Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Precise Object Selection in Dense Virtual Environments Using Single Hand Gesture with 2D Interfaces

Cheol Ho Mun
Kyung Hee University
Tack Woo
twoo@khu.ac.kr

Kyung Hee University

Research Article

Keywords: Virtual Reality, Selection Technique, Human-Computer Interaction, Hand Interaction, Gesture Interaction

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10586409/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

While Raycasting remains the standard for freehand target selection in virtual reality (VR), it fundamentally relies on a single ray casting from the hand, which merely detects the first intersected object. Consequently, Raycasting exhibits significant performance degradation in dense environments and fails to effectively select occluded objects. To overcome these inherent limitations, we introduce Single-Hand Gesture Widget Selection (SHGWS). SHGWS leverages a combination of volumetric area selection, hand gestures and a 2D widget to facilitate target disambiguation. We validated the efficacy of SHGWS through a comparative user study with 27 participants, evaluating it against standard Raycasting and a Double-Hand Widget Selection (DHWS) baseline. Our findings indicate that SHGWS yields superior selection performance compared to both alternatives. This work contributes the design and empirical evaluation of a novel interaction technique that empowers robust, single-handed object selection in bare-hand VR environments.

1 Introduction

Target selection in Virtual Reality (VR) environments represents one of the most fundamental and critical tasks in 3D interaction (LaViola Jr 2017). Today, target acquisition in VR typically operates based on Raycasting, establishing itself as the standard interaction technique in the field (Baloup et al. 2019; Mine 1995). Raycasting functions by detecting the initial intersection point of a ray projected from either a controller or the user's hand. However, its inherent limitations become pronounced when attempting precise selection within dense environments. Because it relies on a single-ray mechanism, acquiring targets that are distant, small, or densely clustered poses significant challenges (Turkmen 2026). Furthermore, a critical constraint exists in that fully occluded objects cannot be selected at all. Therefore, in a dense environment, a strategy of pre-selecting multiple objects at once through area selection and then making a final selection again was adopted rather than Raycasting, a method of selecting one object (Sidenmark et al. 2024; Weller et al. 2021).

To address these challenges, a variety of selection techniques have been investigated, which can be broadly categorized into controller-based and bare-hand interactions. However, previous studies have predominantly focused on controller-based methods, leaving bare-hand selection techniques relatively under-explored (Shi et al. 2024; Meng et al. 2022). Moreover, there is a scarcity of research dedicated to selection techniques that can be fully and effectively executed using only a single hand (Xu et al. 2022; Kim et al. 2025).

Bare-hand interaction offers distinct advantages over controller-based methods, such as enhancing the user’s sense of presence and immersion (Buckingham 2021; Adkins et al. 2021). However, prolonged mid-air hand interactions accumulate fatigue in the arm muscles, a phenomenon known as the “Gorilla Arm Effect” (Vlahovic et al. 2024). Requiring both hands to perform a fundamental task can impose continuous muscular strain on both arms. Furthermore, because both hands are occupied during each selection event, the user’s capacity for concurrent interactions is restricted (Zavichi et al. 2025; Hu et al.

2026)). In contrast, a single-handed technique allows users to alternate hands, providing opportunities to rest each arm, while simultaneously enabling them to perform other interactions during the selection process.

Accordingly, this study proposes Single-Hand Gesture Widget Selection (SHGWS), a precise selection technique for dense environments that can be performed using only one hand. SHGWS achieves disambiguation by dividing the selection process into a pre-selection phase and a post-selection phase. Both of which are driven by single-hand gestures and a 2D widget. Through this approach, we present a method for accurately selecting a single object among dense environments using bare hands, specifically offering a technique that requires only one hand to operate. The primary goal of this research is to contribute to securing freedom of interaction while maintaining higher presence and immersion inherent to bare-hand interaction.

2 Related Works

This research proposes a target selection technique that enables precise manipulation using bare hands and, specifically, single-handed interaction. To overcome the inherent difficulties of target selection in virtual reality, a wide range of selection techniques has been extensively investigated.

2.1 VR Object Selection Techniques

Previous Research on target selection techniques in VR is generally categorized into controller-based and hand-tracking-based approaches, with additional studies exploring the use of multimodal inputs (Yu et al. 2024).

In controller-based research, various attempts have been made to overcome the limitations of the traditional Raycasting method. For instance, Weller et al. (2021) proposed the LenSelect technique, which dynamically increases the size of potentially selectable target objects to reduce selection difficulty according to Fitts' Law. This method successfully achieved a 44% lower error rate compared to standard Raycasting. However, since it arbitrarily modifies object's visual within the virtual content, it has the disadvantage of potentially lowering the overall sense of immersion. Wu et al. (2023) proposed ClockRay, a technique that performs disambiguation by manipulating a cursor along the ray through wrist rotation. By controlling the depth of selection via wrist movements, this approach facilitated the selection of occluded objects. In another study, Yu et al. (2020) proposed a total of seven control-based techniques for selecting fully occluded objects and systematically evaluated how factors such as the degree of occlusion, target depth, and object density impact technique performance.

Multimodal techniques combining gaze and controllers have also been studied. Sidenmark et al. (2024) proposed Cone&Bubble, a two-stage technique that pre-selects candidate objects using gaze-based cone pointing and performs the final selection via the controller's bubble raycasting. Their experimental results confirmed that gaze is suitable for rapid candidate selection, while the controller is appropriate for precise final selection. Furthermore, Chen et al. (2023) introduced GazeRayCursor, which estimates

the depth of an object by utilizing the intersection of two rays emitted from the gaze and the controller. In dense environments, this method achieved a 45% reduction in selection time and a 10-fold decrease in manual depth adjustments compared to conventional techniques.

In addition, techniques utilizing various other modalities have been proposed. For example, Turkmen et al. (2026) introduced VoiceRay, a technique that combines controller-based Raycasting with voice recognition to perform disambiguation through voice commands in dense or occluded VR Environments.

In comparison, research on bare-hand interaction remains relatively scarce. Despite the inherent instability and precision limitations of vision-based hand tracking (Kim et al. 2025; Liu et al. 2024), bare-hand interaction continues to be steadily investigated due to its significant advantages. Specifically, it enables interaction through natural movements without the need for extra devices and enhances user presence (Polydorou et al. 2021; Moon et al. 2023) and provides an intuitive interaction experience (Chua et al. 2024; Song et al. 2023). Driven by these benefits, the following studies have been conducted to develop bare-hand selection techniques.

For bare-hand target selection in dense VR environments, Shi et al. (2023) implemented an existing technique, MultifingerBubble, alongside proposing and comparatively evaluating two novel techniques: HandDepthCursor and HandConeGrid. MultiFingerBubble is a single-handed technique that emits a ray from each of the four fingers and selects a target by bending the corresponding finger (Delamare et al. 2022). However, because it is inherently based on Raycasting, it is ineffective for precise selection in dense or occluded environments. HandDepthCursor overcomes this by utilizing hand depth information to enable precise selection, while HandConeGrid performs disambiguation by arranging objects detected within a cone volume into a grid format. The user study results demonstrated that both HandDepthCursor and HandConeGrid achieved faster and more accurate selection performance compared to MultiFingerBubble. In particular, HandConeGrid showed advantages in high-density environments. However, the grid-based arrangement presents limitations. Visually relocating the objects disrupts their original spatial context, and the grid itself can become visually cluttered as object density increases.

Sun et al. (2025) focused on the ‘Heisenberg Effect’, where the confirmation action (e.g., pinching) during bare-hand selection shakes the tracking pose and inadvertently alters the intended target. To address this, they proposed the Single-Hand Frozen Gesture (SHFG), a technique where the user makes a fist to freeze the target before performing the selection action. They developed two variant techniques based on SHFG, LiftFrozenFinger and PianoFrozenFinger. The results confirmed an improvement in selection accuracy and a reduction in hand fatigue compared to MultiFingerBubble.

However, bare-hand techniques still remain relatively under-investigated compared to controller-based techniques. Furthermore, even among the existing studies, single-handed techniques exhibit clear limitations in dense or occluded environments, including hand fatigue, hand tremor, and difficulty in selecting occluded objects.

While MultiFingerBubble has been utilized as a comparative condition in previous literature, raycasting remains the prevailing standard across the contemporary VR ecosystem. Given the persistent lack of a widely accepted alternative, a direct empirical comparison against Raycasting is required, independent of the outcomes of past studies.

Therefore, there is a need for a novel interaction design that enabled single-handed interaction while maintaining accessibility, while simultaneously facilitating intuitive and rapid disambiguation among dense or occluded targets. Addressing this need is the goal of this research.

2.2 Widget-based Selection Techniques

Widgets serve as visual interface elements, provided in either 2D or 3D formats, to assist users with complex selection tasks. The primary advantages of widget-based selection techniques can be summarized in two aspects. First, from the perspective of Fitts' Law, widgets effectively reduce the selection difficulty by displaying target objects at a larger scale or with greater spacing with the 2D interface space (Weller et al. 2021). In other words, objects that are difficult to select in 3D environments due to their small size or proximity can be magnified and rearranged in the widget space, thereby enhancing selection precision and resolving ambiguity (Han et al. 2022). Second, widgets provide a separate visualization space for candidate objects without requiring the manipulation or deformation of the actual 3D objects in the scene. This allows users to clearly identify selection targets while maintaining the spatial context of the VR environment. However, a common limitation of these techniques is that widgets may obstruct the user's field of view, potentially reducing immersion or creating visual distractions by occluding parts of the scene (Kim et al. 2022).

Debarba et al. (2013) proposed the Disambiguation Canvas to support precise object selection in VR environments. This technique operates in two stages: first, candidate objects are pre-selected via a volume casting approach driven by the orientation sensor of a mobile device, and then a 2D widget on the device's touchscreen is employed to confirm the final selection. This process enables rapid and accurate disambiguation, even among hundreds of objects. However, the technique possesses a structural limitation as it requires an external mobile device, preventing it from functioning with a VR/XR devices alone.

Maslych et al. (2023) introduced a technique that integrates Cone-Casting, the World-in-Miniature metaphor, and grasping to visualize occluded objects as a Disocclusion Mini-map widget placed around the user. This approach offers the advantage of visualizing spatial relationships between objects within a dedicated widget space, thereby avoiding direct modifications to the 3D scene. Comparative experiments demonstrated superior selection performance for this technique compared to other selection methods. However, a notable limitation remains: the mini-map widget occupies a portion of the user's field of view which can visually obstruct parts of the VR scene and subsequently reduce the user's immersion. Furthermore, this technique still uses raycast and controller for the post-selection, and needs both hands.

In summary, while existing widget-based selection techniques have demonstrated effectiveness in target disambiguation, they are often constrained by dependency on external hardware or visual obstruction of the user's field of view. The SHGWS technique proposed in this study differentiates itself from these prior approaches in several ways. First, it operates solely via vision-based hand tracking integrated into the VR headset. Second, it minimizes visual obstruction by placing the 2D widget within a limited space adjacent to the user's hand. Finally, it allows users to perform the entire process form pre-selection to post-selection using only a single hand.

3 Technique Design

The proposed technique, Single-Hand Gesture Widget Selection (SHGWS), employs a two-stage disambiguation pipeline. Initially, it utilizes volumetric selection to simultaneously pre-select a cluster of candidate objects. Subsequently, the user executes a final target confirmation leveraging a 2D widget and single-hand gestures. All gesture recognition is seamlessly processed through the VR headset's built-in hand-tracking capabilities.

3.1 Pre Selection

The pre-selection phase of our proposed technique operates based on a Cone Detection mechanism. This mechanism defines a conical detection volume centered around the ray emitted from the user's hand (Liang et al. 1994; Sidenmark et al. 2020). Objects detected within this cone are highlighted with a white outline. Upon performing a pinch gesture (bringing the thumb and index finger together), a 2D widget is activated, which visualizes the detected objects as a 2D list. Here, we operate under the assumption that the closer an object is to the ray, the higher the probability that it is the user's intended target (Weller et al. 2021; Luo et al. 2026). Therefore, objects positioned closer to the center of the detection cone are placed at the top of the list. The topmost item is then designated as the preliminary selection candidate and highlighted with a red outline in both the widget list and the corresponding 3D object. This layout is expected to minimize the number of scrolling actions required by the user. The activated widget remains visible as long as the pinch gesture is maintained, tracking the position of the user's hand. Releasing the pinch gesture deactivates the widget.

3.2 Post Selection

After the Pre-Selection phase via Cone Detection, users can refine the candidate list by scrolling through the 2D widget. Scrolling is controlled by the three remaining fingers (middle, ring, and little fingers) not involved in the pinch gesture. An upward movement of these fingers beyond a 30° threshold triggers an upward scroll, while a downward movement beyond 70° triggers a downward scroll. This threshold asymmetry accounts for the physiological limitations of finger flexion compared to extension. During scrolling, the currently focused item is highlighted with a red outline on both the widget list and the corresponding 3D object in the virtual environment. To minimize tracking noise and instability caused by finger occlusion, the system calculates the average angular value of the three fingers. Additionally, a state-based filtering mechanism and a 0.2-second temporal refractory period are implemented to

prevent unintended reverse scrolling caused by finger recoil. Final selection is confirmed when the user rotates their wrist outward by at least 50° relative to the orientation recorded at the moment the pinch gesture was initiated.

4 Research Methodology

To verify the effectiveness of the proposed technique, we conducted a user study in a VR environment. The study was designed as a comparative experiment against existing selection methods. The primary task required participants to accurately select a target sphere marked in green among numerous white spheres positioned in front of them, and participants repeatedly performed only this task. The spheres were generated within a spherical volume with a radius of 2.5 meters, the center of which was positioned 3.6 meters directly in front of the participant. To prevent severe overlapping and excessive clustering of the objects, Poisson disk sampling was applied during the generation of the spheres (Robert 2007).

The independent variables were Interaction Technique (Raycasting, DHWS, and SHGWS) and Target Density (Low, Normal, and High). To minimize the effects of presentation order, the sequence of experimental conditions was counterbalanced using a Latin Square design (Richardson 2018). The number of generated spheres corresponding to each density level is detailed in the Table 1.

Table 1
Sphere generation by target density

Target Density	Number of spheres	Distance between spheres (m)
Low	50	1.0 ~ 1.5
Normal	75	1.0 ~ 1.25
High	100	0.75 ~ 1,0

We measured Task Completion Time (TCT), Error Rate (ER), Workload, Usability, and Preference as dependent variables. TCT and ER were automatically logged through the client, while data for workload, usability, and preference were collected via questionnaires. Specifically, workload was assessed using the NASA Task Load Index (NASA-TLX) (Hart 2006), and usability was evaluated based on the System Usability Scale (SUS) (Lewis 2018). The questionnaire concluded by asking participants to rank the interaction techniques in order of preference.

We selected Raycasting as the primary baseline technique for comparison. While many target selection studies tend to adopt methods proposed in recent academic literature (Sun et al. 2025; Kim et al. 2022), our study aims to verify whether SHGWS is effective in the actual VR market by comparing it against Raycasting, which is the most widely used technique in real-world VR environments. Furthermore, we designed Double Hand Widget Selection (DHWS) as an intermediate technique between Raycasting and SHGWS. DHWS operates similarly to SHGWS but utilizes both hands by adding the non-dominant hand. The pre-selection phase works identically to SHGWS; however, to scroll and select items from the

generated widget, the user uses the opposite hand via Raycasting. We adopted DHWS as a comparative condition because Raycasting is the conventional method for interacting with widgets in VR, and it allows us to clearly analyze the differences between unimanual and bimanual interactions.

4.1 User Study

The primary task of the experiment was to accurately select a target object, indicated by a green sphere, among numerous white distractor spheres positioned in front of the user. Participants performed this task repeatedly throughout the experiment. A total of 27 participants (13 males, 13 females, 1 other) volunteered for the study. Among them, 26 were right-handed and one was left-handed. Twenty-six participants had prior experience with VR, while one had no prior experience. No monetary compensation was provided for participation.

Prior to the experiment, participants were provided with a brief overview of the study to help them understand the research objectives. Following this, they completed a short demographic questionnaire and were given clear instructions regarding the experimental procedure.

The experiment was conducted using a within-subjects design involving three interaction techniques across three target density levels. To ensure participants thoroughly understood each interaction technique, they were required to watch a demonstration video prior to the trials. Following the demonstration, participants underwent a standardized tutorial session. The tutorial required a minimum of 10 successful target selections to ensure proficiency, with participants given the option to continue practicing until they felt comfortable with the technique.

Once participants confirmed they were proficient with the technique, the main experiment commenced. The core task involved selecting the green target sphere among the white distractor spheres, with each condition repeated 15 times. Given the three interaction techniques and three target density levels, a total of 3,645 trials were collected (27 participants × 15 trials × 3 techniques × 3 densities). Upon completing each interaction technique, participants were asked to fill out NASA-TLX, SUS questionnaires. After completing all interaction techniques, participants were asked to complete a final preference questionnaire and provide additional subjective feedback. Additionally, participants were invited to engage in a brief exit interview to share their personal insights. The entire experiment took approximately 30 to 40 minutes per participant. Prior to the experiment, informed consent was obtained from all individual participants. The study protocol was approved by the Institutional Review Board (IRB) of Kyung Hee University (Approval No. KHGIRB-25-250(EA)).

4.2 Hypotheses

Based on the design rationale of SHGWS and the selection of our baseline techniques, we formulated the following two hypotheses for user study:

- **H1**: SHGWS will achieve faster and more accurate selection performance compared to Raycasting, while resulting in higher user satisfaction. This trend is expected to become more pronounced as

target density increases.

- **H2**: Despite being a unimanual technique, SHGWS will demonstrate selection performance and user experience comparable to the bimanual DHWS.

The formulation of H1 is based on the distinct operational characteristics of Raycasting and SHGWS. Raycasting, as a single-step selection method, is efficient in low-density environments but prone to increased task completion time and error rates in high-density scenarios. In contrast, we anticipate that SHGWS will maintain consistent performance regardless of target density, owing to its use of volumetric area selection. Nevertheless, as SHGWS involves a three-stage selection process, our study specifically focuses on quantifying the extent of the performance gap in task completion time between SHGWS and Raycasting within low-density environments.

H2 was formulated to determine whether reducing the number of hands required for interaction—comparing SHGWS to DHWS—maintains comparable performance and satisfaction. Demonstrating that a unimanual technique can achieve the efficacy of a bimanual method while satisfying user needs would signify a meaningful contribution toward increasing interaction freedom in virtual environments.

4.3 Apparatus

The experimental environment was constructed using a Meta Quest 3 VR headset, utilizing its built-in vision-based hand-tracking system. Participants were required to wear only the HMD, with no additional external devices. The HMD was connected directly to a Windows 11 PC via a USB-C cable. The PC configuration included an AMD Ryzen 7 processor, 32 GB of RAM, and an NVIDIA RTX 5060 Laptop GPU. The experimental software was developed in-house using Unity 6000.0.43f1. We utilized the Meta Interaction SDK for hand gesture recognition and developed custom C# scripts for tracking finger articulation and logging interaction data.

5 Result

To verify the effectiveness of the proposed technique for object selection in dense VR environments, statistical analyses were conducted on the data collected from 27 participants. Prior to the main analysis, a Shapiro-Wilk test was performed on the quantitative data, which revealed that the data did not follow a normal distribution. Consequently, we applied the Aligned Rank Transform (ART) procedure, followed by a Repeated Measures ANOVA (RM-ANOVA) (Wobbrock et al. 2011). Post-hoc analyses were conducted using Bonferroni corrections. The datasets generated during the current study are available in the Supplementary Information.

5.1 Task Completion Time

A significant main effect of interaction technique on Task Completion Time was observed ($F(2, 52) = 4.73$, $p = 0.013$). The main effect of target density was also found to be significant ($F(2, 52) = 18.81$, $p < 0.001$). However, there was no significant interaction effect between technique and density ($F(4, 104) =$

1.97, p = 0.105). Post-hoc pairwise comparisons revealed that SHGWS significantly reduced the Task Completion Time compared to Raycasting (p = 0.010). In contrast, no statistically significant differences were found when comparing DHWS to either Raycasting or SHGWS (p = 0.357).

5.2 Error Rate

The statistical analysis of Error Rate revealed significant main effects for both interaction technique ($F(2, 52) = 100.44$, $p < 0.001$) and target density ($F(2, 52) = 35.35$, $p < 0.001$). Furthermore, a significant interaction effect between technique and density was observed ($F(4, 104) = 15.99$, $p < 0.001$). Post-hoc pairwise comparisons showed that both SHGWS and DHWS exhibited significantly lower error rates compared to Raycasting ($p < 0.001$). However, no statistically significant difference was found between SHGWS and DHWS ($p = 0.198$).

5.3 Workload

Regarding user workload, a significant main effect of interaction technique was observed on the overall NASA-TLX score ($F(2, 52) = 60.46$, $p < 0.001$). Compared to Raycasting, both SHGWS and DHWS demonstrated significantly lower workload across all six subscales: Mental Demand ($p < 0.001$), Physical Demand ($p < 0.01$), Temporal Demand ($p < 0.01$), Performance ($p < 0.001$), Effort ($p < 0.001$), and Frustration ($p < 0.001$). Furthermore, no statistically significant differences were found between SHGWS and DHWS across any of the workload metrics.

5.4 Usability

The results of the usability evaluation revealed a significant main effect of interaction technique ($F(2, 52) = 8.79$, $p < 0.001$). Post-hoc pairwise comparisons showed that both DHWS ($p < 0.001$) and SHGWS ($p < 0.05$) exhibited significantly higher usability scores compared to Raycasting. Furthermore, no statistically significant difference in usability was found between DHWS and SHGWS ($p = 0.479$).

5.5 Preferences

At the conclusion of the survey, we investigated the overall user preference for each interaction technique. SHGWS emerged as the most preferred method among participants, followed by DHWS. Raycasting was ranked as the least preferred technique.

6 Discussion

In this study, we proposed a novel single-hand precision selection technique for dense VR environments. Through an empirical user study, its significant effectiveness is confirmed.

The results of this study demonstrate that both SHGWS and DHWS exhibited superior objective performance, statistically outperforming Raycasting in selection tasks. Specifically, compared to

Raycasting, both SHGWS and DHWS showed statistically significant improvements in Task Completion Time, Error Rate, and Workload.

Despite requiring more operational steps than Raycasting, SHGWS achieved faster task completion times even in the low-density condition. Raycasting operates in two steps: pointing at an object and performing a pinch gesture. In contrast, SHGWS involves a four-step process: pointing at the object, performing the pinch gesture, scrolling via finger gestures, and finally twisting the wrist to confirm the selection. Even though SHGWS entails two additional steps, it did not exhibit any increase in execution time and instead demonstrated higher accuracy across all density levels. This suggests that SHGWS provides a highly accurate selection mechanism while operating at a speed comparable to or faster than Raycasting. Furthermore, we analyze that the introduction of the widget and Area Detection significantly reduced the time required to visually locate and target the object, thereby completely offsetting the time penalty incurred by the additional interaction steps.

Furthermore, when compared to DHWS, which separates the interaction tasks into both hands, SHGWS exhibited statistically equivalent performance across all objective metrics. Essentially, SHGWS compresses the bimanual operations of DHWS—which incorporates the traditional raycast-based widget selection—into a single-handed technique. This finding indicates that SHGWS can perform the exact same functions as DHWS using only one hand, without experiencing any significant performance degradation.

In the system usability evaluation, DHWS scored significantly higher than Raycasting. SHGWS also showed a statistically significant improvement over Raycasting, while no significant difference was found between SHGWS and DHWS. However, according to our observations during the experiment and subjective questionnaires, users occasionally experienced situations where gestures did not execute as intended. For instance, P17 stated, "Personally, I liked SHGWS the most, but the three-finger scrolling motion for the selected objects is an unfamiliar action, making it feel a bit difficult at first." Similarly, P20 reported, "The DHWS technique seemed the most intuitive and the fastest to adapt to." Based on these responses, we attribute this to the inherent instability of vision-based hand tracking systems and the presence of a learning curve. In other words, tracking instability and the initial learning load are identified as the primary factors that can occasionally degrade usability.

Although SHGWS did not achieve higher usability scores than DHWS, its primary significance lies in enabling the unimanual execution of bimanual tasks. Being able to perform complex selection processes with only one hand frees up the other hand for concurrent secondary tasks, thereby contributing to increased interaction freedom in virtual environments.

However, participants selected SHGWS as their most preferred method. Positive feedback was also reported. For example, P16 shared, "Having to snap my wrist to make a selection provided a sense of physical feedback, which made the interaction feel quite fun." Furthermore, as mentioned earlier, P17 also explicitly stated that SHGWS was their favorite technique.

6.1 Application of SHGWS

SHGWS is a technique that operates by detecting hand movements through the built-in cameras and sensors of a VR HMD. Therefore, it can run on various devices that support vision-based hand tracking, such as VR HMDs, XR glasses, and AR glasses, without requiring any external hardware. This implies that SHGWS can be utilized as a versatile selection technique that is independent of any specific platform.

Due to its unimanual nature, SHGWS provides the distinct advantage of enabling interaction with virtual spaces even when one hand is occupied with another task. For example, while wearing a VR HMD or XR glasses, a user can hold a physical object or perform a manual task in the real world with one hand, while effortlessly operating the virtual interface with the other. As previously discussed, this capability significantly contributes to increasing the overall degree of interaction freedom in virtual environments.

Furthermore, due to its widget-based structural characteristics, SHGWS has the potential to be extended beyond 3D object selection to 2D interface environments. In tasks such as selecting one of multiple windows or picking a specific item from a menu, SHGWS is expected to provide a more consistent interaction method compared to conventional Raycasting or bimanual techniques. However, since the experiments in this study were strictly conducted in a dense 3D virtual environment, verifying its practical effectiveness in 2D interface scenarios remains a subject for future work.

7 Limitation and Future works

In this study, we proposed SHGWS, a single-handed interaction technique designed for seamless selection in dense virtual environments. Compared to the conventional Raycasting method, SHGWS achieved shorter task completion times, lower error rates, and reduced user workload. When compared to the bimanual DHWS, although SHGWS did not achieve higher usability scores, it demonstrated comparable performance across all other metrics and recorded a higher overall user preference. However, this study has the following limitations.

First, the target objects used in our evaluation were fixed in size and shape. Real-world VR content requires interactions with objects of various sizes. However, target size was not set as an independent variable in our experiment, meaning the targets were identical across all conditions. Since target size is known to directly affect the accuracy and task completion time of selection techniques (Bergström et al. 2021), we cannot rule out the possibility that our findings are only valid under this specific target size condition. Future studies should test the technique in more complex VR environments where targets vary in size and shape, and potentially include moving targets.

Second, the scope of the comparative techniques was limited. In this study, Raycasting, the most widely used method in VR environments, was selected as the primary baseline. Additionally, DHWS—the bimanual version of SHGWS that utilizes Raycasting for post-selection—was included as a comparative condition. However, the field of VR target selection features various prior techniques, such as

HandConeGrid, MultiFingerBubble, and PianoFinger. Because a direct comparison with these methods was not conducted, there is a limitation in precisely determining the relative standing of SHGWS as a unimanual selection technique in dense environments. Future research should include a broader range of existing techniques in comparative evaluation to further validate the practical effectiveness of SHGWS.

Third, the gestures proposed in this technique were designed in a top-down manner without incorporating user feedback. Consequently, the lower usability enhancement compared to DHWS and the qualitative feedback indicating that the gestures felt difficult are natural outcomes. Nevertheless, SHGWS achieved comparable performance to DHWS in all metrics, and even showed a higher user preference. This suggests that the cause of the lower usability enhancement is more likely related to the direction of the gesture design rather than a structural flaw in the technique itself. In future work, it is necessary to design an improved, bottom-up version of SHGWS by conducting a gesture elicitation study (Villarreal et al. 2020) to collect gestures that users inherently feel are appropriate for each interaction step. By introducing user-reflected gestures while maintaining the base structure of SHGWS—which already demonstrated high preference—we expect to bridge the gap between the current usability scores and user preference.

8 Conclusion

In this study, we proposed SHGWS, a unimanual selection technique designed for dense, hand-tracked virtual environments. To enable precise selection in dense spaces, SHGWS provides a single-handed 2D widget interface. The technique performs disambiguation by dividing the process into pre-selection via Cone Detection and post-selection using gestures and a 2D widget, ultimately finalizing the selection with an outward wrist twist.

To validate the effectiveness of SHGWS, we conducted a comparative study against the conventional Raycasting method and the intermediate bimanual technique, DHWS. The results demonstrated that SHGWS delivered improved performance over Raycasting for object selection tasks in dense spaces, while showing comparable performance to DHWS.

Through this research, we addressed the limitations of the standard Raycasting technique and filled a research gap by introducing a method that completes the entire selection process in a dense environment using only one hand. Furthermore, we demonstrated that reducing a bimanual task to a unimanual one does not significantly degrade performance and can lead to higher user preference.

Although SHGWS showed better results across most performance metrics and subjective feedback than other techniques, improvements are still needed to address its relatively lower usability and to further optimize Task Completion Time. In addition, future work requires user testing in more complex VR environments and direct comparative experiments with other prior techniques to objectively determine the relative standing of user-friendly version of SHGWS.

SHGWS is a versatile technique that can be applied to various hand-tracking-enabled XR devices without requiring any additional hardware. Furthermore, its unimanual operation allows the user's other hand to perform concurrent physical or virtual tasks, providing a higher degree of interaction freedom. Through these characteristics, we hope that SHGWS can contribute to enhancing the bare-handed interaction experience and user freedom in VR environments. As XR devices become widely adopted, we expect that this approach could be utilized as a practical selection method for next-generation virtual environments.

Declarations

Author Contribution

C.H.M. conceptualized the study, developed the interaction technique and virtual environment, conducted the user study, and wrote the main manuscript text. T.W. supervised the research, refined the methodology, and provided critical revisions. All authors reviewed and approved the final manuscript.

Acknowledgement

- This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2026 (Project Name: Development of Visual and Haptic XR-AI Convergence Content and Specialist Training Program, Project Number: RS-2026-25520006, Contribution Rate: 50%)- This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) under the virtual convergence support program to nurture the best talents (IITP-2026-RS-2024-00425383) grant funded by the Korea government (MSIT)

Data Availability

The datasets generated during the current study are available in the Supplementary Information.

References

1. LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). 3D user interfaces: theory and practice. Addison-Wesley Professional.
2. Baloup, M., Pietrzak, T., & Casiez, G. (2019). Raycursor: A 3d pointing facilitation technique based on raycasting. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-12). https://doi.org/10.1145/3290605.3300331
3. Mine, M. R. (1995). Virtual environment interaction techniques. UNC Chapel Hill CS Dept.
4. Turkmen, R., Rasoulifar, Z., & Batmaz, A. U. (2026). VoiceRay: 3D Object Selection Technique for Occluded and Dense Environments in Virtual Reality. In Proceedings of the 2026 CHI Conference on

Human Factors in Computing Systems (pp. 1-19). https://doi.org/10.1145/3772318.3790399

5. Sidenmark, L., Sun, Z., & Gellersen, H. (2024). Cone&Bubble: Evaluating Combinations of Gaze, Head and Hand Pointing for Target Selection in Dense 3D Environments. In 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 642-649). IEEE. https://doi.org// 10.1109/VRW62533.2024.00126
6. Weller, R., Wegele, W., Schröder, C., & Zachmann, G. (2021). Lenselect: Object selection in virtual environments by dynamic object scaling. Frontiers in Virtual Reality, 2, 684677. https://doi.org/10.3389/frvir.2021.684677
7. Shi, R., Wei, Y., Hu, X., Liu, Y., Yue, Y., Yu, L., & Liang, H. N. (2024). Experimental analysis of freehand multi-object selection techniques in virtual reality head-mounted displays. Proceedings of the ACM on Human-Computer Interaction, 8(ISS), 93-111. https://doi.org/10.1145/3698129
8. Meng, X., Xu, W., & Liang, H. N. (2022). An exploration of hands-free text selection for virtual reality head-mounted displays. In 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 74-81). IEEE. https://doi.org/10.1109/ISMAR55827.2022.00021
9. Xu, W., Meng, X., Yu, K., Sarcar, S., & Liang, H. N. (2022). Evaluation of text selection techniques in virtual reality head-mounted displays. In 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 131-140). IEEE. https://doi.org/10.1109/ISMAR55827.2022.00027
10. Kim, J., Park, S., Zhou, Q., Gonzalez-Franco, M., Lee, J., & Pfeuffer, K. (2025). Pinchcatcher: Enabling multi-selection for gaze+ pinch. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-16). https://doi.org/10.1145/3706598.3713530
11. Buckingham, G. (2021). Hand tracking for immersive virtual reality: opportunities and challenges. Frontiers in Virtual Reality, 2, 728461. https://doi.org/10.3389/frvir.2021.728461
12. Adkins, A., Lin, L., Normoyle, A., Canales, R., Ye, Y., & Jörg, S. (2021). Evaluating grasping visualizations and control modes in a vr game. ACM Transactions on Applied Perception (TAP), 18(4), 1-14. https://doi.org/10.1145/3486582
13. Vlahovic, S., Skorin-Kapov, L., Suznjevic, M., & Pavlin-Bernardic, N. (2024). Not just cybersickness: short-term effects of popular VR game mechanics on physical discomfort and reaction time. Virtual reality, 28(2), 108. https://doi.org/10.1007/s10055-024-01007-x
14. Zavichi, M., Santos, A., Moreira, C., Maciel, A., & Jorge, J. (2025). Gaze–hand steering for travel and multitasking in virtual environments. Multimodal Technologies and Interaction, 9(6) 61. https://doi.org/10.3390/mti9060061
15. Hu, B., Liu, Y., Liu, X., Xiao, Q., Wang, X., Yuan, Z., ... & Wang, Y. (2026). Glide-in-Place: Foot-Steered Differential-Drive for Hands-Free VR Locomotion. arXiv preprint arXiv:2604.10237. https://doi.org/10.48550/arXiv.2604.10237
16. Yu, D., Dingler, T., Velloso, E., & Goncalves, J. (2024). Object selection and manipulation in VR headsets: Research challenges, solutions, and success measurements. ACM Computing Surveys, 57(4), 1-34. https://doi.org/10.1145/3706417

17. Wu, H., Sun, X., Tu, H., & Zhang, X. (2023). Clockray: A wrist-rotation based technique for occluded-target selection in virtual reality. IEEE Transactions on Visualization and Computer Graphics, 30(7), 3767-3778. https://doi.org/10.1109/TVCG.2023.3239951
18. Yu, D., Zhou, Q., Newn, J., Dingler, T., Velloso, E., & Goncalves, J. (2020). Fully-occluded target selection in virtual reality. IEEE transactions on visualization and computer graphics, 26(12), 3402-3413. https://doi.org/10.1109/TVCG.2020.3023606
19. Chen, D. L., Giordano, M., Benko, H., Grossman, T., & Santosa, S. (2023). Gazeraycursor: Facilitating virtual reality target selection by blending gaze and controller raycasting. In Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology (pp. 1-11). https://doi.org/10.1145/3611659.3615693
20. Liu, H., Zhang, Z., Jiao, Z., Zhang, Z., Li, M., Jiang, C., ... & Zhu, S. C. (2024). A reconfigurable data glove for reconstructing physical and virtual grasps. Engineering, 32, 202-216. https://doi.org/10.1016/j.eng.2023.01.009
21. Polydorou, N., & Edalat, A. (2021). An interactive VR platform with emotion recognition for self-attachment intervention. EAI Endorsed Transactions on Pervasive Health & Technology, 7(29), 1. https://doi.org/10.4108/eai.14-9-2021.170951
22. Moon, H. S., Orr, G., & Jeon, M. (2023). Hand tracking with vibrotactile feedback enhanced presence, engagement, usability, and performance in a virtual reality rhythm game. International Journal of Human–Computer Interaction, 39(14), 2840-2851. https://doi.org/10.1080/10447318.2022.2087000
23. Chua, P., Sasikumar, P., Weerasinghe, Y., & Nanayakkara, S. (2024). Motion as Emotion: Detecting Affect and Cognitive Load from Free-Hand Gestures in VR. arXiv preprint arXiv:2409.12921. https://doi.org/10.48550/arXiv.2409.12921
24. Song, Z., Dudley, J. J., & Kristensson, P. O. (2023). Hotgestures: Complementing command selection and use with delimiter-free gesture-based shortcuts in virtual reality. IEEE Transactions on Visualization and Computer Graphics, 29(11), 4600-4610. https://doi.org/10.1109/TVCG.2023.3320257
25. Shi, R., Zhang, J., Yue, Y., Yu, L., & Liang, H. N. (2023). Exploration of bare-hand mid-air pointing selection techniques for dense virtual reality environments. In Extended abstracts of the 2023 CHI conference on human factors in computing systems (pp. 1-7). https://doi.org/10.1145/3544549.3585615
26. Delamare, W., Daniel, M., & Hasan, K. (2022). Multifingerbubble: A 3d bubble cursor variation for dense environments. In CHI Conference on Human Factors in Computing Systems Extended Abstracts (pp. 1-6). https://doi.org/10.1145/3491101.3519692
27. Sun, M., Guo, D., Yang, F., & Liu, Y. (2025). Enhanced target selection in dense VR environments using single-hand frozen gestures. International Journal of Human–Computer Interaction, 41(21), 13495-13504. https://doi.org/10.1080/10447318.2025.2474473

28. Han, D., Kim, D., & Cho, I. (2022). Portal: Portal widget for remote target acquisition and control in immersive virtual environments. In Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (pp. 1-11). https://doi.org/10.1145/3562939.3565639
29. Kim, W., & Xiong, S. (2022). ViewfinderVR: configurable viewfinder for selection of distant objects in VR. Virtual Reality, 26(4), 1573-1592. https://doi.org/10.1007/s10055-022-00649-z
30. Debarba, H. G., Grandi, J. G., Maciel, A., Nedel, L., & Boulic, R. (2013). Disambiguation canvas: A precise selection technique for virtual environments. In IFIP Conference on Human-Computer Interaction (pp. 388-405). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40477-1_24
31. Maslych, M., Hmaiti, Y., Ghamandi, R., Leber, P., Kattoju, R. K., Belga, J., & LaViola, J. J. (2023). Toward intuitive acquisition of occluded vr objects through an interactive disocclusion mini-map. In 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR) (pp. 460-470). IEEE. https://doi.org/10.1109/VR55154.2023.00061
32. Liang, J., & Green, M. (1994). JDCAD: A highly interactive 3D modeling system. Computers & graphics, 18(4), 499-506. https://doi.org/10.1016/0097-8493(94)90062-0
33. Sidenmark, L., Clarke, C., Zhang, X., Phu, J., & Gellersen, H. (2020). 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). https://doi.org/10.1145/3313831.3376438
34. Luo, X., Moon, H. S., Holz, C., & Oulasvirta, A. (2026). Point & Grasp: Flexible Selection of Out-of-Reach Objects Through Probabilistic Cue Integration. In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (pp. 1-19). https://doi.org/10.1145/3772318.3790836
35. Robert Bridson. 2007. Fast Poisson Disk Sampling in Arbitrary Dimensions. In ACMSIGGRAPH2007 Sketches (San Diego, California) (SIGGRAPH '07). Association for Computing Machinery, New York, NY, USA, 22–es. https://doi.org/10. 1145/1278780.1278807
36. Richardson, J. T. (2018). The use of Latin-square designs in educational and psychological research. Educational Research Review, 24, 84-97. https://doi.org/10.1016/j.edurev.2018.03.003
37. Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications. https://doi.org/10.1177/154193120605000909
38. Lewis, J. R. (2018). The system usability scale: past, present, and future. International Journal of Human–Computer Interaction, 34(7), 577-590. https://doi.org/10.1080/10447318.2018.1455307
39. Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 143-146). https://doi.org/10.1145/1978942.1978963
40. Bergström, J., Dalsgaard, T. S., Alexander, J., & Hornbæk, K. (2021). How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In proceedings of the 2021 CHI

conference on human factors in computing systems (pp. 1-20). https://doi.org/10.1145/3411764.3445193

41. Villarreal-Narvaez, S., Vanderdonckt, J., Vatavu, R. D., & Wobbrock, J. O. (2020). A systematic review of gesture elicitation studies: What can we learn from 216 studies?. In Proceedings of the 2020 ACM designing interactive systems conference (pp. 855-872). https://doi.org/10.1145/3357236.3395511

Figures

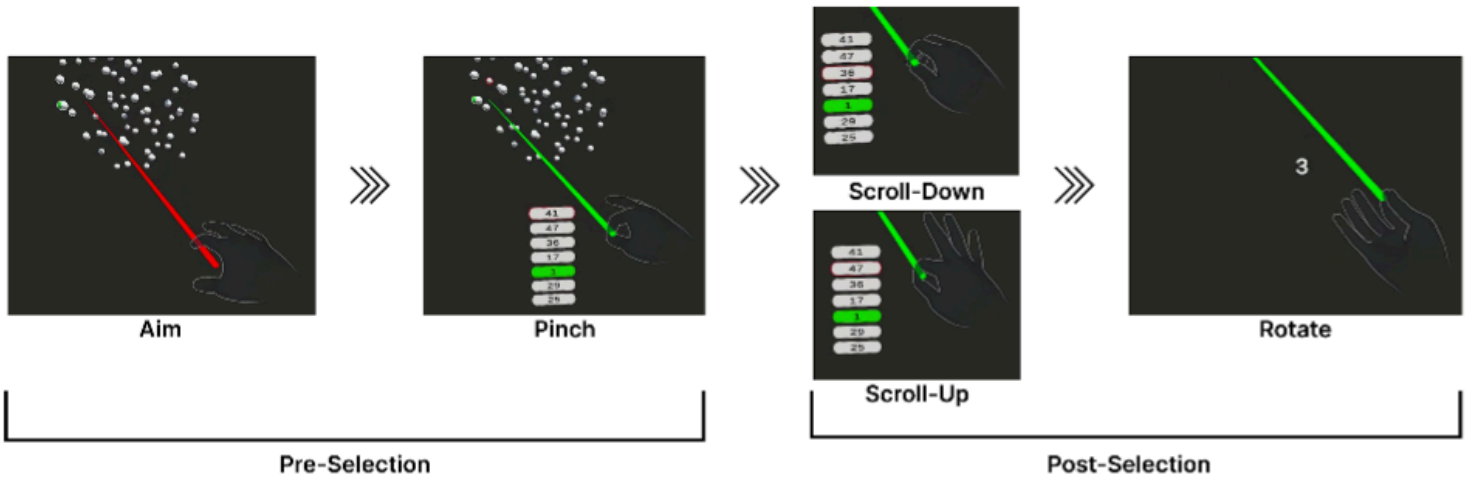

Figure 1

Procedure of Single-Hand Gesture Widget Selection

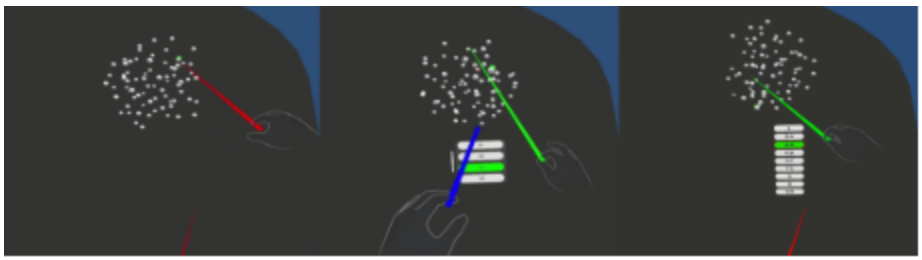

Figure 2

Raycast, Double Hand Widget Selection (DHWS), Single Hand Gesture Wideget Selection (SHGWS)

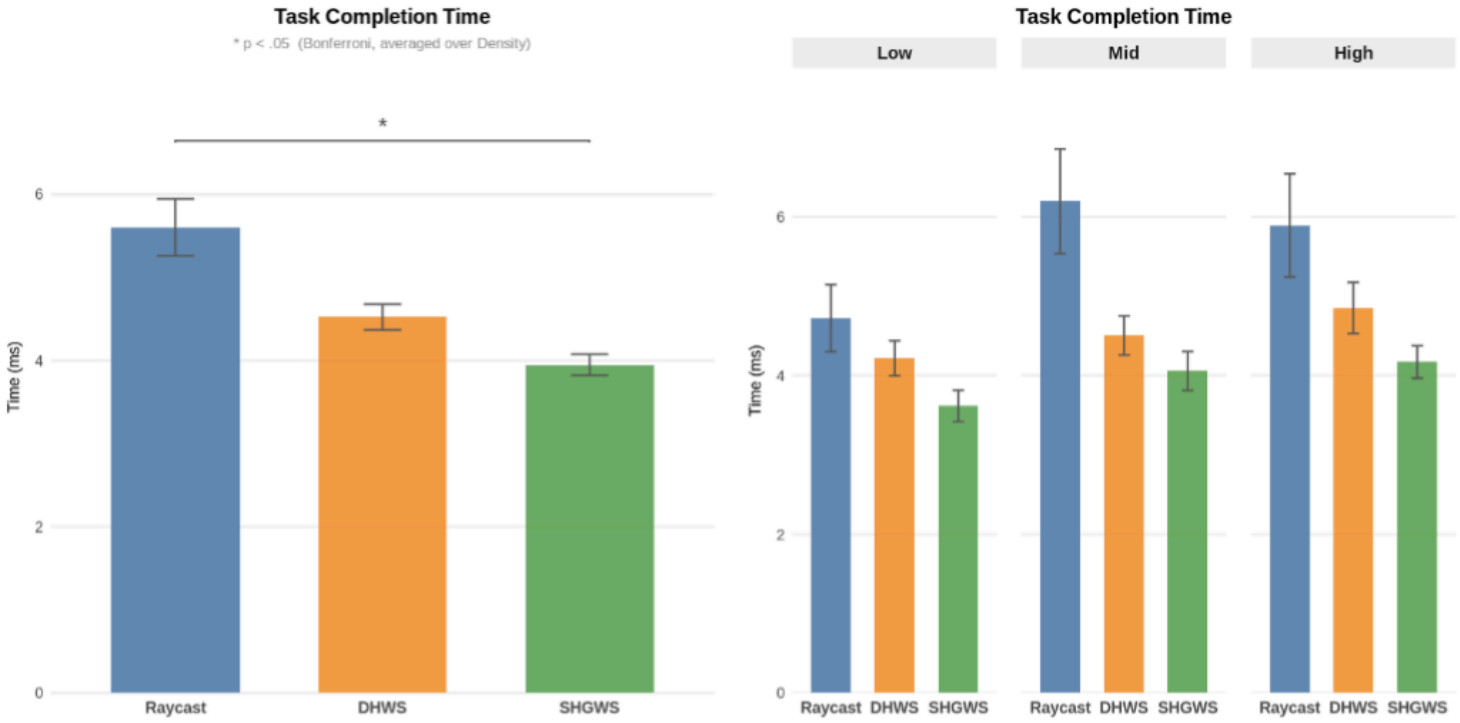

Figure 3

Task Completion Time

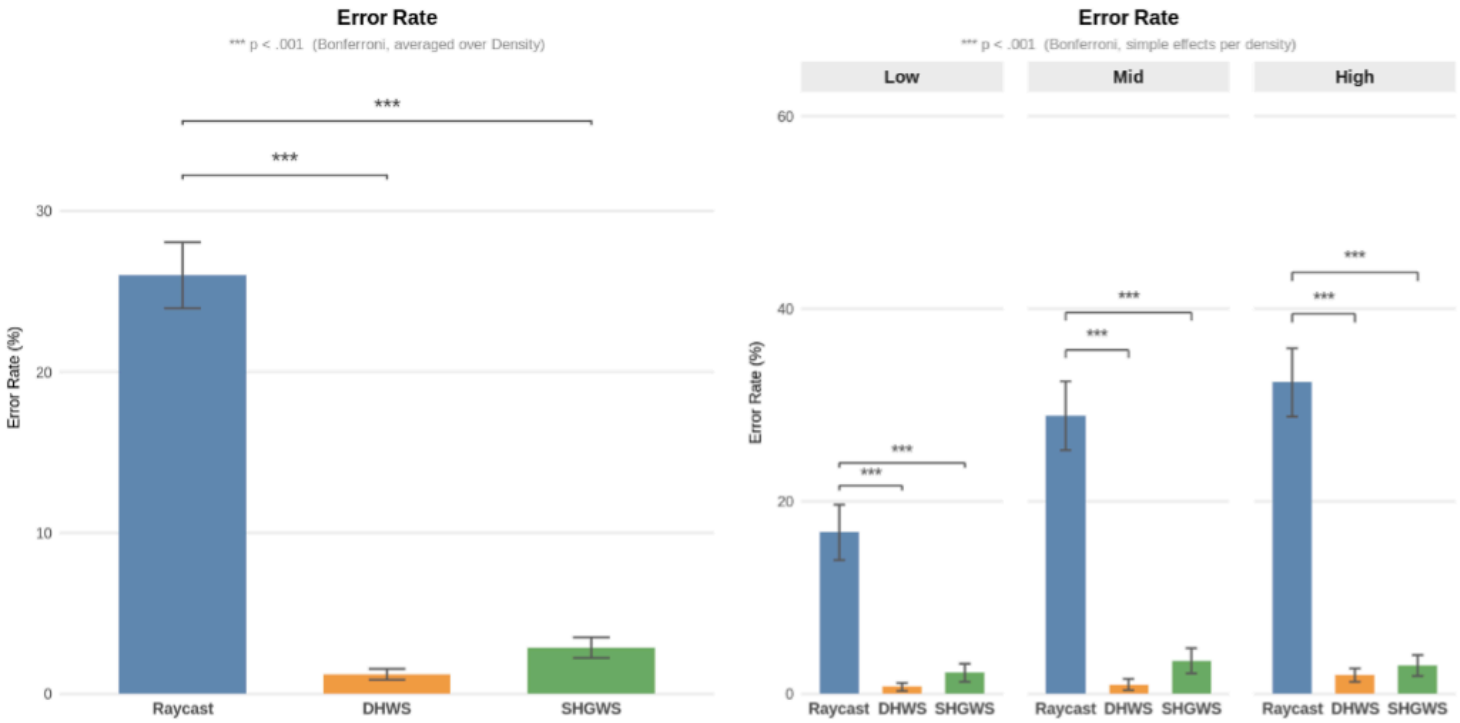

Figure 4

Error Rate

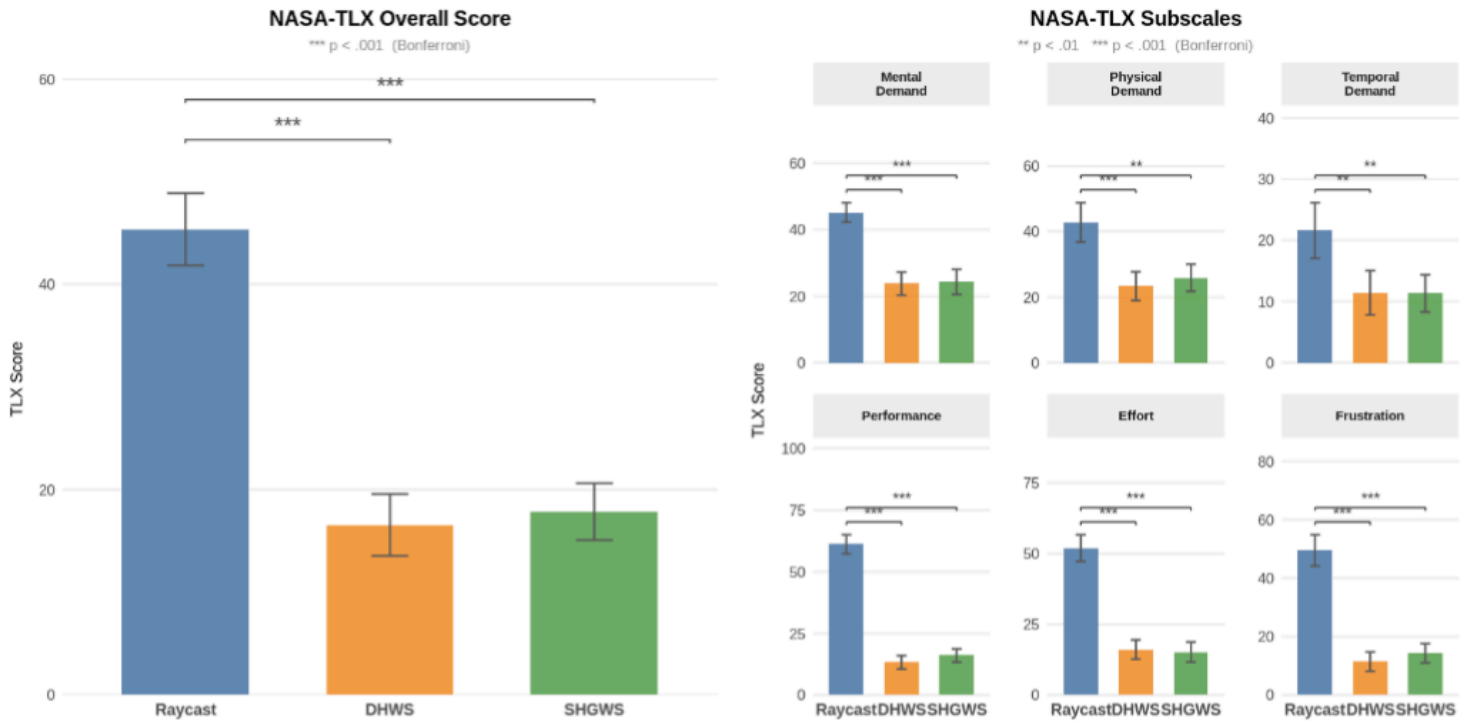

Figure 5

Workload (NASA-TLX Scores)

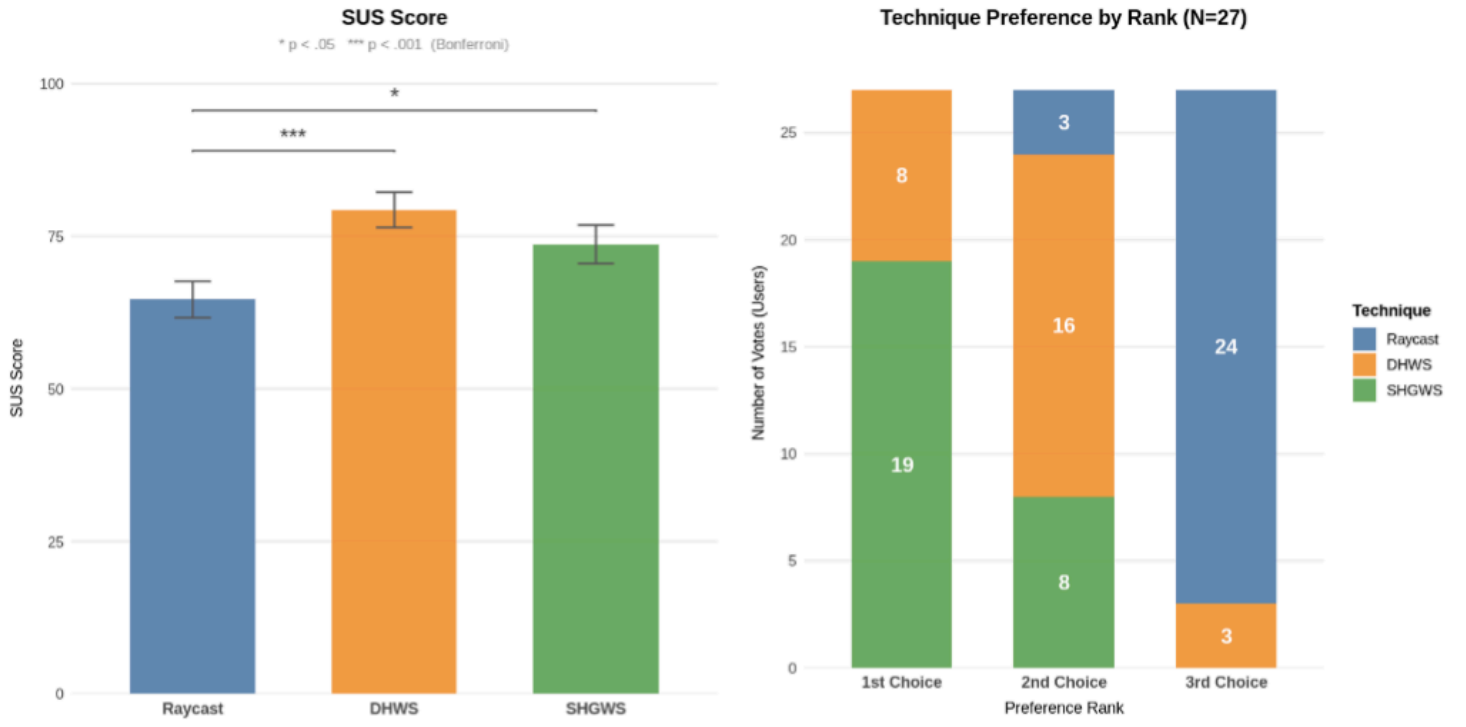

Figure 6

Usability (SUS Score), Preferences

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformations.zip