectively. Assuming that buttons are anchored on objects and users can interact with them directly or indirectly from a distance, we followed the manufacturers' recommended sizes. MRTK suggests a minimum size of 32mm for direct input and 1° for ray or gaze interaction[4]. Similarly, Apple visionOS recommends a minimum button size of 44mm[5]. By adhering to these guidelines, we ensure that our target sizes are representative of real-world use cases and are suitable for both direct and indirect interaction methods.

4.2 Procedure

We conducted an experiment with four sessions involving two distance variants and two methods. The sequence of these sessions was balanced using a Latin square. Before the experiment, we informed the participants about both methods and gave them the opportunity to freely experiment with targets placed at 1.5 meters, the middle distance of the two experiment settings. After the free trials, the participants performed rehearsal trials at the same 1.5 meters setting, with three to five attempts for each method.

In the pilot testing, we observed that participants often unintentionally raised their hands before they found the targets, which could potentially impact the data. To address this, we instructed participants to rest their hands on the chair armrest before the trial and concealed targets. After a countdown, the targets were revealed, and then the trial began. While the targets were not initially displayed, they were all conveniently visible without the need for extensive searching.

Each display position was tested with two different speeds in both directions of Whirling. For ray casting, we ran four trials per position to balance the number of trials with Whirling. This requirement resulted in a total of 48 trials per session. To prevent participant fatigue and frustration, any trials that could not be completed within 15 seconds were counted as failures.

Following each session, participants completed a NASA Task Load Index in Raw TLX form [27]. At the end of the four sessions, a post-experiment interview was conducted, collecting data from the social acceptability survey [52] and asking for general preference between the two methods.

4.3 Result

We collected data on the success rate and completion time for each trial, as well as user feedback on their preferences and perceived workload. This data was analyzed using a two-way repeated measures ANOVA to assess success rate and completion time. In our post-hoc analysis, we used a permutation test with Benjamini-Hochberg correction and an alpha value of 0.05 to determine significance. We utilized a one-tailed permutation test [62]. First, we grouped the results of each participant in the four settings, resulting in sixteen results per setting. Then, we randomly shuffled these sets within each setting and calculated the difference in means of the shuffled pairs. We repeated this process for 10,000 times, counting the differences that surpassed the original paired set to determine if the difference was significant.

4.3.1 Success Rate

Unfortunately, we could not find an effect on the success rate through repeated measure ANOVA. However, the success rate of 'Ray casting Far' showed a relatively low value ($M = 86.58, SD = 15.48$) on average. 'Whirling Near' showed the highest rate ($M = 95.63, SD = 6.46$) followed by 'Whirling Far' ($M = 95.53, SD = 5.74$) and 'Ray casting Near' ($M = 93.59, SD = 7.64$). Fig. 4 illustrates the success rates for each method at different target distances, with error bars representing 95% confidence intervals and asterisks indicating statistically significant differences between the two methods. We also calculated the success rate across different display positions. Fig. 5 illustrates the average success rate for each position, which geometrically corresponds to the displayed positions from the user's perspective, as shown in Fig. 3.

We investigated the types of errors observed for both ray casting and Whirling. In the case of ray casting, there were no false triggered inputs, but all errors were made by timeout. We could find both false target activation and timeout errors for the Whirling. The timeout error rate for 'Whirling Near' ($M = 2.55\%, SD = 4.35$) and 'Whirling Far'($M = 2.80\%, SD = 4.95$) showed higher value than false activation error (Near: $M = 1.81\%, SD = 3.22$ / Far: $M = 1.68\%, SD = 2.06$).

We also counted the number of pending attempts per successful trial for Whirling, as participants could correct their target before confirmation. On average, the pending counts were 1.08 ($SD = 0.08$) for Near targets and 1.12 ($SD = 0.09$) for Far targets. The maximum number of attempts in a trial was 3. In the Near distance condition, 92.4% of the trials were completed on the first attempt, 7.2% on the second attempt, and 0.4% required a third attempt. For the Far distance condition, 88.8% of the trials were completed on the first attempt, 10.5% on the second attempt, and 0.5% on the third attempt.

4.3.2 Completion Time

For the completion time analysis, we included only successful trials to prevent timeout failures from skewing the data. Successful trials encompassed those with multiple attempts across both methods. We measured the total time, including canceled selections, until the participant successfully completed the task. This approach allowed for a fair comparison between the two techniques, accounting for their unique characteristics and Whirling's ability to correct target selection before confirmation.

From the ANOVA, the method did not show the effect, but significant effects were found on distance ($F(1,15) = 32.61, p < .001, \eta^2 = 0.244$) and interaction ($F(1,15) = 25,49, p < .001, \eta^2 = 0.219$). During hypothesis testing, we identified significant findings. 'Ray casting near' was notably faster than 'Ray casting far' ($p_{adj} < .001$), 'Whirling Near' ($p_{adj} = .001$), and 'Whirling Far' ($p_{adj} < .001$). Furthermore, 'Whirling Near' performed quicker than 'Ray casting far' ($p_{adj} = 0.008$). 'Whirling Far' also showed a faster completion time than 'Ray casting far'($p_{adj} = 0.011$). However, the difference between 'Whirling Near' and 'Whirling

[4] Button - Mixed Reality https://learn.microsoft.com/windows/mixed-reality/design/button

[5] Buttons — Apple Developer Documentation https://developer.apple.com/design/human-interface-guidelines/buttons

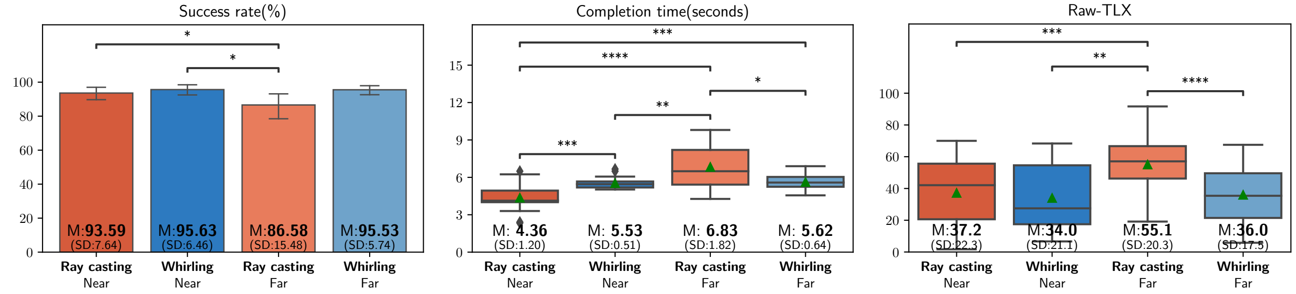
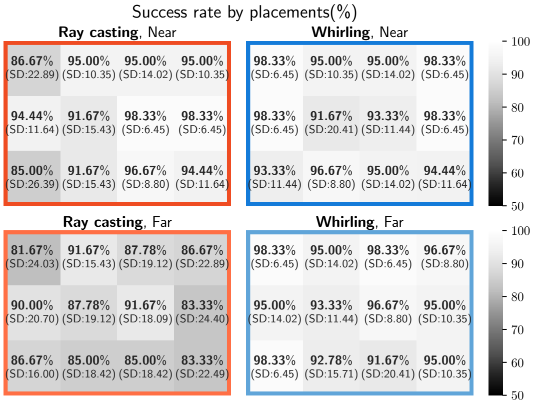


Figure 4: Success rate, completion time, and Raw-TLX values for each method under different conditions.



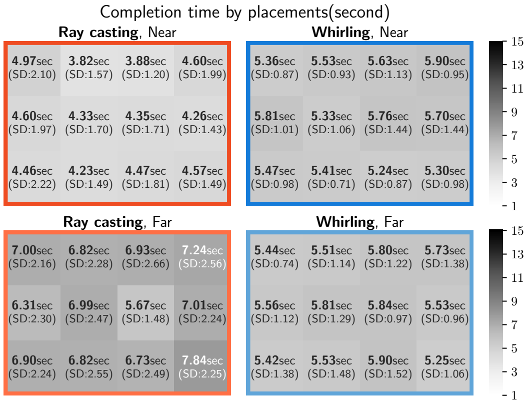


Figure 5: The success rates for each method and target size are distributed according to the presented display position. Each cell contains the results of the trial conducted at that particular position.

Figure 6: The completion times for each method and target size are distributed according to the presented display position. Each cell contains the results of the trial conducted at that particular position.

Far' was rejected. The results revealed that the 'Ray casting Near' ($M = 4.36, SD = 1.20$) showed the fastest completion time, followed by 'Whirling Near' ($M = 5.53, SD = 0.51$) and 'Whirling Far' ($M = 5.62, SD = 0.64$). 'Ray casting far' ($M = 6.83, SD = 1.82$) had the slowest completion time. Completion times for each distance and method are presented in Fig. 4 as boxplots with 1.5 interquartile range and asterisks indicating statistically significant differences. Following the success rate, we calculate the completion time across different display positions. Fig. 6 shows the average completion time on each position across all trials.

4.3.3 Hand position and sensing FOV

We collected hand positions during trials to investigate user behavior and required sensing capability. Following Whirling's reference point, we gathered the wrist joint of the right hand using the left-handed Cartesian coordinate system, which follows the conventions of Unity. For the analysis, we used data from successful trials, the same as the completion time analysis, and we used positions of the last three seconds to avoid noise from starting movements. Average positions were calculated as the arithmetic mean of wrist coordinates per trial, representing the central tendency of hand position. Sensing field of view (FOV) was determined by measuring the angle between the device and wrist, then finding the range from minimum to maximum for each trial, revealing the extent of hand motion.

For the horizontal sensing FOV, we could reveal a significant main effect on the method ($F(1,15) = 76.09$, $p < .001$, $\eta^2 = .208$) and interaction ($F(1,15) = 28.00$, $p < .001$, $\eta_p^2 = .040$), but not in distance. In post hoc analysis, all combinations show significant differences except for the distance difference within the same method. Similar to horizontal, on the vertical sensing FOV, we could reveal a significant main effect on the method ($F(1,15) = 111.06$, $p < .001$, $\eta^2 = .461$) and interaction ($F(1,15) = 54.59$, $p < .001$, $\eta^2 = .132$), but not on distance. In post hoc for vertical, we could also find all combinations have significant differences but between Whirling with different distances. These results indicate that the method significantly influenced both sensing FOV, and this effect varied depending on the distance condition, despite distance alone not having a significant main effect. Regarding the sensing FOV, we found that participants utilized a wider vertical range than the horizontal range for all distance and method combinations. Fig. 9 shows the average and standard deviation values with significance.

4.3.4 Subjective responses

To investigate the perceived workload, we utilized Raw-TLX. Repeated measures ANOVA and paired permutation tests with correction were performed, which was consistent with other results. The ANOVA results showed significant main effects of method ($F(1,15) = 7.01$, $p = .018$, $\eta^2 = .074$) and distance ($F(1,15) = 9.55$, $p = .007$, $\eta^2 = .059$), as well as interaction ($F(1,15) = 9.68$, $p = .007$, $\eta_g^2 = .039$). In the post hoc permutation test, results displayed that 'Ray casting Near' ($M = 37.2, SD = 22.3$), 'Whirling Near'($M = 34.0, SD = 21.1$), and 'Whirling Far' ($M = 36.0, SD = 17.5$) had comparable scores, while 'Ray casting far' ($M = 55.1, SD = 20.3$) was significantly higher than the rest. In statistical terms, 'Ray casting far' had a notably higher Raw-TLX score than 'Whirling Near' ($p_{adj} = .001$), 'Whirling Far'($p_{adj} < .001$), and 'Ray casting near' ($p_{adj} < .001$).

However, no significant differences were detected between the

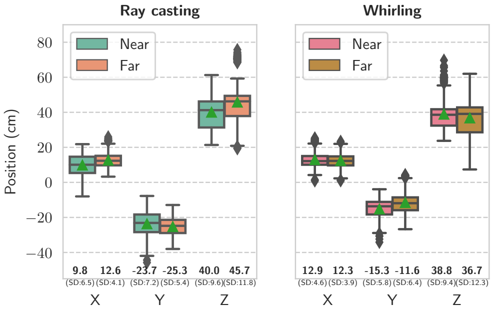


Figure 7: Average hand positions relative to the head during the ray casting and Whirling trials.

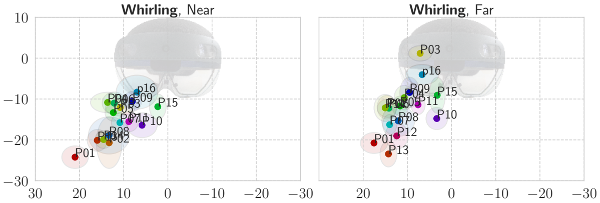


Figure 8: Distribution of hand positions during Whirling per participant. Average X-Y coordinates are indicated by points, and standard deviations are shown in ellipses.

methods for social acceptability counts. For locations, Whirling scored slightly higher (M = 3.00, SD = 1.67) than ray casting (M = 2.88, SD = 1.02). In terms of audience, ray casting received a slightly higher score (M = 4.75, SD = 1.34) than Whirling (M = 4.44, SD = 1.67). Despite these minor differences in social acceptability scores, a majority of the participants, 10 out of 16, expressed a preference for the Whirling technique over ray casting in a general context, with the remaining 6 participants favoring ray casting.

5 Discussion

The results indicate that Whirling Interface is more beneficial than ray casting when interacting with distant targets. Although Whirling exhibited slower completion times compared to ray casting for closer targets, it surpassed for the farther distances. Furthermore, Whirling demonstrated consistent performance across distance and display position, both in terms of overall completion time and success rate, indicating its potential for a more consistent user experience. Based on the comparison study results, our discussion is divided into three main themes: (1) the role of input states as an interface, (2) the advantages of sensing, and (3) addressing limitations and future directions. Our objective is to provide a thorough understanding of the implications and identify opportunities to advance motion matching as an interface for XR displays.

Importance of input states. Expanding on motion matching selection, we enhanced our interface by modifying the input modality and introduced feedback-driven input states to improve interactivity. We also added a cancellation feature for mis-targeted selections, crucial for gestural interfaces. These enhancements were underpinned by our modified detection technique.

Our study revealed that approximately 18% of trials required multiple attempts to succeed, underscoring the importance of the cancellation function in motion-matching interfaces. It could be thought that using Whirling often requires multiple attempts, which the user could feel disturbing. However, subjective responses indicated that participants easily adapted to our input system such that the repeated attempts did not affect the experience much.

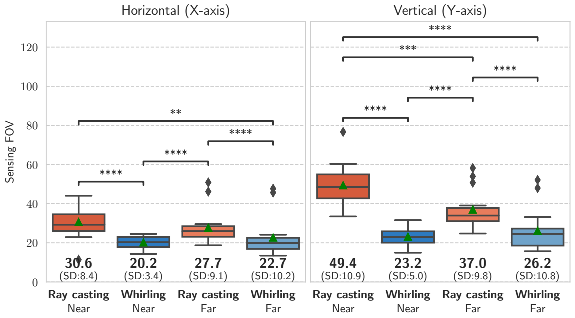


Figure 9: Sensing FOV of the ray casting and Whirling method.

While our current parameters were based on limited pilot testing, there is potential for further optimization through refinement and the application of machine learning techniques for state transitions, potentially replacing our current multiple thresholding approach. Notably, Whirling's completion time includes a 1.5 second pending period for potential cancellations. This duration could be dynamically adjusted based on user performance, target number, and input confidence level, potentially expediting the process.

Despite areas for potential improvement, our successful results indicate that the Whirling Interface can serve as a solid foundation for implementing motion-matching-based selection methods in gestural interfaces.

Advantages of sensing. The ray casting technique presents several challenges, such as difficulty in understanding the cause of failed pinch gestures and performance dependence on tracking resolution or non-trackable areas due to limited sensor field of view. Additionally, minor errors can cause substantial targeting discrepancies when interacting with distant objects. These limitations could also be found from our study. As the wrist is often positioned at a greater depth relative to the sensor when aiming at more distant objects, inaccuracies may arise from limitations in tracking resolution. Furthermore, when pointing at greater distances, the wrist may be partially occluded by the arm, and the posture becomes more difficult for the user to maintain. These factors may partially explain the subpar performance of ray casting at greater distances.

Moreover, several participants expressed dissatisfaction with the unlimited length of the ray in the ray casting method. They found it particularly bothersome as the ray constantly adjusted its size when facing a target, especially at the boundaries, causing visual disturbance. This sentiment seems to contradict findings from previous research [9]. However, this perceived drawback can be interpreted as an advantage in the context of the Whirling method, which does not rely heavily on spatial cognition, a crucial factor in ray casting.

Interestingly, our experiment revealed that Whirling exhibited consistent performance across two target distances and outperformed ray casting in the Far setting. As shown in Fig. 5 and Fig. 6, Whirling maintained steady success rates and completion times across different positions in the user's visual field. This result indicates that users can mentally coordinate synchronous movements and execute gestures consistently regardless of the target's size and position. Moreover, Whirling's performance is likely unaffected by target size, provided the user can perceive the target movement.

Whirling offers several additional advantages. It operates within a smaller sensing field of view and maintains consistent performance across target positions. Being a wrist-based method, it avoids self-occlusion issues. Minor inconsistencies, such as jittering or transient tracking loss, have minimal influence on its performance. These benefits could become more pronounced when using fewer sensors, less accurate sensors, or in outdoor conditions with lower tracking accuracy due to solar interference. In addition, vision-based hand

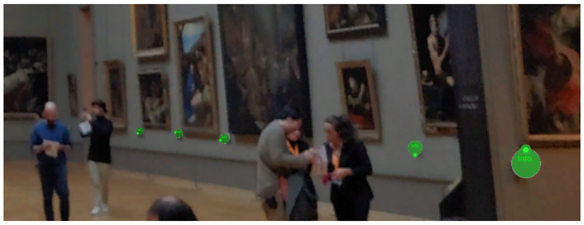


Figure 10: Whirling Interface concept in a museum. Small whirling buttons select information on distant locations, minimizing visual distraction compared to larger interfaces.

tracking could potentially be replaced with IMU sensors or radar, as our experiment showed that users could effectively mimic the orbit.

Subjective responses. When participants first saw the Whirling method, they expressed concern that it may be difficult. However, at the end of the experiment, they believed that the method was superior to ray casting. This qualitative observation is supported by the quantitative NASA-TLX scores, which show that Whirling for both near and far targets had less workload than the far condition for ray casting. A similar workload was reported for ray casting for near targets as for all Whirling targets. One of the participants reported that because Whirling has feedback, it informs their performance, which helps learning. In contrast, the pinch gesture required for ray casting was difficult to learn by just success or failure.

As evidenced in Fig. 7, we noticed a uniformity in the execution of the Whirling gesture by the participants, whether the target was near or far. Participants were able to align the movement of their hands with the target relatively well, demonstrating that the size of the target does not influence the ability to match movements. Indeed, even with variations in target size, users were able to maintain consistent gesture dimensions.

Participants were not specifically guided to make their gestures smaller; they were allowed to perform according to their preferences. Interestingly, five participants managed to execute the Whirling gesture within a diameter of approximately 10 cm. This finding indicates the potential for Whirling to work with smaller gestures, which could enhance its social acceptability.

Applications. Practically, Whirling could serve as an alternative selection method in certain scenarios or integrate with button-based interfaces. For instance, imagine walking through a museum equipped with smart glasses; rather than needing large buttons or moving close to an object to make a selection, users could simply select an augmented name tag of an exhibited object by performing a Whirling gesture. Alternatively, Whirling could be combined with traditional selection techniques to enable interaction with significantly smaller buttons, providing quick shortcut menus. Since Whirling is tolerant to small perturbations, it may be possible to select a target with Whirling while walking toward it.

Limitations And Future Work. Both ray casting and Whirling have the potential for improvement, and the optimal choice depends on the specific context. Further comparisons are necessary to understand their strengths and weaknesses fully. 'Ray casting' performance could be enhanced through attention modeling or modern pinch gesture detection techniques, which could also be applied to Whirling. Whirling could potential improve reaction time or accuracy through machine learning such as the Siamese network to calculate correlation.

We did not conduct a comparative study with an input method that combines eye tracking for target setting and pinch for confirmation, a well-known technique adapted by Apple Vision Pro. This decision was made for several reasons. First, many VR headsets or smart glasses do not feature an eye-tracking module, and research papers reported that they are not fit for small target selection. Still, we could have combined head movement with the pinch gesture. However, we chose not to do so, as it does not fully adhere to Norman's fundamental principles. Our focus is on scenarios involving the everyday use of smart glasses, where multiple interactable objects are likely present. This scenario could highlight numerous potential interactable targets, potentially bothering the user with animations or false activations. Consequently, we concentrated on hand input exclusively, which explicitly signals the intention to make an input by raising the hand. Nonetheless, the Whirling Interface could be extended by substituting hand tracking with a smartwatch's IMU. In such a scenario, some external activation, like the hotword of voice input, might be needed. In this case, a comparison with head (or eye) plus pinch input could be a feasible future direction. However, there are some possible ways to combine an accurate eye-tracking system and an intention recognition system to address our concerns. It could be a possible way to compare them. There is research adapting intention prediction with ray casting to improve user experience. For this reason, combining an intention-based threshold on Whirling and comparing it with ray casting will be an important future work.

Another important consideration is the scalability issue. Unlike traditional pointing techniques where throughput ($TP = ID_e/MT$) can be calculated using Fitts' law ($ID_e = \log_2(D/W_e + 1)$), the Whirling Interface is not significantly affected by distance (D) or target width (W_e). This unique characteristic makes it challenging to estimate the system's throughput using conventional methods. Nevertheless, there are other potential difficulty factors beyond target size and distance. For instance, an increased number of targets can reduce the distinction between the designated target and others, thereby increasing input difficulty. Additionally, while we only tested two speeds in our study, other speeds might prove more challenging to follow. Consequently, modeling an index of difficulty (ID) for the Whirling Interface remains a crucial task to ensure its widespread adoption, similar to traditional pointing methods.

The current study did not test Whirling's performance in scenarios involving user mobility or occupied hands, where ray casting would likely be infeasible. Future work should explore these contexts, as Whirling has the potential to function effectively in such situations.

6 Conclusion

We introduced the Whirling Interface, a novel bare-hand input method for XR displays using motion matching. Our comparative study with ray casting revealed that the Whirling Interface offers superior accuracy, consistency, and completion time for small or distant objects. It improves upon previous motion-matching methods by incorporating distinct functional input states and features flexible sensing requirements. Moreover, an advantage of Whirling is the number and type of sensors that may successfully sense the required movement. As we continue to explore and refine this interface, we anticipate that it will play a substantial role in shaping the future of XR interfaces for everyday use.

Acknowledgments

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2024-00397663, Real-time XR Interface Technology Development for Environmental Adaptation) and Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0012746, The Competency Development Program for Industry Specialist).

References

[1] S. Ahn and G. Lee. Gaze-assisted typing for smart glasses. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 2019. doi: 10.1145/3332165.3347883 2

[2] M. Al-Kalbani, I. Williams, and M. Frutos-Pascual. Analysis of medium wrap freehand virtual object grasping in exocentric mixed

reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2016. doi: 10.1109/ISMAR.2016.14 2

[3] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013. doi: 10.1016/j.cag.2012.12.003 1

[4] S. V. Babu, H.-C. Huang, R. J. Teather, and J.-H. Chuang. Comparing the fidelity of contemporary pointing with controller interactions on performance of personal space target selection. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 404–413, 2022. doi: 10.1109/ISMAR55827.2022.00056 1

[5] T. Bafna, P. Bækgaard, and J. P. P. Hansen. EyeTell: Tablet-based Calibration-free Eye-typing using Smooth-pursuit movements. *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–6, 2021. doi: 10.1145/3448018.3458015 2

[6] M. Baloup, V. Oudjail, T. Pietrzak, and G. Casiez. Pointing techniques for distant targets in virtual reality. In *Proceedings of the 30th Conference on l'Interaction Homme-Machine*, p. 100–107. ACM, 2018. doi: 10.1145/3286689.3286696 1

[7] M. Baloup, T. Pietrzak, and G. Casiez. Raycursor: A 3d pointing facilitation technique based on raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 1–12. ACM, 2019. doi: 10.1145/3290605.3300331 1, 2

[8] A. U. Batmaz and W. Stuerzlinger. Effects of 3D Rotational Jitter and Selection Methods on 3D Pointing Tasks. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1687–1692, 2019. doi: 10.1109/vr.2019.8798038 1

[9] A. U. Batmaz and W. Stuerzlinger. Effect of fixed and infinite ray length on distal 3d pointing in virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, p. 1–10. ACM, 2020. doi: 10.1145/3334480.3382796 7

[10] A. Bellino. SEQUENCE: a remote control technique to select objects by matching their rhythm. *Personal and Ubiquitous Computing*, 22(4):751–770, 2018. doi: 10.1007/s00779-018-1129-2 3

[11] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, p. 35–ff. ACM, 1997. doi: 10.1145/253284.253301 2

[12] Y. Cha and R. Myung. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics*, 43(4):350–355, 2013. doi: 10.1016/j.ergon.2013.05.005 1

[13] C. Clarke, A. Bellino, A. Esteves, and H. Gellersen. Remote control by body movement in synchrony with orbiting widgets: An evaluation of tracematch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), sep 2017. doi: 10.1145/3130910 3

[14] C. Clarke and H. Gellersen. MatchPoint: Spontaneous Spatial Coupling of Body Movement for Touchless Pointing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 179–192. ACM, 2017. doi: 10.1145/3126594.3126626 2

[15] M. Dhuliawala, J. Lee, J. Shimizu, A. Bulling, K. Kunze, T. Starner, and W. Woo. Smooth eye movement interaction using EOG glasses. *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 307–311, 2016. doi: 10.1145/2993148.2993181 2

[16] H. Drewes, M. Khamis, and F. Alt. Smooth Pursuit Target Speeds and Trajectories. *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pp. 139–146, 2018. doi: 10.1145/3282894.3282913 2

[17] T. J. Dube, Y. Ren, H. Limerick, I. S. MacKenzie, and A. S. Arif. Push, Tap, Dwell, and Pinch: Evaluation of Four Mid-air Selection Methods Augmented with Ultrasonic Haptic Feedback. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):207–225, 2022. doi: 10.1145/3567718 1, 3

[18] E. Dykstra-Erickson, M. Tscheligi, J. Williamson, and R. Murray-Smith. Pointing without a pointer. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, pp. 1407–1410, 2004. doi: 10.1145/985921.986076 2

[19] A. Esteves, E. Bouquet, K. Pfeuffer, and F. Alt. One-handed Input for Mobile Devices via Motion Matching and Orbits Controls. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–24, 2022. doi: 10.1145/3534624 2, 3

[20] A. Esteves, Y. Shin, and I. Oakley. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human-Computer Studies*, 139:102414, 2020. doi: 10.1016/j.ijhcs.2020.102414 2

[21] A. Esteves, E. Velloso, A. Bulling, and H. Gellersen. Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, pp. 457–466. ACM, 2015. doi: 10.1145/2807442.2807499 2, 3

[22] A. Esteves, D. Verweij, L. Suraiya, R. Islam, Y. Lee, and I. Oakley. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 167–178. ACM, 2017. doi: 10.1145/3126594.3126616 2

[23] M. Fukumoto, Y. Suenaga, and K. Mase. "finger-pointer": Pointing interface by image processing. *Computers & Graphics*, 18(5):633–642, 1994. doi: 10.1016/0097-8493(94)90157-0 2

[24] T. Grossman, X. A. Chen, and G. Fitzmaurice. Typing on glasses: Adapting text entry to smart eyewear. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, p. 144–152. ACM, 2015. doi: 10.1145/2785830.2785867 2

[25] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T.-H. Yu, C.-J. Tai, M. Akbay, Z. Wang, A. Nitzan, G. Dong, Y. Ye, L. Tao, C. Wan, and R. Wang. Megatrack: Monochrome egocentric articulated hand-tracking for virtual reality. *ACM Trans. Graph.*, 39(4), aug 2020. doi: 10.1145/3386569.3392452 2

[26] S. Han, P.-C. Wu, Y. Zhang, B. Liu, L. Zhang, Z. Wang, W. Si, P. Zhang, Y. Cai, T. Hodan, R. Cabezas, L. Tran, M. Akbay, T.-H. Yu, C. Keskin, and R. Wang. Umetrack: Unified multi-view end-to-end hand tracking for vr. In *SIGGRAPH Asia 2022 Conference Papers*. ACM, 2022. doi: 10.1145/3550469.3555378 2

[27] S. G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909 5

[28] Y. Igarashi, K. Futami, and K. Murao. Silent speech eyewear interface: Silent speech recognition method using eyewear with infrared distance sensors. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers*, p. 33–38. ACM, 2022. doi: 10.1145/3544794.3558458 1

[29] J. Kaye, A. Druin, C. Lampe, D. Morris, J. P. Hourcade, M. Carter, E. Velloso, J. Downs, A. Sellen, K. O'Hara, and F. Vetere. PathSync. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3415–3427, 2016. doi: 10.1145/2858036.2858284 2, 3

[30] M. Khamis, C. Oechsner, F. Alt, and A. Bulling. VRpursuits: interaction in virtual reality using smooth pursuit eye movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, pp. 1–8. ACM, 2018. doi: 10.1145/3206505.3206522 2

[31] W. Kim and S. Xiong. ViewfinderVR: configurable viewfinder for selection of distant objects in VR. *Virtual Reality*, 26(4):1573–1592, 2022. doi: 10.1007/s10055-022-00649-z 2

[32] N. Kimura, M. Kono, and J. Rekimoto. Sottovoce: An ultrasound imaging-based silent speech interaction using deep neural networks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019. doi: 10.1145/3290605.3300376 1

[33] M. Krüger, T. Gerrits, T. Römer, T. Kuhlen, and T. Weissker. Intenselect+: Enhancing score-based selection in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2829–2838, 2024. doi: 10.1109/TVCG.2024.3372077 2, 3

[34] S. Kuznetsov, D. Saakes, R. Wakkary, L. Geurts, L. Hayes, M. Lau, D. Verweij, A. Esteves, S. Bakker, and V.-J. Khan. Designing Motion Matching for Real-World Applications. *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 645–656, 2019. doi: 10.1145/3294109.3295628 2

[35] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head- and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 1–14. ACM, 2018. doi: 10.1145/3173574.3173655 2

[36] J. Lee, S. Aggarwal, J. Wu, T. Starner, and W. Woo. SelfSync: exploring

self-synchronous body-based hotword gestures for initiating interaction. In *Proceedings of the 23rd International Symposium on Wearable Computers*, pp. 123–128. ACM, 2019. doi: 10.1145/3341163.3347745 2

[37] P. Lukowicz, A. Krüger, A. Bulling, Y.-K. Lim, S. N. Patel, C. Clarke, A. Bellino, A. Esteves, E. Velloso, and H. Gellersen. TraceMatch. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 298–303, 2016. doi: 10.1145/2971648.2971714 2, 3

[38] M. Ma, K. Merckx, P. Fallavollita, and N. Navab. [poster] natural user interface for ambient objects. In *2015 IEEE International Symposium on Mixed and Augmented Reality*, pp. 76–79, 2015. doi: 10.1109/ISMAR.2015.25 2

[39] R. Mandryk, M. Hancock, M. Perry, A. Cox, M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018. doi: 10.1145/3173574.3173655 1, 3

[40] G. Mark, S. Fussell, C. Lampe, J. P. Hourcade, C. Appert, D. Wigdor, D. Verweij, A. Esteves, V.-J. Khan, and S. Bakker. WaveTrace. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2180–2186, 2017. doi: 10.1145/3027063.3053161 2, 3

[41] A. Marquardt, M. Steininger, C. Trepkowski, M. Weier, and E. Kruijff. Selection performance and reliability of eye and head gaze tracking under varying light conditions. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 546–556. IEEE Computer Society, mar 2024. doi: 10.1109/VR58804.2024.00075 3

[42] F. Mattern, S. Santini, J. F. Canny, M. Langheinrich, J. Rekimoto, M. Vidal, A. Bulling, and H. Gellersen. Pursuits. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 439–448, 2013. doi: 10.1145/2493432.2493477 2

[43] M. R. Mine, F. P. Brooks, and C. H. Sequin. *Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction*, p. 19–26. ACM Press/Addison-Wesley Publishing Co., 1997. 2

[44] D. A. Norman. Natural user interfaces are not natural. *Interactions*, 17(3):6–10, may 2010. doi: 10.1145/1744161.1744163 2, 3

[45] D. R. Olsen, R. B. Arthur, K. Hinckley, M. R. Morris, S. Hudson, S. Greenberg, J.-D. Fekete, N. Elmqvist, and Y. Guiard. Motion-pointing: target selection using elliptical motions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 289–298, 2009. doi: 10.1145/1518701.1518748 2

[46] F. Paternò, G. Jacucci, M. Rohs, C. Santoro, H. Drewes, M. Khamis, and F. Alt. DialPlates. *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pp. 1–10, 2019. doi: 10.1145/3365610.3365626 2

[47] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst. Exploring Natural Eye-Gaze-Based Interaction for Immersive Virtual Reality. *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 36–39, 2017. doi: 10.1109/3dui.2017.7893315 1

[48] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. *The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR*, p. 79–80. ACM, 1996. 2

[49] I. POUPYREV and T. ICHIKAWA. Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing*, 10(1):19–35, 1999. doi: 10.1006/jvlc.1998.0112 2

[50] Y. Y. Qian and R. J. Teather. 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*, p. 91–98. ACM, New York, NY, USA, 2017. doi: 10.1145/3131277.3132182 1, 2

[51] G. Reyes, J. Wu, N. Juneja, M. Goldshtein, W. K. Edwards, G. D. Abowd, and T. Starner. SynchroWatch: One-Handed Synchronous Smartwatch Gestures Using Correlation and Magnetic Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–26, 2018. doi: 10.1145/3161162 2

[52] J. Rico and S. Brewster. Usable gestures for mobile interfaces: Evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 887–896. ACM, 2010. doi: 10.1145/1753326.1753458 5

[53] J. Schjerlund, K. Hornbæk, and J. Bergström. Ninja hands: Using many hands to improve target selection in vr. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 2021. doi: 10.1145/3411764.3445759 2

[54] R. Shi, J. Zhang, Y. Yue, L. Yu, and H.-N. Liang. Exploration of Bare-Hand Mid-Air Pointing Selection Techniques for Dense Virtual Reality Environments. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2023. doi: 10.1145/3544549.3585615 2, 3

[55] J. Shimizu, J. Lee, M. Dhuliawala, A. Bulling, T. Starner, W. Woo, and K. Kunze. Solar system: Smooth pursuit interactions using eog glasses. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, p. 369–372. ACM, 2016. doi: 10.1145/2968219.2971376 2

[56] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen. Outline Pursuits: Gaze-assisted Selection of Occluded Objects in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, 2020. doi: 10.1145/3313831.3376438 2

[57] T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, R. W. Picard, and A. Pentland. Augmented Reality through Wearable Computing. *Presence: Teleoperators and Virtual Environments*, 6(4):386–398, 1997. doi: 10.1162/pres.1997.6.4.386 2

[58] Y. Tian, H. Bai, S. Zhao, C.-W. Fu, C. Yu, H. Qin, Q. Wang, and P.-A. Heng. Kine-appendage: Enhancing freehand vr interaction through transformations of virtual appendages. *IEEE Transactions on Visualization and Computer Graphics*, 2022. doi: 10.1109/TVCG.2022.3230746 2

[59] E. Velloso, M. Carter, J. Newn, A. Esteves, C. Clarke, and H. Gellersen. Motion Correlation: Selecting Objects by Matching Their Movement. *ACM Transactions on Computer-Human Interaction*, 24(3):1–35, 2017. doi: 10.1145/3064937 2

[60] D. Vogel and R. Balakrishnan. 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*, p. 33–42. ACM, 2005. doi: 10.1145/1095034.1095041 2

[61] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbæk, K. Pfeuffer, and H. Gellersen. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 2023. doi: 10.1145/3544548.3581423 1, 2, 3

[62] W. J. Welch. Construction of permutation tests. *Journal of the American Statistical Association*, 85(411):693–698, 1990. doi: 10.1080/01621459.1990.10474929 5

[63] D. Wolf, J. Gugenheimer, M. Combosch, and E. Rukzio. Understanding the heisenberg effect of spatial interaction: A selection induced error for spatially tracked input devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, p. 1–10. ACM, 2020. doi: 10.1145/3313831.3376876 2

[64] J. Wu, C. Colglazier, A. Ravishankar, Y. Duan, Y. Wang, T. Ploetz, and T. Starner. Seesaw: rapid one-handed synchronous gesture interface for smartwatches. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pp. 17–20. ACM, 2018. doi: 10.1145/3267242.3267251 2

[65] Y. Yan, C. Yu, X. Yi, and Y. Shi. Headgesture: Hands-free input approach leveraging head movements for hmd devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4), dec 2018. doi: 10.1145/3287076 2

[66] S. Yi, Z. Qin, E. Novak, Y. Yin, and Q. Li. Glassgesture: Exploring head gesture interface of smart glasses. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, 2016. doi: 10.1109/INFOCOM.2016.7524542 2

[67] C. Yu, K. Sun, M. Zhong, X. Li, P. Zhao, and Y. Shi. One-dimensional handwriting: Inputting letters and words on smart glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, p. 71–82. ACM, 2016. doi: 10.1145/2858036.2858542 2

[68] D. Yu, Q. Zhou, J. Newn, T. Dingler, E. Velloso, and J. Goncalves. Fully-occluded target selection in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3402–3413, 2020. doi: 10.1109/TVCG.2020.3023606 2