Is the Jedi Force Pull Method Effective? Evaluating Eye-Tracked Object Grabbing in VR

Naznin Sharmin Shishir
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada
naznin.shishir1@ucalgary.ca

Richard Zhao
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada
richard.zhao1@ucalgary.ca

Abstract

Virtual reality (VR) has become an increasingly popular medium for many purposes, including training, education, and video games. While handheld controllers are traditionally used for interaction in VR environments, alternative methods such as hand tracking, eye tracking, and speech recognition have gained traction. Eye tracking, in particular, has seen adoption with the release of commercial VR headsets like the Meta Quest Pro and Apple Vision Pro. This paper addresses the question: is eye tracking an effective method for grabbing objects from a distance, allowing for an interaction method similar to "Force Pull" seen in *Star Wars*? We examine the accuracy and effectiveness of eye-tracked object grabbing in VR, comparing it to non-eye-tracking methods. We develop an interactive VR environment to measure time, accuracy, and task load for eye-tracked object grabbing, distance hand grabbing, and close-up object grabbing interactions with 3D objects. Our results provide the indications that eye tracking outperformed the other two methods across multiple metrics.

CCS Concepts

• **Human-centered computing → Virtual reality**; **Empirical studies in interaction design**.

Keywords

eye tracking, virtual reality, VR, interactions, object grabbing, distance grabbing

ACM Reference Format:
Naznin Sharmin Shishir and Richard Zhao. 2025. Is the Jedi Force Pull Method Effective? Evaluating Eye-Tracked Object Grabbing in VR. In *International Conference on the Foundations of Digital Games (FDG '25), April 15–18, 2025, Graz, Austria.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3723498.3723695

1 Introduction

Eye tracking is among the most extensively studied hands-free interaction methods in virtual reality (VR) environments [22]. With the release of newer commercial VR headsets such as the Meta Quest Pro and Apple Vision Pro, eye tracking has seen broader adoption as a method of interaction in VR. This technology has shown promise in creating adaptive VR systems and enhancing user experiences [17], with applications across fields such as behavioral science, education, medicine, design, and virtual reality [21]. Eye tracking integrated into head-mounted displays (HMDs) has become especially popular for interaction in VR-based education and learning contexts [21]. Prior research indicates that eye gaze reaches objects before other interaction methods are applied [26], making eye tracking a faster option for selection and pointing tasks in VR [22, 27].

Through individual gaze calibration, eye tracking can achieve the accuracy needed for precise interaction [23], potentially serving as an alternative for users with physical disabilities [6]. Recently, hand tracking (without holding a tracking device such as a controller) has also gained acceptance as a viable replacement for traditional hand controllers [18]. However, while its performance varies across contexts compared to hand controllers [15], the efficiency of hand tracking techniques as a selection method relative to eye tracking remains under-explored. Furthermore, few studies have examined the effectiveness of eye-tracking interactions in VR compared to hand controllers [8, 14, 19], and none, to date, have directly compared eye tracking with hand tracking methods for grabbing objects.

The goal of this paper is to examine the effectiveness of eye tracking as an object grabbing method in VR compared to two hand-tracking-based grabbing techniques. In summary, the contributions of this paper are as follows:

- We design a VR environment to test the effectiveness of various object interaction methods, and evaluate the effectiveness of eye tracking as an object-grabbing method within a VR system using eye-finger multimodal interaction.
- We provide a comparative analysis of three methods, eye-tracked object grabbing, distance hand grabbing, and close-up object grabbing interactions in VR by examining user preferences, accuracy, task completion time, and task load.
- We analyze participant results across different genders, highlighting equitable access to eye-tracked interactions in VR.

2 Related Works

2.1 VR Interactions

Although controllers are traditionally the most common interaction method in VR, alternative methods are gaining wider acceptance among users. The most prevalent of these are hand tracking, head tracking, speech interaction, and eye tracking. In a systematic review, Monteiro et al. [22] identify speech interaction as the most widely used hands-free interface, followed by eye and head gaze

FDG '25, Graz, Austria

ACM ISBN /25/04
https://doi.org/10.1145/3723498.3723695

interactions. Previous studies also compare the usability of hand tracking and controller-based interactions in virtual environments [12, 15, 16, 18, 36]. For example, Hameed et al. [12] evaluate play length, click frequency, selection frequency, and mental effort in a reach-pick-place VR task, finding that handheld controllers outperform hand tracking across all metrics. Similarly, Johnson et al. [15] compare hand tracking and controller-based interactions in a ball-sorting task and observe that hand tracking leads to poorer performance and a reduced sense of naturalness. Bothén et al. [5] explore head gaze as a means for implementing common game interactions such as aiming and walking. Additionally, sensors such as electroencephalogram (EEG) sensors [34] and cameras [31] have been used in VR for various purposes.

Another study assesses various VR interaction methods and finds that participants prefer controller interactions with controller visualization over hand tracking in different tasks [36]. Luong et al. [18] examine the effects of VR controllers and free-hand interaction on effort, performance, and motor behavior in selection and trajectory tracing tasks. Their findings show that participants feel more in control, experience less physical effort, and perform faster and more accurately using VR controllers in a raycast setting. However, in mid-air settings, hands-free interaction proves to be more effective than controllers. Khundam et al. [16] compare VR controllers and hand tracking across four interaction types and find no significant differences in accuracy, preference, or usability between the methods. Although controllers slightly outperform hand tracking on the System Usability Scale, hand tracking is perceived as more useful and realistic, particularly in medical training contexts.

Aslam et al. [3] develop a voice-augmented virtual interface, comparing natural language commands to hand controllers for VR interactions. The voice-augmented interface achieves lower error rates, higher precision, and similar efficiency to hand controllers. The mixed results in prior studies indicate that there is no conclusive evidence favoring one interaction method over another across all applications.

2.2 Eye Tracking in VR

The integrated eye-tracking capabilities in the latest HMDs are gaining popularity and admiration. Several surveys reference different methods and applications of eye tracking in existing research [1, 17, 21, 23, 29]. Mikhailenko et al. [21] provide an overview of fields where eye tracking is applied, including behavioral science, education, medicine, design, and virtual reality, with a particular focus on eye tracking in VR for educational applications. Eye tracking can assist in developing adaptive VR systems and enhance the user experience [17].

Some studies evaluate the performance of eye tracking for selection when combined with hand tracking for manipulation [6, 7, 25, 26, 28]. Pfeuffer et al. [28] develop an experimental interface enabling participants to perform tasks by selecting 3D objects with eye gaze and manipulating them using hand pinch gestures. Based on user feedback and observations, the proposed method is both useful and innovative, surpassing traditional real-world interactions. Cecotti et al. [6] present a virtual keyboard designed for users with disabilities, where users can point to specific keys using eye gaze and select them by performing one of eight hand gestures. Chen et al. [7] examine eye-hand coordination patterns among children and adults while playing two different video games. Mutasim et al. [26] explore gaze behavior in a VR eye-hand coordination training system for sports, finding that participants locate an object with their gaze approximately 0.25 seconds before touching it with their finger.

Plopski et al. [29] include multimodal interactions along with eye-only interactions in their survey. Wei et al. [37] test the performance of eye tracking on the Meta Quest Pro headset under both head-free and head-restrained conditions, finding the Meta Quest Pro to be a viable option for eye tracking, with signal quality comparable to existing augmented / virtual reality eye-tracking headsets.

Some studies compare the efficiency of eye tracking in selection or aiming tasks with other interaction methods, such as hand controllers [14, 19], head tracking [4, 8], and gamepad input [27]. Luro et al. [19] compare gaze-based aiming in VR to traditional controller-based aiming in an "aim and shoot" task. The usability scores show no significant difference in performance between gaze and controller aiming, although gaze performs better overall in terms of speed and when targeting unpredictable paths. Blattgerste et al. [4] contrast head-gaze and eye-gaze aiming to assess the advantages of eye-gaze-based engagement techniques in VR and AR. Their findings show that eye-gaze outperforms head-gaze in speed, task load, head movement, and user preference. Pai et al. [27] propose a system combining gaze tracking with forearm electromyography for VR selection tasks and compare it with gamepad input, motion-tracked controllers, gaze direction with dwell time, and eye-gaze direction with dwell time. Their results show that eye-gaze selection with forearm contraction is a quick and efficient targeting method for VR applications.

Choe et al. [8] evaluate the performance of head tracking, manual controllers, and gaze input techniques to determine the best timing and input method for VR. They find that manual controllers yield the shortest task completion times, while gaze input provides the highest accuracy. Hou et al. [14] investigate eye-based selection in a VR environment using Fitts' Law modeling and compare three techniques: controller selection with confirmation, eye-gaze selection with dwell confirmation, and eye-gaze selection with controller confirmation in 2D and 3D environments. Their findings indicate that eye-gaze accuracy is higher for 3D objects, whereas controller interaction is more accurate for 2D objects. However, eye dwell performs worse in terms of both speed and accuracy. Mutasim et al. [25] design a task to examine hand pinch performance under Fitts' Law, where objects are targeted with eye gaze and activated through pinching, button clicks, or dwell. Their results suggest that pinch gestures serve as a viable alternative to button clicks in eye-gaze-based VR systems, though participants prefer button clicks and dwell due to pinch recognition issues with the Leap Motion device.

Danion et al. [10] investigate gaze behavior while tracking a moving target with either eye tracking or hand tracking, finding that eye tracking produces more accurate results than hand tracking. Additionally, they note that participants keep their gaze closer to the target in both types of experiments, even when not instructed to do so during hand tracking. Vertegaal [35] compares two eye-tracking input techniques—manual click and dwell time click—using Fitts' law, concluding that eye tracking combined with manual click outperforms dwell time click in both speed and accuracy. Sonntag et al. [33] introduce a virtual traffic environment to examine the effectiveness of eye gaze interaction in various pedestrian activities at traffic signals. A systematic review by Monteiro et al. [22] identifies eye gaze as especially well-suited for pointing and selection tasks among hands-free interfaces.

Most studies on eye tracking use satisfaction, efficiency, and efficacy as evaluation metrics, often through custom questionnaires [22]. Accuracy and error rates are the primary metrics for efficacy, while interaction times are the primary metrics for efficiency. However, eye tracking metrics in most studies are limited to fixation durations and counts, indicating a need for more extensive research to improve performance [17, 30]. Rappa et al. [30] propose an analytical framework for extended reality research using eye tracking, which includes six distinct aspects that could support future studies in this area.

Eye calibration is essential for each participant prior to eye tracking experiments to ensure an accurate gaze vector [23]. However, recalibration may be necessary during experiments to maintain the highest accuracy, which can be time-consuming and disruptive for participants [28]. Real-time recalibration is therefore valuable, allowing eye tracking systems to recalibrate autonomously without interrupting the user experience [23]. Sidenmark et al. [32] investigate the timing and likelihood of gaze fixations on objects interacted with during hand interaction in VR. They find that the optimal approach for real-time recalibration is to set the object as a calibration point, recording gaze data when participants fixate on it.

Clay et al. [9] analyze participants' eye pointing in a VR environment to assess its effectiveness. Meißnera et al. [20] review the advantages of different eye-tracking technologies for desktop, natural, and virtual environments, highlighting the benefits of mobile eye tracking specifically in VR. Duchowski et al. [11] describe advancements in binocular eye tracking within VR, where user gaze direction, head position, and orientation are tracked to better understand user actions. Moustafa et al. [24] investigate gaze position and eye-hand coordination, particularly for individuals with vision impairments, by analyzing how they read and interact in VR. Other applications of eye-tracking data include predicting cybersickness, as explored in studies by Andrei et al. [2]. Additionally, Andrei et al. [2] find that eye tracking can contribute to making VR-based conversations feel more realistic for users.

While some prior studies investigate the efficiency of eye tracking combined with hand tracking, and others compare eye tracking as a pointing method against hand controllers and head tracking, none specifically examine the usability and accuracy of eye tracking in comparison to hand tracking in VR when performing a grabbing task. Previous experiments have not explored the use of a distant hand grab gesture in combination with eye tracking as a multimodal interaction method.

This research aims to show the advantages of a multimodal interaction technique that combines eye tracking with finger movements, allowing users to select an object using eye tracking and grab it with a pinch gesture. We compare user preference, time, accuracy, and task load of the eye-tracked method against two types of hand tracking methods in VR: distance hand grabbing and close-up object grabbing. In our research, we deliberately choose not to include the use of controllers. Our ultimate goal is to improve accessibility for users in VR, and we believe that enabling VR interactions with minimal additional tools (using hands rather than controllers) is a step in the right direction.

3 Object Interaction Methods

In this research, we demonstrate the effectiveness of incorporating eye tracking into 3D object interactions in VR. We focus on one fundamental interaction as our starting point: grabbing an object. This interaction is both common and essential in VR environments. For our study, we define the following three object interaction methods:

Eye-tracked (ET) object grabbing is a multimodal interaction method in which objects are selected in a scene using the user's eye gaze. Once an object is selected, a small finger motion—a pinch gesture in this case—causes the selected object to move to the user's hand. This method minimizes body movement compared to the other methods, as the user does not need to move their arm throughout the process, evoking scenes from *Star Wars* where the Jedi characters can summon objects to their hands with minimal motion. Figure 1 shows ET object grabbing in action for a user wearing a VR headset.

Distance hand (DH) object grabbing requires the user to raise and extend their hand to point at an object. When an object is pointed at, it becomes selected, and a pinch or grab motion brings the object to the user's hand. Figure 2 shows DH grabbing.

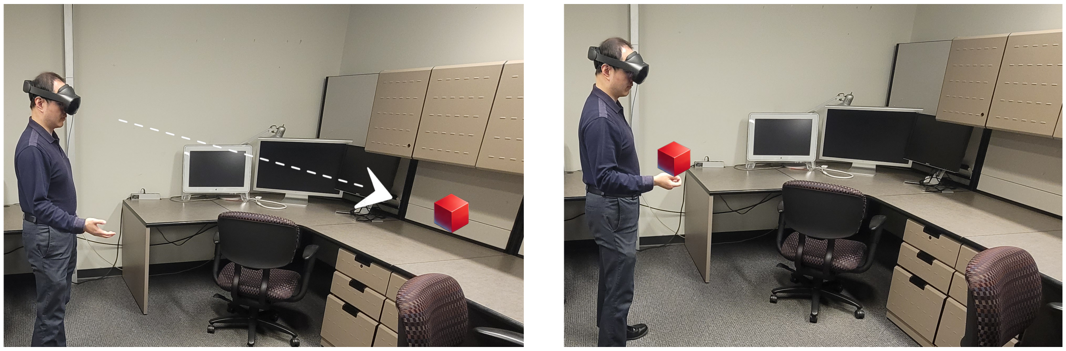
Close-up (CU) object grabbing involves grabbing an object with the hand at a close distance, similar to how one would interact with an object in the real world. This method serves as our baseline for comparison. Additionally, we aim to observe whether participants prefer this method due to its resemblance to real-life interactions. Figure 3 shows CU object grabbing. In this example, the boundary of the VR environment is set to match the boundary of the physical world, so the user is safe when performing VR tasks.

4 VR Environment for Object Interactions

To implement the VR setup for this study, we developed the experimental project using Unity Engine, leveraging the Oculus Integration SDK (version 57.0.1) to obtain eye tracking and hand tracking data. The project was primarily developed in C#. By integrating the OVR eye gaze script with custom eye interactor and eye interactable scripts, we implemented new functionalities and tailored the setup within our VR scenes.

The eye tracking used in this study is based on the Oculus Integration SDK, with two eye interactors developed to support left and right eye gaze. An eye-tracking interactor script is attached to each eye, responsible for managing eye tracking behaviors. To allow interaction with distant objects, the ray distance was set to 20. The eye-tracking raycast was made transparent to avoid additional visual effort in targeting objects. An eye interactable script was written and attached to each 3D interactable object, enabling appropriate functionalities when the eye-tracking ray intersects with an interactable object.

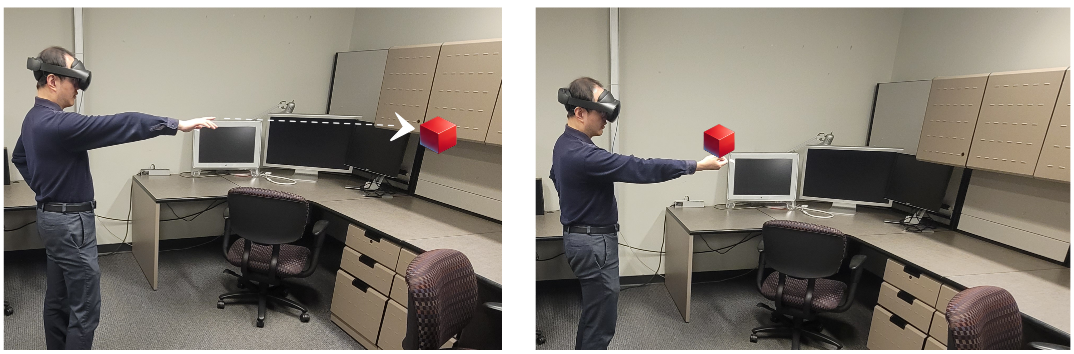
In our eye tracking setup, we implemented a multimodal interaction technique where object selection occurs via eye tracking and



(a) Eye-tracked object selection.

(b) Grabbing an eye-tracked object.

Figure 1: Eye-tracked object grabbing. (a) illustrates that an object can be selected by the user looking at the object. No arm movement is needed. (b) illustrates that with a pinch motion of the hand, the selected object flies to the user's hand.



(a) Distance hand object selection.

(b) Grabbing a selected object.

Figure 2: Distance hand grabbing. (a) illustrates that an object can be selected by the user extending the hand to point at the object. (b) illustrates that with a pinch or grab motion of the hand, the selected object flies to the user's hand.

Figure 3: Close-up object grabbing. The user needs to walk up to the object and grab the object similar to grabbing an object in real life.

manipulation is managed through hand gesture. Users can select a 3D object by gazing at it and then grab it by performing a pinch gesture. Here are the functions of the eye interactable script:

- Selection (when eye gaze hits an eye interactable)
- Grab (when user grabs an object)
- Idle (when neither selection nor grab happens)

The DH grab tracking is implemented by attaching a distance hand grab interactor prefab to each hand interactor, for both the left and right hands. The DH grab interactor is responsible for detecting the hand grab posture and facilitating distance grabbing with bare hands. Additionally, a DH grab interactable script is attached to each interactable object to recognize grab gestures directed toward it and to manage the corresponding functionalities.

When an object is selected, a "SELECTED" text message appears above it to provide feedback to the user (this can be turned off as needed). An object is deselected when the user looks away (in ET) or when the user moves the arm away (in DH). After selection, the user can grab the object by performing a grab gesture with either the left or right hand. The grab gesture is recognized using data from the OVR hand. Once the system detects a grab gesture, it calculates the distance and direction between the interactable object and the user's hand. The object's position is then updated from its previous location to align with the user's hand's grab point. A rigidbody component is attached to each interactable 3D object to enable physics-based movements. After a successful grab, the object moves from a distance to the user's hand automatically, simulating a natural grasping experience.

The CU grab tracking is implemented by attaching a hand grab interactor prefab to each hand interactor, for both the left and right hands. The CU grab interactor detects the hand grab posture and enables grabbing with bare hands. Additionally, a hand grab interactable script is attached to each interactable object to recognize grab gestures when an object is touched and grabbed with bare hands, allowing for the execution of the intended functionalities.

The VR environment is designed to support the study and comparison of different interaction methods. It can spawn three types of objects: cubes with varying colors, cubes of different sizes, and objects with diverse shapes. This setup enables a range of object interactions across multiple task types. Figure 4 illustrates one possible setup where cubes of different colors are created around a user.

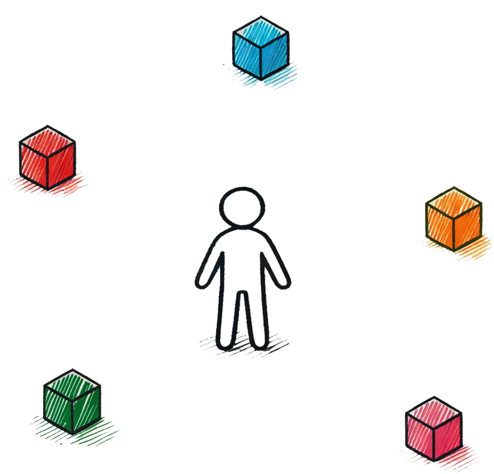


Figure 4: An illustration of one experiment setup where five cubes with different colors are randomly generated around the user. They can be at different angles and different heights.

5 User Study

5.1 Study Setup

For our study, we test three distinct object grabbing methods: ET object grabbing, DH grabbing, and CU grabbing. Each interaction method is evaluated using three tasks:

- The Color Task: select among five cubes of different colors (red, blue, orange, green, and pink) but identical sizes.
- The Size Task: select among five cubes of varying sizes, each labeled with a unique identifier (1 to 5).
- The Shape Task: select among five differently shaped objects (cube, rectangular prism, triangular prism, sphere, and star) with identical colors.

We use these tasks to provide participants a variety of visual stimuli. Each task includes five sub-tasks, where for each sub-task, users are instructed to select a specific object based on predefined criteria.

Figure 5 shows a few examples of the tasks from the user's first-person view. Figure 5 (a) and (b) show the Color Task where the user grabbed a red cube; (c) and (d) show the Size Task where the user grabbed a size-3 cube; (e) shows the Shape Task where the user grabbed a sphere; finally, (f) shows an instruction board that informs the user what their next task is.

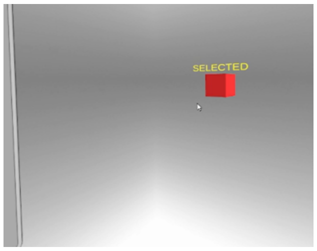
We utilized a laptop (8th Gen Intel(R) Core (TM) i5-8350U CPU @ 1.70GHz 1.90 GHz) and a desktop PC featuring a 12th Gen Intel(R) Core (TM) i7-12700H @ 2.10GHz with Nvdia GeForce GTX 1060 graphics. Meta Quest Pro was used for this study which was connected to the desktop PC. The laptop was used for collecting survey responses.

5.2 Study Procedure

The user study was conducted in a controlled, in-person lab environment, with identical equipment provided to each participant to ensure consistent testing conditions. To enhance the sense of immersion, participants were asked to stand in the center of a circular, object-free space. Only the experiment facilitator, located outside the VR boundary, was present to monitor and process the hand-tracking data. Participants were informed that questions during the study were limited to exceptional cases and that any uncertainties were clarified before starting the study.

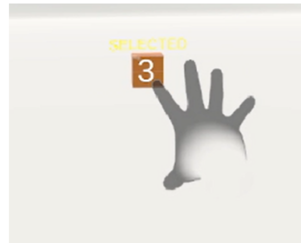
After a brief introduction to the study procedure and equipment, participants were immersed in the virtual environment. Participants were asked to complete all three tasks. The order of the tasks was randomized for each participant to minimize the learning effect from one task to another. A short break and additional briefing were provided prior to each new task. Participants were told to try to complete the tasks as fast as they could. Following the study, participants completed a questionnaire assessing their preferences and task load. The entire process, including preparation, the study, and the post-study questionnaire, took approximately 60 minutes per participant.

In all tasks of the study, a practice session preceded the main task to ensure participants felt comfortable and confident with each interaction method. No time limits were imposed on the practice sessions, allowing participants to fully understand the interactions, regardless of their prior familiarity with VR.



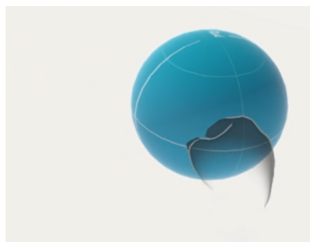
(a) Eye-tracked object selection (Color Task).

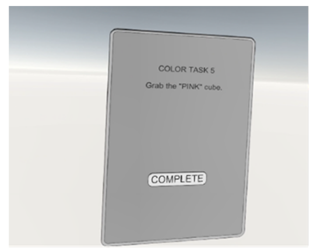

(b) Grabbing the eye-tracked object (Color Task).



(c) Distance hand object selection (Size Task).


(d) Grabbing the distance-hand selected object (Size Task).


(e) Close-up object grabbing (Shape Task).



(f) Instruction board.

Figure 5: Examples of the tasks involved in the user study from the user's view: (a) Selecting the red colored cube as per the instruction from five different colored cubes using eye tracking. (b) Grabbing the red cube after selecting it with eye. (c) Selecting the size-3 cube from five different sized cubes using distance hand selection. (d) Grabbing the selected cube with hand after selecting it. (e) Grabbing the sphere with close-up object grabbing from five different shapes. (f) Example of an instruction board in a color task.

The user study was reviewed and approved by the Research Ethics Board (REB) at the University of Calgary.

6 Results and Discussions

A total of 21 participants were recruited through the University of Calgary mailing lists of students. Each participant was assigned a unique ID to ensure the anonymity of the study data. The participants were between the ages of 22 and 49. 15 participants reported some previous experiences with VR. We asked participants to self-identify their gender; 10 self-identified as men and 11 self-identified as women, with no other gender reported. Participants reported their ethnic backgrounds as South Asian, East Asian, and Middle Eastern. Only one participant indicated "somewhat" sensitivity to VR sickness, while all others reported "no" sensitivity. All participants identified as right-handed and confirmed that they did not have color blindness.

6.1 Questionnaire Results

After completing the study, participants filled out a post-study questionnaire. In total, 19 of the 21 participants agreed that interacting with 3D models was easier using ET; one participant disagreed, and another responded with "maybe." Additionally, 19 participants believed that tasks took the least time to complete with the ET method, while two answered "maybe."

Participants rated their levels of discomfort, fatigue, and dizziness on a scale from 1 to 10, with 10 indicating the highest level of discomfort and 1 the lowest. For ET, 19 participants rated their experience as 1, while two rated it 2. Responses for the DH method varied between 1 and 4, and for the CU method, ratings ranged from 1 to 6. These results suggest that participants experienced higher levels of discomfort, fatigue, and dizziness with DH and CU interactions compared to ET.

Lastly, 12 participants reported feeling less confused when completing tasks with ET compared to the other two methods. Of the remaining participants, five disagreed, two responded "maybe," and two noted that they needed extra time to understand the interaction method.

6.2 Time

For the quantitative analysis, we analyze the time measures: Selection time refers to the time a user takes from the start of a task to making the correct selection in ET and DH interactions. Selection and Grab (Selection-Grab) time refers to the time a user takes from the start of a task to successfully select and grab the correct object (or in the case of CU grabbing, select and grab are considered the same action).

The comparative analysis between ET and DH selection times reveals that ET is significantly faster at selecting a 3D object from a distance compared to DH. Figure 6 shows the selection time comparisons for each of the three tasks using both interaction methods. ET outperforms DH, with results statistically significant at the 95% confidence level, as determined by two-tailed paired t-tests (Table 1). For instance, in the Color task, the average selection

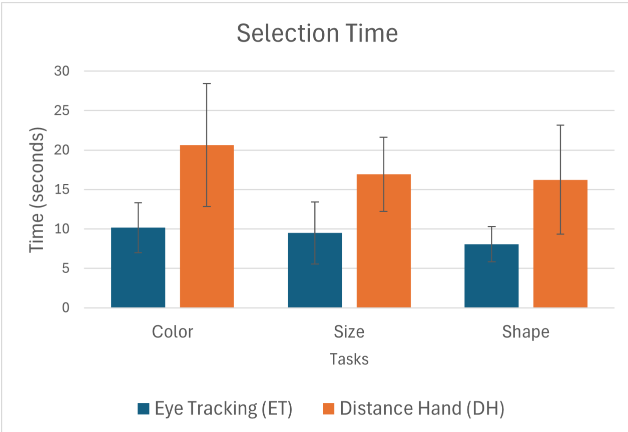


Figure 6: Selection time comparison between ET and DH in three different tasks. Error bars represent one standard deviation.

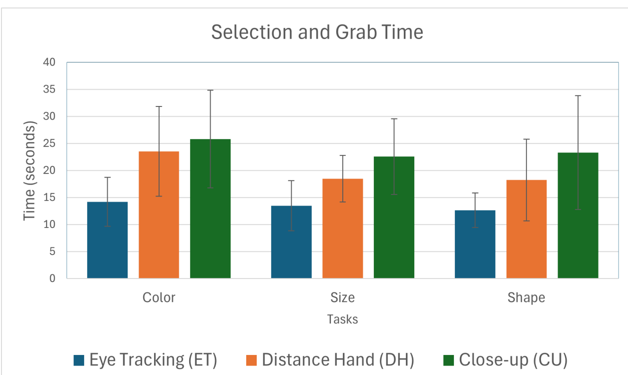


Figure 7: Selection-Grab time comparison between ET, DH, and CU grabbing in three different tasks. Error bars represent one standard deviation.

time with ET is 10.16 seconds, significantly lower than the 20.64 seconds required for DH.

We use ANOVA (Analysis of Variance) to assess whether there are statistically significant differences in the Selection-Grab times across the three interaction methods (ET, DH, and CU). By comparing the mean performance metrics for each interaction type, ANOVA helps determine if any observed differences in times or error rates are unlikely to have occurred by chance. When ANOVA indicates significant results, we conduct post-hoc t-tests to identify specific pairs of interactions that differ.

Results show that participants take less time to select and grab a 3D object using ET compared to both DH and CU interactions. Figure 7 shows the select and grab time comparisons for each task across the three interaction methods. ET outperforms DH and CU in all tasks, with results statistically significant at the 95% confidence level, as determined by two-tailed paired t-tests, with the Bonferroni correction used to account for the multiple t-tests (Table 2).

6.3 Accuracy

We measure accuracy across ET, DH, and CU interactions by calculating the total number of mistakes participants make in each task. Mistakes are defined as grabbing an object different from the one specified for that task. Table 3 shows the number of mistakes made by participants in each task. As expected, CU grabbing is the most accurate, as participants must walk up to the object, giving them extra time to confirm its correctness. Participants using ET and DH demonstrate similar accuracy.

Tasks	ET	DH	p-value ET and DH
Color	10.16 (3.18)	20.64 (7.78)	< 0.001
Size	9.48 (3.95)	16.92 (4.70)	< 0.001
Shape	8.07 (2.23)	16.24 (6.90)	< 0.001

Table 1: Selection time performance (mean and standard deviation) across all participants, measured in seconds, with the p-values of t-tests.

Tasks	ET	DH	CU	p-value ET and DH	p-value ET and CU
Color	14.20 (4.53)	23.52 (8.32)	25.78 (9.04)	< 0.001	< 0.001
Size	13.46 (4.66)	18.44 (4.31)	22.54 (7.00)	0.003	< 0.001
Shape	12.62 (3.20)	18.22 (7.55)	23.31 (10.53)	0.005	< 0.001

Table 2: Selection-Grab time performance (mean and standard deviation) across all participants, measured in seconds, with the p-values of t-tests.

6.4 Gender Analysis

We present the results of a comparative analysis between women and men participants interacting with ET, DH, and CU grabbing methods. A research question is whether there are any differences in performance when examining each gender separately using the eye-tracked method. Tables 4 and 5 show that ET consistently outperforms DH and CU in both Selection and the Selection-Grab times across all tasks, with statistical significance for each gender.

When comparing across gender, in ET tasks, women participants required less time than men participants for both Selection and Selection-Grab times. In contrast, man participants generally took less time than women participants for DH and CU tasks, except in the CU Color task. While none of these differences are statistically significant, this finding provides some indications that women participants were not disadvantaged by the ET interaction, and that eye tracking is equally accessible and user-friendly for both women

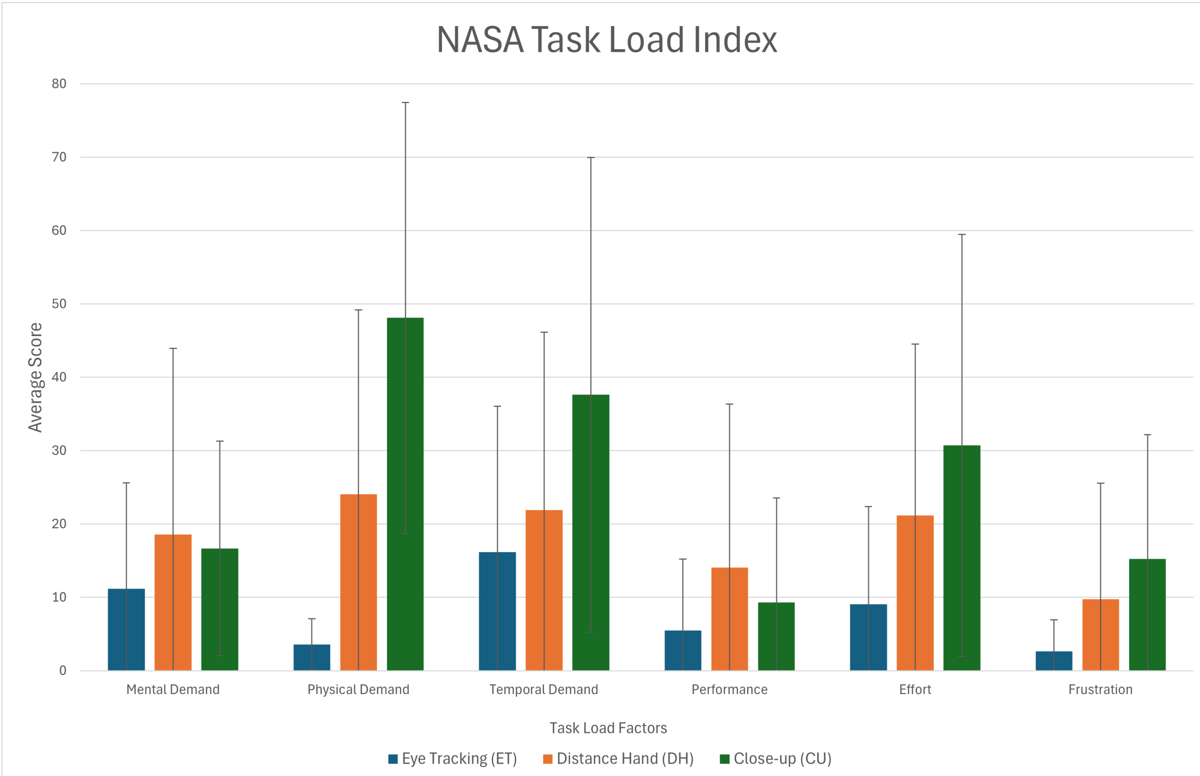


Figure 8: Task Load survey comparison between ET, DH, and CU grabbing. A lower score is more desirable in all factors. Error bars represent one standard deviation.

and men users, reducing entry barriers and enhancing inclusivity for women in using VR environments.

Mistakes	Interaction Type		
	ET	DH	CU
Color Task	2	4	0
Size Task	2	3	0
Shape Task	3	5	1

Table 3: Mistakes made by all participants for each task across the three grabbing methods.

Tasks	Men		Women	
	ET	DH	ET	DH
Color	10.46 (3.75)	19.88 (6.51)	9.88 (2.72)	21.33 (9.04)
Size	10.46 (3.77)	16.22 (3.35)	8.59 (4.07)	17.57 (5.75)
Shape	8.25 (2.08)	14.97 (4.69)	7.90 (2.46)	17.39 (8.50)

Table 4: Selection time performance (mean and standard deviation) for Men and Women using ET and DH, measured in seconds.

6.5 Task Load

We asked the participants to fill out a NASA Task Load Index (TLX) survey in the post-study questionnaire. The NASA TLX survey [13] is a subjective workload assessment tool that was developed to allow users to assess subjective workload assessments on various human-machine interface systems. The NASA TLX survey tracks six factors: mental demand, physical demand, temporal demand, performance, effort, and frustration.

We chose this survey to measure the feasibility of the eye-tracked interaction method as an alternative to hand tracking, in terms of how much load it puts on the users. A participant provided a rating for each of the six factors. The factor of Performance was rated on a scale of "perfect" to "failure" while the other five factors were rated on a scale of "low" to "high", representing scores from 0 to 100. Participants completed the TLX survey for each of the interactions they experienced and marked the task loads based on their own experience.

Figure 8 shows the average adjusted score for each of the six factors, based on user post-participation responses. We received a large range of responses as the standard deviations show. While ET received the least average score for the six factors, the significant results came in the factors of physical demand, effort, and frustration, indicating that participants experienced lower physical stress during the ET interactions while having less effort and frustration.

Tasks	Men			Women		
	ET	DH	CU	ET	DH	CU
Color	14.61 (5.14)	22.33 (7.80)	26.34 (9.89)	13.82 (4.11)	24.60 (9.00)	25.27 (8.64)
Size	14.48 (4.68)	17.70 (3.67)	22.08 (3.64)	12.53 (4.65)	19.12 (4.89)	22.97 (9.26)
Shape	13.40 (3.29)	17.01 (5.88)	21.35 (6.80)	11.91 (3.10)	19.32 (9.28)	25.08 (13.15)

Table 5: Selection-Grab time performance (mean and standard deviation) for Men and Women using the three grab methods, measured in seconds.

7 Conclusions, Limitations, and Future Work

In this paper, we conduct a study comparing three VR interaction methods to evaluate the In this paper, we conduct a study comparing three VR interaction methods to evaluate the effectiveness of combining eye tracking with finger movements for a multimodal method for object grabbing. Our results provide some evidence that the eye-tracked method demonstrates superior efficiency and user preference, reducing the time required for interactions with 3D models and providing a more favored alternative to hand-tracking-only interactions in VR for object grabbing. Additionally, our findings show that women perform as well as men when using the eye-tracked method, suggesting that it is equally accessible and effective across genders. These results indicate that eye tracking is an efficient interaction technique in VR, particularly valuable for applications in learning and gaming, where fast and accurate 3D interactions are essential.

Despite the promising outcomes, several limitations impact the generalizability of our results. First, our study sample was relatively small and drawn from a university population, which may not fully represent broader demographics or skill levels. Additionally, the limited range of tasks (selection and grabbing) and the relatively simple environments confine our conclusions to these specific interactions. Incorporating other common VR interactions (e.g. moving, scaling, rotating, and repositioning objects) would be a natural next step. More complex environments would certainly add difficulty to any of the interaction methods presented.

We plan to analyze additional user physical attributes to expand the scope of our findings. Future research could explore fully hands-free interactions by integrating eye tracking with speech recognition, providing a practical alternative for users with physical disabilities. Furthermore, eye-tracking selection could be extended to support more complex object manipulations, such as scaling and rotation. Studies on long-term usability, fatigue, and performance in collaborative VR environments would offer insights into the method's effectiveness across various settings. Testing with diverse populations beyond a university sample could also generalize findings. Broadening the application of eye tracking to other domains beyond object selection and distance grabbing could contribute to more accessible VR interfaces for educational and professional environments, fostering innovation in virtual learning and gaming applications.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant. We thank members of the Serious Games Research Group and the anonymous reviewers for their feedback. The Writefull tool within the Overleaf editor was used to improve the grammar in this paper.

References

[1] Isayas Berhe Adhanom, Paul MacNeilage, and Eelke Folmer. 2023. Eye tracking in virtual reality: a broad review of applications and challenges. *Virtual Reality* 27, 2 (2023), 1481–1505.

[2] Razvan Andrei, Oliver Benjamin Engermann, Christian Glerup Sørensen, and Markus Löchtefeld. 2023. Examining the effects of eye-tracking on dyadic conversations in virtual reality. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia*. 398–408.

[3] Farhan Aslam and Richard Zhao. 2024. Voice-Augmented Virtual Reality Interface for Serious Games. In *2024 IEEE Conference on Games (CoG)*. IEEE, 1–8.

[4] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*. 1–9.

[5] Simon Bothén, Jose Font, and Patrik Nilsson. 2018. An analysis and comparative user study on interactions in mobile virtual reality games. In *Proceedings of the 13th International Conference on the Foundations of Digital Games*. 1–8.

[6] Hubert Cecotti, Yogesh Kumar Meena, and Girijesh Prasad. 2018. A multimodal virtual keyboard using eye-tracking and hand gesture detection. In *2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 3330–3333.

[7] Yuping Chen and Meng-Jung Tsai. 2015. Eye-hand coordination strategies during active video game playing: An eye-tracking study. *Computers in Human Behavior* 51 (2015), 8–14.

[8] Mungyeong Choe, Yeongcheol Choi, Jaehyun Park, and Hyun K Kim. 2019. Comparison of gaze cursor input methods for virtual reality devices. *International Journal of Human–Computer Interaction* 35, 7 (2019), 620–629.

[9] Viviane Clay, Peter König, and Sabine U König. 2019. Eye tracking in virtual reality. *Journal of eye movement research* 12, 1 (2019).

[10] Frederic R Danion and J Randall Flanagan. 2018. Different gaze strategies during eye versus hand tracking of a moving target. *Scientific reports* 8, 1 (2018), 10059.

[11] Andrew T Duchowski, Vinay Shivashankaraiah, Tim Rawls, Anand K Gramopadhye, Brian J Melloy, and Barbara Kanki. 2000. Binocular eye tracking in virtual reality for inspection training. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 89–96.

[12] Asim Hameed, Andrew Perkis, and Sebastian Möller. 2021. Evaluating hand-tracking interaction for performing motor-tasks in vr learning environments. In *2021 13th international conference on quality of multimedia experience (QoMEX)*. IEEE, 219–224.

[13] SG Hart. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload/Elsevier* (1988).

[14] Wen-jun Hou and Xiao-lin Chen. 2021. Comparison of eye-based and controller-based selection in virtual reality. *International Journal of Human–Computer Interaction* 37, 5 (2021), 484–495.

[15] Cheryl I Johnson, Nicholas W Fraulini, Eric K Peterson, Jacob Entinger, and Daphne E Whitmer. 2023. Exploring Hand Tracking and Controller-Based Interactions in a VR Object Manipulation Task. In *International Conference on Human-Computer Interaction*. Springer, 64–81.

[16] Chaowanan Khundam, Varunyu Vorachart, Patibut Preeyawongsakul, Witthaya Hosap, and Frédéric Noël. 2021. A Comparative Study of Interaction Time and Usability of Using Controllers and Hand Tracking in Virtual Reality Training. Informatics 8, 3 (2021).

[17] Fan Li, Ching-Hung Lee, Shanshan Feng, Amy Trappey, and Fazal Gilani. 2021. Prospective on eye-tracking-based studies in immersive virtual reality. In *2021 IEEE 24th international conference on computer supported cooperative work in design (CSCWD)*. IEEE, 861–866.

[18] Tiffany Luong, Yi Fei Cheng, Max Möbus, Andreas Fender, and Christian Holz. 2023. Controllers or Bare Hands? A Controlled Evaluation of Input Techniques on Interaction Performance and Exertion in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[19] Francisco Lopez Luro and Veronica Sundstedt. 2019. A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In *Proceedings of the 11th ACM Symposium on eye tracking research & applications*. 1–9.

[20] Martin Meißner, Jella Pfeiffer, Thies Pfeiffer, and Harmen Oppewal. 2019. Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *Journal of Business Research* 100 (2019), 445–458.

[21] Maria Mikhailenko, Nadezhda Maksimenko, and Mikhail Kurushkin. 2022. Eye-tracking in immersive virtual reality for education: a review of the current progress and applications. In *Frontiers in Education*, Vol. 7. Frontiers Media SA, 697032.

[22] Pedro Monteiro, Guilherme Gonçalves, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2702–2713.

[23] Jesús Moreno-Arjonilla, Alfonso López-Ruiz, J Roberto Jiménez-Pérez, José E Callejas-Aguilera, and Juan M Jurado. 2024. Eye-tracking on virtual reality: a survey. *Virtual Reality* 28, 1 (2024), 38.

[24] Rasha Sameer Moustafa, Harri Karhu, Sami Andberg, and Roman Bednarik. 2023. Seeing Through Their Eyes-A Customizable Gaze-Contingent Simulation of Impaired Vision and Other Eye Conditions Using VR/XR Technology. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*. 1–2.

[25] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality. In *Acm symposium on eye tracking research and applications*. 1–7.

[26] Aunnoy K Mutasim, Wolfgang Stuerzlinger, and Anil Ufuk Batmaz. 2020. Gaze tracking for eye-hand coordination training systems in virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–9.

[27] Yun Suen Pai, Tilman Dingler, and Kai Kunze. 2019. Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality* 23 (2019), 119–131.

[28] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th symposium on spatial user interaction*. 99–108.

[29] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Computing Surveys (CSUR)* 55, 3 (2022), 1–39.

[30] Natasha Anne Rappa, Susan Ledger, Timothy Teo, Kok Wai Wong, Brad Power, and Bruce Hilliard. 2022. The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review. *Interactive Learning Environments* 30, 7 (2022), 1338–1350.

[31] Michael Shomoye and Richard Zhao. 2024. Automated emotion recognition of students in virtual reality classrooms. *Computers & Education: X Reality* 5 (2024), 100082.

[32] Ludwig Sidenmark and Anders Lundström. 2019. Gaze behaviour on interacted objects during hand interaction in virtual reality for eye tracking calibration. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–9.

[33] Daniel Sonntag, Jason Orlosky, Markus Weber, Yecheng Gu, Sergey Sosnovsky, Takumi Toyama, and Ehsan Nadjaran Toosi. 2015. Cognitive monitoring via eye tracking in virtual reality pedestrian environments. In *Proceedings of the 4th international symposium on pervasive displays*. 269–270.

[34] Christoph Tremmel, Christian Herff, Tetsuya Sato, Krzysztof Rechowicz, Yusuke Yamani, and Dean J Krusienski. 2019. Estimating cognitive workload in an interactive virtual reality environment using EEG. *Frontiers in human neuroscience* 13 (2019), 401.

[35] Roel Vertegaal. 2008. A Fitts Law comparison of eye tracking and manual input in the selection of visual targets. In *Proceedings of the 10th international conference on Multimodal interfaces*. 241–248.

[36] Jan-Niklas Voigt-Antons, Tanja Kojic, Danish Ali, and Sebastian Möller. 2020. Influence of hand tracking as a way of interaction in virtual reality on user experience. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–4.

[37] Shu Wei, Desmond Bloemers, and Aitor Rovira. 2023. A preliminary study of the eye tracker in the meta quest pro. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*. 216–221.