Design and Evaluation of User-Centric Gesture-based Selection Techniques for Small Objects of Varying Distances in Dense Virtual Environments of HMD-VR Applications

Thesis
submitted in fulfilment of the
requirements of the degree of

DOCTOR OF PHILOSOPHY

by

Shimmila Bhowmick

ROLL NO. 166105005

Supervisor

Dr. Pratul Chandra Kalita

Department of Design
Indian Institute of Technology
Guwahati, Guwahati 781039, India

January 2024

CERTIFICATE

This is to certify that the thesis entitled "**Design and Evaluation of User-Centric Gesture-based Selection Techniques for Small Objects of Varying Distances in Dense Virtual Environments of HMD-VR Applications**" submitted by Shimmila Bhowmick to the Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy in Design is a record of bonafide research work carried out by her under my supervision and guidance. The thesis work, in my opinion, has reached the requisite standard fulfilling the requirement for the degree of Doctor of Philosophy. The results contained in this thesis have not been submitted in part or full to any other University or Institute for award of any degree or diploma.

Dr. Pratul Chandra Kalita
Professor, Department of Design
Indian Institute of Technology Guwahati
Guwahati-781039

Guwahati
Date:

STATEMENT

I do hereby declare that the matter embodied in this thesis is the result of investigations carried out by me in the Department of Design, Indian Institute of Technology Guwahati, Guwahati, Assam, India.

In keeping with the general practice of reporting scientific observations, due acknowledgements have been made wherever the work described is based on the findings of other investigators.

(Shimmila Bhowmick)

Guwahati,

Date:

Abstract

The global virtual reality market was worth USD 21.83 billion in 2021 and is expected to grow at a compound annual growth rate of 15.0% from 2022 to 2030 (Virtual Reality Market Size and Share Report, 2022). Virtual reality (VR) is a computer-generated simulation of a three-dimensional environment that can be interacted with in a seemingly real or physical way through the use of electronic devices, such as head-mounted displays. In VR, users have various options for interacting with the virtual world, such as controllers, gloves, bare-hand gestures, haptic devices, and bodysuits. VR has a multitude of applications, ranging from gaming to healthcare, education and training, real estate and architecture, and tourism and travels to name a few. These VR applications provide the ability to select, manipulate (such as move, rotate, scale, and transform) virtual objects, navigate through the virtual environment, and perform application-specific tasks (Argelaguet and Andujar, 2013). Object selection is an essential and crucial task in any of the VR applications because it allows users to identify, interact with and manipulate virtual objects in the virtual environment (VE). In VR, VEs contain objects of varying densities, sizes, and distances from the user. For example, VR applications designed for drug discovery molecular visualization enable scientists to physically interact with small-sized molecules, even those as small as nail-size, that are located at varying distances within a dense VE. However, selecting small objects placed at varying distances in a dense VE can be challenging, and can lead to inaccurate selection, increased task completion times, and higher levels of fatigue and frustration. Therefore, it is necessary to develop effective object selection techniques that enable accurate and quick selection, particularly for small objects placed at varying distances within dense VE.

Several selection techniques have been developed to facilitate accurate and quick object selection. However, as far as we are aware, there has been insufficient research aimed at developing an object selection technique capable of accurately and quickly selecting small objects located at varying distances within a dense virtual environment. Therefore, there is a need for a selection technique that is both precise and efficient, and capable of selecting small objects located at different distances within a dense virtual environment without introducing errors. The objective of this study is to enhance the current techniques for selecting objects in dense virtual environments by introducing an accurate, error-free, and efficient method that is specifically designed to select

small objects placed at varying distances in these environments. Our thesis is focused on controller-less object selection techniques. The rationale behind using a controller-less selection technique is that it can reduce the complexity of virtual reality setups by eliminating the need for extra hardware, such as controllers. This approach also improves accessibility and user comfort since users do not have to hold or manipulate additional equipment. Additionally, a controller-less technique enables the use of natural and intuitive selection methods, such as body gestures that allow users to interact with the VR environment using their own hands and body movements, thereby increasing the chances to enhance their sense of presence in the virtual world.

The thesis is based on five primary experiments, with the first two studies being gesture elicitation studies. The initial study focuses on investigating user-centric gestures for selecting small objects within arm's reach and at a distance in a dense VE. The second study concentrates on evaluating user-elicitated gestures for selecting an individual gesture for each VE condition. The gestures were evaluated based on their ease of use, appropriateness, suitability for the gesture's function, preferences, and effort. Thereafter three studies were conducted to evaluate the techniques with existing techniques in the literature. The third study aimed to design and evaluate gesture-based object selection techniques for dense VE where targets are small and placed within arm's reach. We designed a technique called Locked Dwell Time based Point and Tap (LDTPT) and compared it with existing techniques from the literature. The results showed that LDTPT was the most natural and efficient technique for selecting small targets at arm's reach. However, participants found it difficult to use and learn. To address the limitations, in the fourth study, we designed and evaluated a new technique called Tiny hands, which scales down the size of virtual hands to select small targets. Results showed that Tiny hands technique was significantly faster, more accurate, and easy to use, learn, and preferred over other techniques. In the fifth study, we designed and evaluated AMAZE and AMAZE-X techniques (A Multi-finger Approach to Zoom in dense Environments) for the selection of small targets placed at a distance in dense VE. This technique offers zoom using multiple fingers in VR. Results show that these techniques outperform existing techniques in task completion time, accuracy, easy to use, ease of learning, naturalness, preference and effort. Lastly, we are able to present a set of design recommendations that can be used by designers and developers to design efficient and effective gesture-based object selection techniques for small object selection in dense VE where targets are placed within arm's reach and at a distance.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Professor Pratul Chandra Kalita for his guidance, encouragement, and support during the course of my thesis work. His exceptional technical acumen, and his penetrating intellect have inspired me to strive for excellence in my own work and to constantly push the boundaries of my knowledge and skills. He has consistently emphasized the importance of maintaining a healthy work-life balance, encouraging me to prioritize both my personal and professional pursuits ensuring a well-rounded and contented existence. I have been fortunate enough to have him as my professional guide.

Special acknowledgement is due to Professor(s) Saugata Karmakar, Debayan Dhar, and Chandan Karfa for serving on my supervisory committee and providing their valuable inputs that were helpful in improving the thesis. I truly appreciate all of their time and assistance as I navigated through this process.

My deepest appreciation goes to my wonderful husband, Keyur. I cannot thank him enough for his unending patience and understanding throughout this journey and for all the sacrifices he made to enable me to complete this thesis. His willingness to take on additional responsibilities, to make sure I had the time and space to focus on my research, has been invaluable. I am grateful for his constant support and encouragement, even during the most challenging times. I also want to thank him for his willingness to read my thesis draft. His critical feedback and insightful comments have helped me to improve the quality of my work and to present my research in the best possible way. His attention to detail and meticulous proofreading has been invaluable in ensuring my thesis is polished. I appreciate the time and effort he has invested in this process. I thank him for being an exceptional partner and for going above and beyond to help me achieve my goals.

And to my little bundle of joy, my daughter Vyana: this Ph.D. work became her twin sister for the first three years of her life, and I would not have had it any other way. Together, we went through a lot to make this research possible. There were times when I felt overwhelmed and stretched to my limit juggling deadlines, research, and family responsibilities. However, through it all, she was my constant source of motivation and inspiration. Her unwavering love, energy, and enthusiasm has helped me to stay focused and determined, even during the most difficult moments.

I am grateful and thankful to God for her presence in my life, and I know that this experience has made our bond stronger than ever. I am proud to have been able to complete my Ph.D. while also being a dedicated parent. In many ways, motherhood enriched my Ph.D. journey, and I am grateful to have experienced both simultaneously. I am grateful for the child she is and for being the best companion, I would have ever asked for.

Last but certainly not least, I thank my mom and dad for their unconditional love and support. They have instilled in me the values of hard work, determination, and perseverance, which have been essential in this journey. My brothers for always standing by my side, offering words of encouragement and support throughout the process. I thank my parents-in-law, who supported me in countless ways, for which I am eternally grateful. My family has been the bedrock that has facilitated the growth of my career, providing me with much-required support to balance my professional and personal life.

I thank God for this journey and for giving me an expected end.

Shimmila Bhowmick

Table of Contents

List of Figures

shaped closed surface for selection in dense VEs Wonner et al. 2012 (f) Expand Technique uses zoom and SQUAD to select objects.

Figure 5: Controller-based object selection techniques that use heuristics and scoring-based functions (a) IntenSelect employs a score of objects within a conic selection volume (Haan el al., 2005) (b) PRISM technique (Precise and Rapid Interaction through Scaled Manipulation) presented by Frees and Kessler, 2005. (c) The scope technique developed by Cashion and LaViola, 2014 alters the activation area in relation to the velocity of the handheld controller and the environment density.

Figure 6: Virtual Hand extension (Mine et al. 1997) developed as an alternative to controller-based raycasting (b) The Go-Go technique (Poupyrev et al. 1996) attempts to improve on the simple virtual hand by providing users to interactively change the length of the virtual arm. (c) Eye-rooted and hand-rooted techniques (Argelaguet et al. 2008) (d) Multiray (Matulic and Vogel 2018) uses multi-finger raycasting to independently emit unmodified rays from a distance on large-scale displays (e) PRECIOUS, Mendes et al., 2017, uses cone-casting for selecting out-of-reach objects featuring iterative refinement in VR interfaces.

Figure 7: Controller-based object selection using body gestures (a) The Head-Crusher technique (Pierce et al. 1997) in which the object is between the user's finger and thumb in its image plane. (b) The rapMenu technique (Ni et al. 2011 allow menu selection by controlling wrist tilt, point and pull gestures (c)AirTap and Thumb Trigger (Vogel and Balakrishnan 2005) performs finger clicks and thumb trigger taps on the side of the index finger providing kinesthetic feedback.

Figure 8: Expanded view of the designed VEs – VE1 and VE 2 and the targets (in red color).

Figure 9: The study setup used for the elicitation study. The moderator observes the participant while the cameras placed diagonally to the participant record the gestures for further analysis.

Figure 10: Visual representation of the extracted gestures for VE1: (a) Point and tap (b) Grab and rotate (c) Point & bring hand closer

Figure 11: Visual representation of the extracted gestures for VE2: (a) Pinch-out VE and point, and tap (b) Extending a hand and pull hand closer to the body.

Figure 12: The mean and SD values and error bars for ease of performing, body-part suitability, and gesture appropriateness for finalized gestures of VE1.

Figure 13: The mean and SD values and error bars for ease of performing, body-part suitability, and gesture appropriateness for finalized gestures of VE2.

Figure 14: Selection of small object using *LDTPT* technique. (a) Selection of small objects using one finger pad area. (b) The target object changes from red to yellow. (c) The target object changes its color from yellow to green and is in a locked phase (d) User taps anywhere in the environment to confirm the selection. The target object changes its colour from green to blue.

Figure 15: Experimental setup used for the study. The user is wearing an Oculus Rift headset with a mounted Leap motion device. A camera was placed diagonal to the user took the photograph.

Figure 16: The mean and SD values for task completion time for techniques *LDTPT*, Magnetic grasp, Pinch, and Raycasting, respectively.

Figure 17: The mean and SD values for Error rate for techniques *LDTPT*, magnetic grasp, pinch and raycasting, respectively.

Figure 18: The means and SD values for ease of use, easy to learn, and perceived naturalness for techniques *LDTPT*, magnetic grasp, pinch and raycasting respectively.

Figure 19: Interaction technique 1: Fisting and rotating non-dominant hand (a) User brings the virtual hands in front of the user (b) the user makes a fist gesture on the non-dominant hand to trigger *tiny hands*. By rotating the fist clockwise/anticlockwise, the user selects the hand size.

Figure 20: Interaction technique 2: Holding the Palm downwards with non-dominant hand and performing vertical movement (a) Virtual hands are activated, user places non-dominant hand in the camera view static for 2 secs to trigger *tiny hands* (b) Vertical movement of the non-dominant hand up/down increases/decreases the virtual right-hand size.

Figure 21: Interaction Technique 3: Pinching-in/out the non-dominant hand (a) Pinch in/out using index finger and thumb to decrease/increase the hand size. (b) Pinch-in using index finger and thumb to decrease the hand size.

Figure 22: (a) A snapshot of the virtual environment and the virtual hands created for the study. (b) A comparison of the virtual hands size with the size of targets.

Figure 23: (a) User places virtual hands in front of the user for 2 seconds to trigger *tiny hands* (b) Vertical movement of the non-dominant hand downwards reduces the dominant hand size. (d) *Tiny hand* size is finalized (d) *Tiny hands* precisely select the target.

Figure 24: The average task completion time in seconds for the three techniques: *tiny hands*, ray casting, and pinch-to-select

Figure 25: Error rate for the three techniques: *tiny hands*, raycasting, and pinch. *Tiny hands* made zero errors.

Figure 26: The mean and standard deviation for easy to use, ease of learning, naturalness, and playfulness for the three techniques: *tiny hands*, ray casting and pinch-to-select

Figure 27: Selection using AMAZE technique: (a) User initiates the gestures (b) User performs pinch-out with five fingers which zoom the target at 5x zoom. This step is done multiple times to bring the target at a near distance from the user. (c) The user performs one-finger zoom (1x) to granularly place the target near the user. (d) User points using the right-hand index finger to select and confirm the target.

Figure 28: (a) The effective region (white in colour) is designed to initiate the gesture and to avoid unintentional gestures. The virtual hand colour is black when hands are inside the effective region. (b) The virtual hand colour changes to green when the user goes out of the effective region.

Figure 29: Text-based representation of zoom-in function (a) one finger zoom-in is visually presented to 1x (b) all finger zoom-in is indicated using 5x.

Figure 30: The effective region (10 cm in radius) is an unobstructed space created to perform the pinch-in/out gestures, and two operating regions are designed to control the speed of the zoom-in. Operating region 1 (20 cm) zooms at a rate of 1 m/s and operating region 2 (30 cm) zooms in at a speed of 5m/s.

Figure 31: The experimental setup used for the study. For the AMAZE technique, the user performs the pinch-out gesture to zoom in on the VE with a non-dominant hand.

Figure 32: The mean and SD values for task completion time (in seconds) for the techniques AMAZE, Expand, and Pinch-to-select

Figure 33: The mean and SD values for error rate for techniques AMAZE, Expand and Pinch-to-select

Figure 34: The mean and SD values for ease of use, ease of learning, and naturalness for techniques AMAZE, Expand and Pinch-to-Select

Figure 35: An additional visual element (in white) during the corrective phase indicates the distance to the user in the design of AMAZE-X.

Figure 36: Mean and SD values for task completion time for AMAZE-X, AMAZE, Expand, and Pinch. AMAZE-X was the fastest, followed by Pinch, AMAZE, and Expand.

Figure 37: The mean and SD values for error rate for AMAZE-X, AMAZE, Expand, and Pinch-to-select techniques. The participants performed no errors for AMAZE-X, AMAZE, and Expand.

Figure 38: Mean and SD values for ease of use and learning learn for AMAZE-X, AMAZE, Expand, and Pinch

List of Tables

Chapter 1

1. Introduction

1.1 Context and Motivation

The global virtual reality market was worth USD 21.83 billion in 2021 and is expected to grow at a compound annual growth rate of 15.0% from 2022 to 2030 (Virtual Reality Market Size and Share Report, 2022). Virtual reality (VR) is a computer-generated simulation of a three-dimensional environment that can be interacted with in a seemingly real or physical way through the use of electronic devices, such as head-mounted displays. The aim of virtual reality is to provide users with a fully immersive and interactive experience, allowing users to feel as if they are actually present in a digitally created environment (Sutherland, 1965). VR enables users to experience a simulated world that approximates reality in ways that would not be possible in the real world. In VR, users have various options for interacting with the virtual world, such as controllers, gloves, bare-hand gestures, haptic devices, and bodysuits.

VR has a multitude of applications, ranging from gaming to healthcare, education and training, real estate and architecture, and tourism and travels to name a few. In the gaming industry, VR allows for a more engaging and realistic gaming experience, as players can interact with the game environment using VR headsets and controllers. In education and training, VR is used to create simulated environments for practical learning and skill development in a safe and controlled setting. In healthcare, VR is employed for pain management, rehabilitation, and surgical simulation. For instance, VR can help patients undergoing medical procedures by creating a calming environment or assist in training medical professionals by simulating surgical procedures. VR is also used in real estate and architecture to create virtual tours of properties and help designers and architects visualize and refine their designs. The tourism and travel industry leverages VR to offer immersive travel experiences, enabling tourists to explore destinations virtually, make informed travel decisions, and gain a more realistic understanding of what they can expect from a trip. Overall, as technology continues to evolve, the potential of VR will be seen in many industries for various use cases.

VR applications provide the ability to select, manipulate (such as move, rotate, scale, and transform) virtual objects, navigate through the virtual environment, and perform application-specific tasks (Argelaguet and Andujar, 2013). Object selection is an essential aspect of VR applications because it allows users to interact with and manipulate virtual objects in the virtual environment (VE). Object selection in VR involves identifying and choosing a virtual object within a 3D environment. It is a crucial task in VR because it facilitates manipulation, locomotion, and application context. In VR, VEs contain objects of varying densities, sizes, and distances from the user. For example, VR applications designed for drug discovery molecular visualization enable scientists to interact with and analyze small-molecule compounds and protein structures that drive disease. In order to conduct scientific analysis and relational mapping of molecules, scientists require the ability to physically interact with small-sized molecules, even those as small as nail-sized, that are located at varying distances within a dense VE. However, selecting small objects placed at varying distances in a dense VE can be challenging, and can lead to inaccurate selection, increased task completion times, and higher levels of fatigue and frustration. Therefore, it is necessary to develop effective object selection techniques that enable accurate and quick selection, particularly for small objects placed at varying distances within dense VE.

Several selection techniques have been developed to facilitate accurate and quick object selection, such as the ray-casting technique. It is efficient to select moderately sized nearby objects. However, the performance of ray-casting selection suffers a significant decline due to the need for high angular precision, as well as the magnification of hand and tracker vibrations at longer distances (Poupyrev et al., 1998). Other selection methods, such as Flashlight, SQUAD, and Expand, employ cone casting (Liang and Green, 2000; Kopper et al., 2011; Cahion et al., 2011). These techniques use a progressive refinement approach to achieve high accuracy, but they can be time-consuming and cause fatigue. Additionally, they are not effective when targets are small or in highly dense VEs. Researchers have developed gesture-based object selection methods that rely on finger movements (Vogel and Balakrishnan, 2002) or pulling the hand back while pointing (Ren and O'Neill, 2012) for selecting objects at a distance. Hand-based selection techniques (Mendes et al., 2017) for distant objects have also been introduced, but these methods have lower accuracy and can cause arm fatigue, especially when selecting small objects in crowded virtual environments. As far as we are aware, there has been insufficient research aimed at developing an object selection technique capable of accurately and quickly selecting small objects located at varying distances within a dense virtual environment. Therefore, there is a need for a selection technique that is both precise, efficient, and capable of selecting small objects located at different distances within a dense virtual environment without introducing errors.

The objective of this study is to enhance the current techniques for selecting objects in dense virtual environments by introducing an accurate, error-free, and efficient method that is specifically designed to select small objects placed at varying distances in these environments. Our thesis is focused on controller-less object selection techniques. The rationale behind using a controller-less selection technique is that it can reduce the cost and complexity of virtual reality setups by eliminating the need for extra hardware, such as controllers. This approach also improves accessibility and user comfort since users do not have to hold or manipulate additional equipment. Additionally, a controller-less technique enables the use of natural and intuitive selection methods, such as body gestures that allow users to interact with the VR environment using their own hands and body movements, thereby increasing the chances to enhance their sense of presence in the virtual world.

The thesis is based on five primary experiments, with the first two studies being gesture elicitation studies. The initial study focuses on investigating user-centric gestures for selecting small objects situated within arm's reach and at a distance in a dense virtual environment. In contrast, the second study concentrates on elucidating user-centric gestures for selecting small objects located beyond arm's reach in a dense VE. The results of these studies yield an individual gesture for each VE condition by evaluating their ease of use, appropriateness, suitability for the gesture's function, preferences, and effort. Thereafter three studies were conducted to evaluate the techniques with existing techniques in the literature. The third study aimed to design and evaluate gesture-based object selection techniques for dense VE where targets are small and placed within arm's reach. We designed a technique called Locked Dwell Time based Point and Tap (LDTPT) and compared it with existing techniques from the literature. The results showed that LDTPT was the most natural and efficient technique for selecting small targets at arm's reach. However, participants found it difficult to use and learn. To address the limitations, in the fourth study, we designed and evaluated a new technique called Tiny hands, which reduces the size of virtual hands to select small targets. Results showed that Tiny hands was significantly faster, more accurate, and easy to use, learn, and preferred over other techniques. In the fifth study, we designed and evaluated the AMAZE technique (A Multi-finger Approach to Zoom in Dense Environments) for selection of small targets placed at distance in dense VE. This technique offers zoom using multiple fingers in VR. The results showed that AMAZE was accurate and natural but took more selection time than other techniques. We redesigned AMAZE to improve its selection time and presented AMAZE-X, which improved the zoom factor and added subtle dash lines as visual elements during the corrective phase. A second comparative study was conducted, and results showed that AMAZE-X was significantly faster, easier to perform, and easier to learn than other techniques and was the most preferred technique. Results show that these techniques outperform existing techniques in terms of task completion time, accuracy, easy to use, easy to learn, naturalness, preference and effort. We are able to present a set of design recommendations that can be used by designers and developers to design efficient and effective gesture-based object selection techniques for small object selection in dense VE where targets are placed within arm's length and at a distance.

1.2 Research Questions

RQ1: What are the user-centric controller-less gestures for selecting small objects in a dense VE?

RQ 1.1: What are the user-centric gestures for the selection of small objects within arms' reach in a dense VE?

RQ 1.2: What are the user-centric gestures for the selection of small objects at a scaled lenth in dense VE?

RQ2: How do the proposed controller-less gestures impact accuracy, naturalness, intuitiveness, error rates, and task completion times to existing techniques?

RQ 2.1: Which of the designed controller-less gestures are natural, easy to use, easy to learn, and efficient for small objects selection within arms' reach in a dense VE?

RQ 2.2: Which of the designed controller-less gestures are natural, easy to use, easy to learn, and efficient for small object selection at scaled length a dense VE?

1.3 Contributions

Following are the specific contributions of this thesis.

1. User-centric gestures for controller-less selection of small object placed at arm's reach in a dense virtual environment. We elicitated 196 gestures (considering the frequency of the gestures) and 23 unique gestures (after categorization) for the specific VE where targets are small and kept within arms' reach in dense VE. (Chapter 3)

2. User-centric gestures for controller-less selection of small object placed at a scaled length in the dense virtual environment. We elicitated 194 gestures (considering the frequency of the gestures) and 29 unique gestures (after categorization) for the specific VE where targets are small and at a distance in dense VE. (Chapter 3)

3. Categorization and taxonomy of a set of natural and intuitive whole-body gestures for small targets placed at arms' reach and distant position in a dense virtual environment. We proposed the gesture taxonomies from 52 (23+29) unique gestures extracted from study 1. Gestures were classified based on hand dominance in performing the task and the motion performed by hands. We also observed the posture and the sequence of gestures to propose the taxonomies. We introduce two taxonomies: hand-dominance and multiple body-part movement gestures, each with sub-categories (dominant hand only, non-dominant hand first, equal hand dominance). (Chapter 3)

4. Design and evaluation of Locked Dwell Time-based Point and Tap (LDTPT) and Tiny hands, two novel gesture-based selection techniques for small objects in arm's reach in a dense virtual environment. We improved upon the user-centric gesture identified for VE1 (Study 1 and Study 2) and designed the two techniques to accurately select small objects in dense VE where objects are placed within arms' reach. We were able to establish that these two techniques proved efficient in terms of task completion time and error rate. LDTPT and Tiny hands technique was also natural, easy to use and easy to learn, most preferred and low in effort in the selection of small targets in dense VE where objects are within arms' reach. (Chapter 4 and Chapter 5).

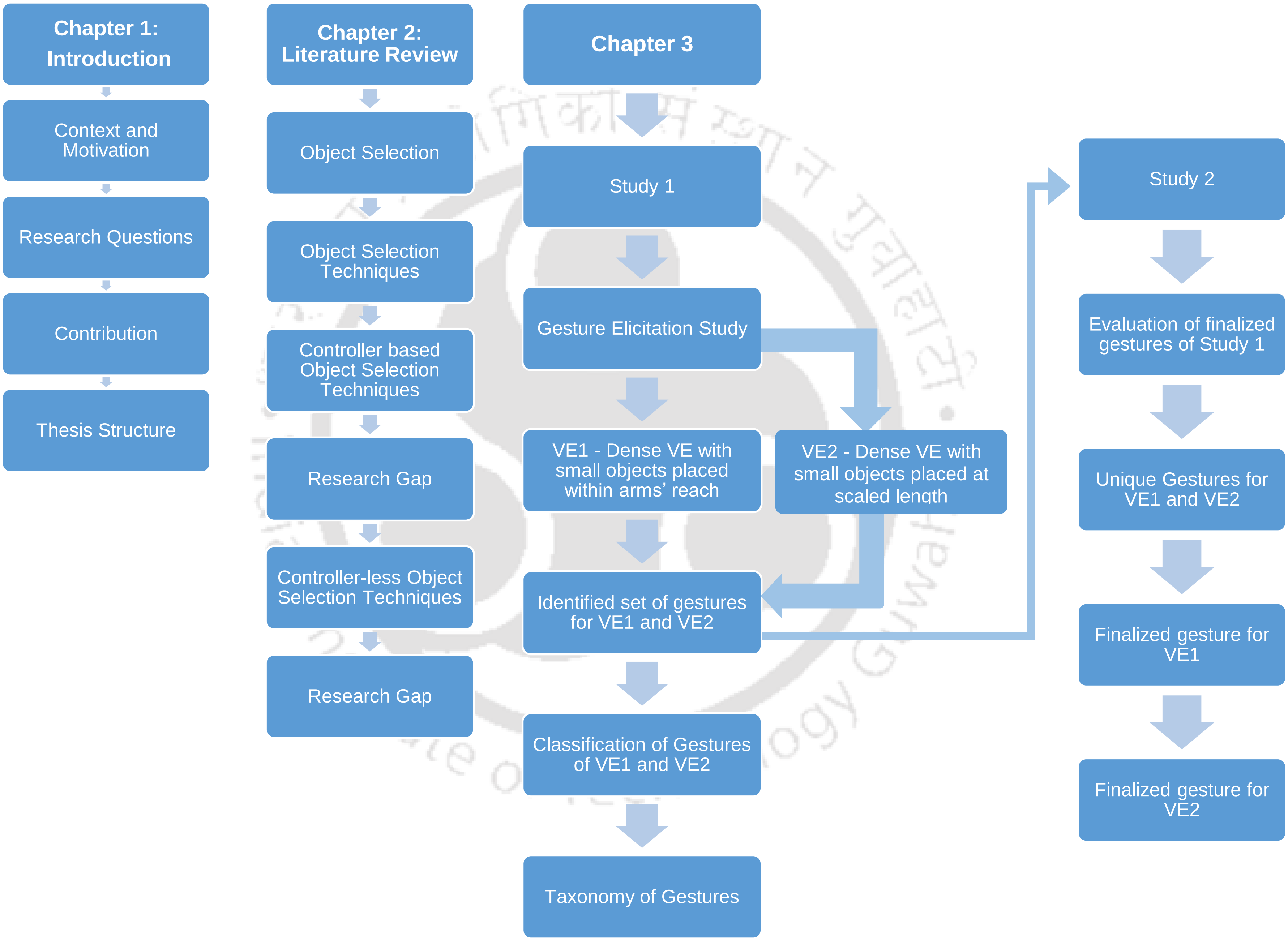
5. Design and evaluation of AMAZE and AMAZE-X, a novel gesture-based object selection technique for small, distant object selection in a dense virtual environment. We designed and evaluated AMAZE for the accurate selection of small distant targets. We established that AMAZE and AMAZE-X proved to be better in terms of task completion time and error rate. It was also easy to use, easy to learn, natural and the most preferred among the techniques. (Chapter 6).
6. Recommendations and design guidelines for designers and developers to design efficient gesture-based selection techniques for small object selection in a dense virtual environments for targets placed within arm's reach and at a distance. Depending on the context and different application scenarios designers can consider using different techniques. For example, considering a medical surgery context (objects are small placed in a dense VE within arms' reach) where accuracy is of utmost importance, the tiny hands gesture could be used as the user is able to navigate with the small size of the hand and accurately select a target. (Chapter 6).

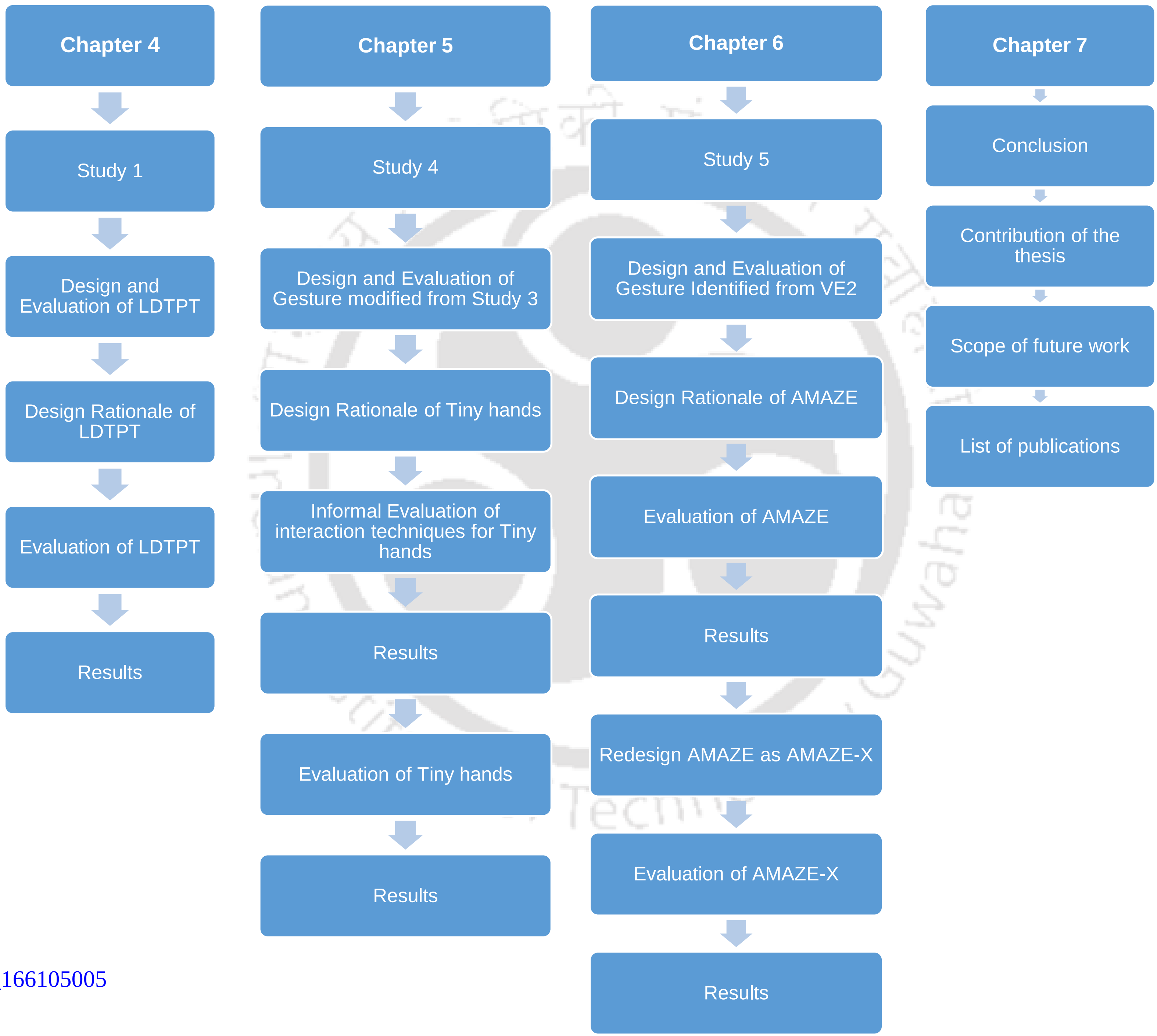
1.4 Thesis Structure

The thesis is divided into seven chapters:

Chapter 1: Introduction presents the context and motivation of the thesis. It starts by describing the global market of VR, the use of VR in a multitude of applications, ranging from gaming to healthcare, education and training, real estate and architecture, tourism and travel. This is followed by providing the importance of object selection and object selection techniques, limitations and the significance of investigating controller-based object selection techniques (section 1.1). The research questions are presented in section 1.2. Further, it presents the contributions of the research work to the existing body of knowledge in section 1.3. At last, section 1.4 details the structure of this thesis.

Chapter 2: **A Review and Analysis of Existing Research** details the focus areas of the literature review. It first provides an overview of object selection in VR (section 2.1) and object selection techniques (section 2.2). Section 2.3 provides an extensive review of existing literature on controller-based object selection methods for small, dense targets placed at varying distances.


Chapter 1: Introduction
Context and Motivation
Research Questions
Contribution
Thesis Structure
Chapter 2: Literature Review
Object Selection
Object Selection Techniques
Controller based Object Selection Techniques
Research Gap
Controller-less Object Selection Techniques
Research Gap
Chapter 3
Study 1
Gesture Elicitation Study
VE1 - Dense VE with small objects placed within arms' reach
VE2 - Dense VE with small objects placed at scaled length
Identified set of gestures for VE1 and VE2
Classification of Gestures of VE1 and VE2
Taxonomy of Gestures
Study 2
Evaluation of finalized gestures of Study 1
Unique Gestures for VE1 and VE2
Finalized gesture for VE1
Finalized gesture for VE2


Chapter 4
Study 1
Design and Evaluation of LDTPT
Design Rationale of LDTPT
Evaluation of LDTPT
Results
Chapter 5
Study 4
Design and Evaluation of Gesture modified from Study 3
Design Rationale of Tiny hands
Informal Evaluation of interaction techniques for Tiny hands
Results
Evaluation of Tiny hands
Results
Chapter 6
Study 5
Design and Evaluation of Gesture Identified from VE2
Design Rationale of AMAZE
Evaluation of AMAZE
Results
Redesign AMAZE as AMAZE-X
Evaluation of AMAZE-X
Results
Chapter 7
Conclusion
Contribution of the thesis
Scope of future work
List of publications

The limitations and challenges of controller-based object selection techniques, specifically for selecting small objects in arm's reach and distant objects in dense VEs is reported in Section 2.3.1. It further discusses the need for new approaches for object selection of small targets in dense VE for targets placed at varying distances. Section 2.4 provides an overview of the existing literature on controller-less object selection techniques. It elaborated on the various object selection techniques using body gestures for small targets in dense VE for targets placed at varying distances. The challenges and the research gap of controller-less object selection techniques are mentioned in section 2.4.1. The chapter is summarized in section 2.5.

Chapter 3: User-Generated Gesture Elicitation Study presents the methodology for identifying suitable body gestures for small object selection in dense HMD-based VR, taking into account varying distances. A comprehensive review of elicitation methodologies for gesture design is discussed in section 3.1. Section 3.2, provides the methodology for our first study, aimed at generating natural and intuitive object selection gestures for small object selection in two VEs (i) *Dense VE where targets are within arms' reach* and (ii) *Dense VE with targets at a scaled distance*. The objective of the study, VE design, participants' details, study set-up and apparatus, and study procedure are presented in sections 3.2.1 to 3.2.5, respectively. The results and gestures collected during the study are reported in section 3.2.5. The final gesture for VE 1 is *point and tap* gesture and the final gesture for VE 2 is *pinch in/out to zoom in/out the VE and point and tap* gesture. Section 3.2.8 provides a discussion on the findings of VE1 and VE2. Section 3.3 presents the evaluation of the subjective ratings of gestures for VE1 and VE2. The details of the participants' information and methodology, as presented in section 3.3. The results of the study are presented in section 3.4, followed by results on each of the environment VE1 and VE2 which is presented in 3.3.4.1 and 3.3.4.2. Section 3.3.5 discussion of Results of Study 2.

Section 3.4 presents a classification of gestures from Study 1. Sections 3.4.1 and section 3.4.2 presents the detailed gesture classification of gestures from VE 1 and VE2. Gesture Taxonomy in VR is presented in 3.5. This is followed by a proposal of two new gesture taxonomy which is presented in section 3.5.2. The chapter is summarized in section 3.6. The elicitation study presented in this chapter presents the base for all our future studies.

Chapter 4: **Design and Evaluation of Locked Dwell Time-Based Point and Tap for Small Object Selection Within Arm's Reach in Dense Environment** proposes a novel technique *Locked Dwell Time-Based Point and Tap (LDTPT),* for small object selection at arms' reach in dense VE. Section 4.1, presents the design rationale for *LDTPT* technique. This is followed by a walkthrough of the *LDTPT* technique in section 4.2. Section 4.3 presents a user study to investigate the efficiency of the technique with techniques in the literature. The objective of the study, baseline techniques, design of VE, participants, study procedure, study tasks and data collection method are presented in sections 4.3.1-4.3.8. The results of the study are presented in section 4.4. The findings and discussions of the study are elaborated in section 4.5. The chapter is summarized in section 4.6.

Chapter 5: Design and Evaluation of Tiny Hands: An Object Selection Technique to Select Small Objects Within Arm's Reach in Dense VE proposes a novel technique *Tiny hands* for object selection at arm's reach refined by suggestions provided in the earlier chapter. The chapter starts with the design rationale for the technique and presents the three interaction techniques used to trigger the tiny hands' gesture in section 5.2. The three interaction techniques are: *fisting and rotating the non-dominant hand*, *holding the palm downwards with the non-dominant hand and performing vertical movement*, and *pinching-in/out the non-dominant hand* (section 5.2.1-5.2.3). Section 5.3 presents an informal evaluation of interaction techniques of tiny hands. The insights and findings of the study are presented in section 5.4. Based on the results of the study *holding the palm downwards with a non-dominant hand and performing a vertical movement* interaction technique was finalized to trigger tiny hands. Section 5.6 presents the second study to evaluate *tiny hand* for small object selection within arm's reach in a dense VE. Section 5.7 presents the tiny hands technique using the palm downwards with non-dominant hand and performing vertical movement to trigger tiny hands. The user study is presented in section 5.8. The finding and results of the study are presented in section 5.9. The findings are elaborated on and discussed in section 5.10. Finally, section 5.11 provides recommendations for designers to design object selection techniques for small objects placed within arm's reach. Section 5.12 summarizes the chapter, emphasizing the effectiveness and versatility of the tiny hands' technique for small object selection in dense VEs, and its potential to improve user experience in virtual reality applications.

Chapter 6: Design and Evaluation of AMAZE Technique to Select Small, Distant Objects in Dense Virtual Environments proposes a novel technique AMAZE and AMAZE-X for small object selection at a distance. We present the design and evaluation of the gesture identified for VE2, Dense VE, where targets are small and placed at a distance. The finalized gesture for VE 2 was *pinch out the VE and point and tap*. We adopted the elucidated gesture to redesign it to suit the challenges of VE2. We present a novel technique, AMAZE (A Multi-finger Approach to Zoom in dense Environments) that offers zoom using multiple fingers in a dense VE where objects are small and placed at a distance (beyond arm's reach). The design details of the AMAZE technique are elaborated in section 6.1. The design rationale for AMAZE is presented in 6.1.1. Design details of AMAZE is presented in 6.1.2. This is followed by evaluating AMAZE to investigate its accuracy and task completion times with existing selection techniques. In the first study, we evaluate AMAZE with two techniques from the literature: Expand and Pinch-to-Select. The study details are presented in section 6.2. Section 6.2.1- 6.2.8 presents the hypothesis, baseline techniques, VE design, participants, study set-up and apparatus, tasks, study procedure, and data-collection method. Section 6.3 presents the results. The findings are discussed in section 6.4. Our results indicate that the AMAZE technique was significantly accurate, and natural. However, it took a higher selection time compared to other techniques. With our analysis supported by participants' comments, we redesigned AMAZE and presented AMAZE-X. Section 6.5 presents the design of the AMAZE-X technique. We improved upon AMAZE-X's zoom factor with the increased CD ratio of 1:5 and added subtle dash lines as visual elements during the corrective phase to improve distance to the user for faster selection Section 6.6 presents the evaluation of AMAZE-X with techniques from study 1. The results of the second study are presented in section 6.7. The results indicate that AMAZE-X was significantly faster, easier to perform and easier to learn than other techniques. We also found AMAZE-X to be the most preferred technique. The findings are discussed in section 6.8. We also provide a set of design recommendations elaborated in section 6.9. Finally, we summarize the chapter in section 6.10. The development of AMAZE and AMAZE-X is an important step toward improving the selection of small, distance targets in dense virtual environments. In addition, from the two experiments, we have extracted recommendations that can help the design of object selection techniques for 3D virtual environments

Chapter 7: Conclusion is the final chapter, which provides an overview of the thesis. This is followed by presenting the contributions of the research. The limitations and scope of future research directions for small object selection in dense VE are discussed in the last section.

Chapter 2

2. A Review and Analysis of Existing Literature

Object selection is the process by which a user can interact and choose a single object or a group of objects within a virtual environment. Object selection is the primary and the most important task in an immersive virtual environment. It is the initial task for most common user interactions in a VE of the four tasks manipulation, locomotion and application context (Argelaguet and Andujar, 2013). These tasks are often preceded by an object selection task. As indicated designing a selection technique is of utmost importance.

Object selection in VR can be done by controller-based and controller-less methods. In controller-based object selection, the user holds a controller device in their hand, which is tracked in the VR space and can use to point at and select objects. In Controller-less object selection, the user can use hand gestures to select, manipulate, and interact with objects in the virtual world. To implement gesture-based object selection in VR, developers can use a variety of techniques such as hand tracking, gesture recognition, and motion sensing to interpret the user's hand gestures as commands for object selection and manipulation. Gesture-based object selection can greatly enhance the user experience in VR by providing a more natural and immersive way to interact with objects in the virtual world. It can also provide users with a greater sense of control and agency within the virtual environment.

This chapter is an intersection of existing literature on the focus areas of this research: object selection (section 2.1), controller-based object selection techniques (section 2.2) and controller-less object selection techniques (section 2.3). Within these two domains object selection for small targets positioned at arms' reach and distant in dense VE are presented. Figure 1 demonstrates a visual representation of the intersection for this research.

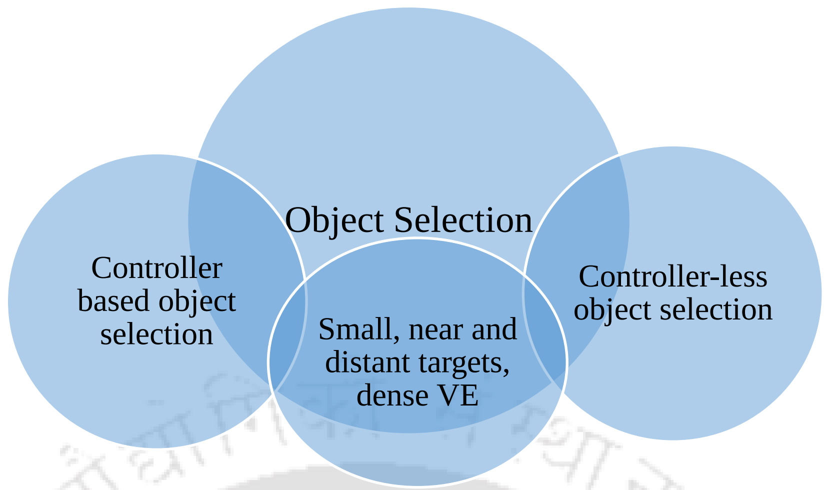


Figure 2: A figure demonstrating the intersection of three focus areas for literature review: object selection, controller based object selection techniques and controller-less object selection techniques for small object selection techniques placed at varying distance from user in dense VE

This chapter aims to provide a comprehensive understanding of object selection and object selection techniques in Virtual Reality (VR). Section 2.2 provides an overview of the existing literature related to controller-based object selection techniques. This section outlines various methods to select objects using controllers, followed by a discussion of the challenges and limitations of these techniques, particularly for selecting objects in small and dense VR environments (section 2.2.3). The next section, section 2.3, presents a review of the existing literature on controller-less object selection techniques, along with the limitations and challenges associated with these techniques. Finally, section 2.4 discusses the overall limitations, challenges, and research gaps in designing a natural object selection technique for small objects in dense VR environments, considering both near and distant objects. The research questions are presented in section 2.5. It further justifies the need to explore natural interaction modalities, with body-gestures. This section forms the basis of our research direction with gestures as interaction modality being the core focus of this research.

2.1 Introduction to Object Selection in VR

Object Selection is the task of acquiring or identifying a particular object or subset of objects from the entire set of objects available. It is the process of indicating an object for final selection. It is also called target acquisition (Zhai et al. 1994). Object selection tasks precede object manipulation such as rotation, scaling, movement, deletion, etc. for the VR application. Object selection tasks also precede virtual locomotion such as determining the position to travel to destination, travel towards to desired destination or anything related to virtual locomotion. If the user cannot select virtual objects effectively, object manipulation and locomotion tasks cannot be performed accurately and effectively. Overall, object selection is a primary task performed frequently (Bowman et al., 2001) and is the most crucial task to use, operate, and efficiently experiencing VR applications.

A selection task is comprised of three sub-tasks: indication of object, feedback, and confirmation of selection. Before interacting with virtual objects, it requires a set of selectable objects. Depending on the number of targets, a technique has to be designed to indicate the object for selection that could be single-object selection, multiple-object selection or multiple single-object selections. After which the user needs to specify the target for the final selection. Furthermore, the user should get adequate feedback, e.g., visible, audible or tactile, about a possible or an already performed object selection. This selection is generally considered as an interaction technique itself and its direct input interaction is the main focus of this research.

VR offers unique possibilities for object selection due to its 360-degree, 3D simulation, and 6 DoF possibilities. A variety of applications for object selection are designed with different complexities including objects presented in a dense environment, small objects and the selection of small objects within the user's arm's reach and at a distance. Applications that leverage these possibilities for object selection can include a wide range of industries, such as healthcare, education, engineering, and entertainment. For example, medical simulations can use VR object selection to train surgeons and medical professionals to precisely select small objects, while education applications can use object selection to teach concepts in physics, chemistry, and biology. Industrial applications can use object selection to train workers in manufacturing and maintenance processes, while gaming applications can use object selection to create unique and immersive experiences for players.

2.2 Object Selection Techniques

A 3D object selection technique requires the user to provide input in 3D space by a gesture, speech or audio e.g. grabbing an object using fingers or pointing to something and saying, "put that there" for the selection to happen. The effectiveness of 3D selection techniques greatly depends on the selection tasks to which they are applied. The same technique could be intuitive and easy to use in some task conditions and utterly inadequate in others. For example, the techniques needed for the selection of virtual objects in immersive gaming applications could be very different from the selection techniques used for surgical training in a medical simulator. Also, the effectiveness of the techniques varies depending on object selection in different VEs including dense VE, varied object sizes and distances.

In the following section, we present Controller based object selection techniques for small, dense VEs for targets positioned near and at a distant from the user.

2.3 Controller-based Object Selection in HMD-based Interfaces

In controller-based object selection, the user holds a controller device in their hand, which is tracked in the VR space and can use it to point at and select objects. For e.g. the Oculus Quest HMD, have a trigger button, a grip button, two thumbsticks, and several other buttons that can be used for object selection and interaction. To implement controller-based object selection in VR, developers can use a variety of input techniques. For example, a common technique for selection within arms' reach is to cast a ray from the controller to determine the object that the user is pointing at developed by Mine et al., (1997). It involves using a virtual ray projected from the user's controller to intersect with virtual objects in the VR environment. When the ray intersects with an object, it can trigger an event that selects the object by pressing a button on the controller (Figure 2a). This technique is perhaps the simplest and most efficient selection technique for targets that are moderately/large-sized objects placed at close range. However, this technique could be difficult to use, when a high precision of selection is required, such as when selecting small or distant objects (Poupyrev et al., 1998; Bowman et al., 1999). In such task scenarios, raycasting selection performance decreases significantly because of the high angular accuracy required and the amplification of hand and tracker jitter with increased distance (Liang and Green 1994; Poupyrev et al., 1998). To overcome the mentioned challenges Grossman and Balakrishnan, (2006), implemented three variations of raycasting: the depth ray, lock ray and flower ray technique. The depth ray improves the ray with a depth marker, visualized as a small sphere, along the length of the ray (Figure 2b). The position of the depth marker can also be controlled dynamically. The distance between the hand and the surface is mapped to the position of the depth marker, using absolute mapping. Moving the hand forwards and backward will move the depth marker in the same manner. With the depth ray, all targets which are intersected by the ray are highlighted green. Of these intersected targets, the one which is closest to the depth marker is highlighted red, indicating that it will be selected with a button press. While depth ray reduces the task completion times by integrating selection and disambiguation phases, it is neither designed nor experimented in dense VE where targets are small. Further, the challenges of specifying the depth and position of the depth marker make the technique unsuitable for dense environments. An improved version of the depth ray is the Lock ray by Grossman and Balakrishnan, (2006), which locks the position of the ray to avoid selection and disambiguation. The user presses and holds a button to highlight the intersected target. The release of the button selects the closest target. However, in a dense VE, there might be ambiguity as there will be too many objects in close proximity for final selection. The third ray-based technique is the flower ray technique by Grossman and Balakrishnan, (2006). In this technique, when the user clicks and holds the button, all intersected targets animate towards the user's viewpoint and flower out into a marking menu. The rationale behind this design is that a marking menu selection should be faster than the disambiguation phase of the lock ray, which is selecting an item from a linear menu. While the marking menu will potentially make the flower ray faster than the lock ray, a possible drawback is that users will need to follow the animation and find their intended object in the marking menu. Also, it increases visual clutter in a dense VE and there will be issues of ambiguity regarding the intersected targets. These ray cursor-based techniques improve movement time however these techniques are implemented in volumetric display for targets that are near. However, these techniques might not perform well for small objects in dense environments. Ninja hands by Schjerlund et.al (2021), investigates the use of many hands for target selection within arms' reach and at distance (Figure 2c). The technique outlines the benefits over existing hand-based techniques in terms of faster completion time however, different ratios between the placement and density of hand and target arrangements might cause more problems with disambiguation. In very dense clusters of targets, more hands would enter the queue and the queue would take longer to resolve. Hence, these techniques need to be effectively designed especially for the selection of small targets and dense VEs.

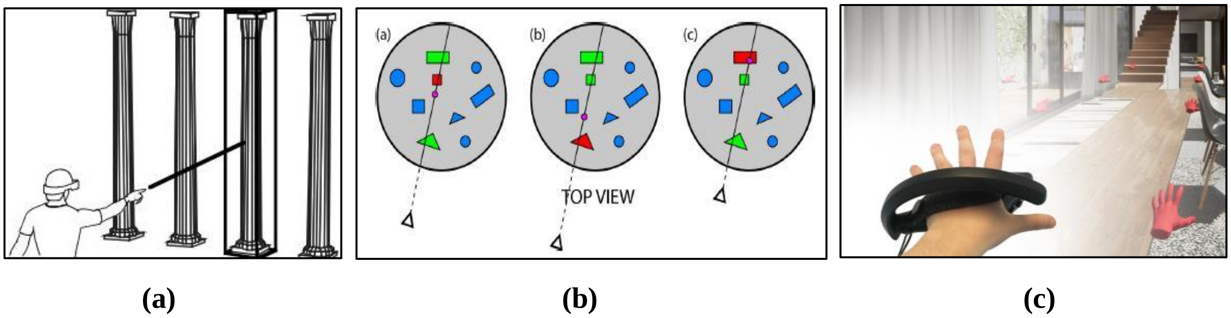
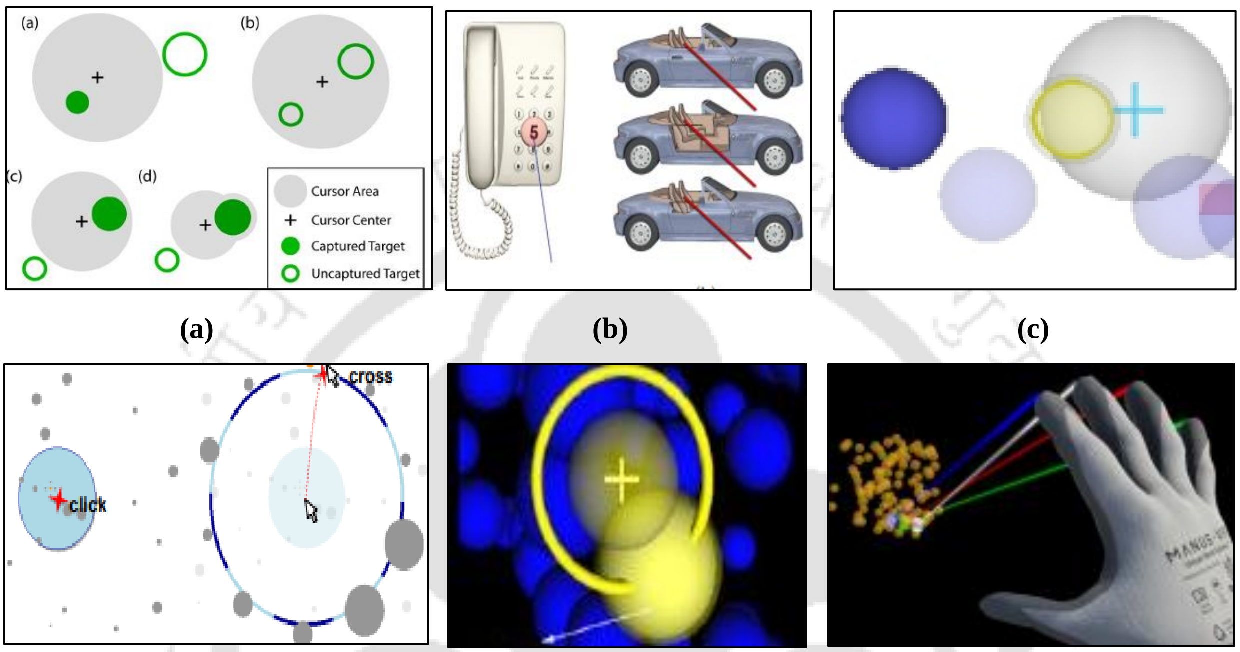


Figure 2: Controller-based object selection techniques within arms' reach and at a distance (a) Raycasting using data gloves for selection of targets within arms' reach (Mine et al. 1997) (b) The Depth ray and Lock ray technique selects objects using a depth marker and locks the position of the ray to avoid accidental selection (Grossman and Balakrishnan, 2006) (c) The Ninja hands technique (Schjerlund et al. 2021) brings the possibility of using many hands for target selection within arms' reach and at a distance.

Some techniques address the selection of small objects by increasing the size of the selection tool at the expense of requiring disambiguation mechanisms to guess the object the user aims to select. The Silk Cursor by Zhai et al. (1994) creates a region within the volume cursor to select small objects. With this, the cursor activation area is increased, reducing the acquisition times for small targets. However, the problem with this technique is that multiple objects could be brought under the cursor at a time, hence making it difficult to precisely select a single small object. An area cursor is a cursor that has a large activation area and has been found to perform better than other regular cursors for target acquisition tasks. Worden et al. (1997) propose an enhanced area cursor by including a single-point hotspot centered within the area cursor, which takes effect when more than one target is within the cursor's boundary. The enhanced area cursor performed identically to regular point cursors when targets were close together and outperformed point cursors when targets were far apart. However, two significant problems with area cursors are that large area cursors can obscure underlying data. It can be difficult, if not impossible, to use area cursors to select one target, small targets from several targets closely grouped together. Cockburn and Firth (2004) developed a similar technique based on expanding targets called bubble targets. Instead of increasing the size of the entire target, a bubble would appear around the target as the cursor approached. The Bubble cursor developed by Grossman and Balakrishnan (2005) improves upon area cursors by dynamically resizing its activation area depending on the proximity of surrounding targets, such that only one target is selectable at any time (Figure 3a). In the 2D technique, it dynamically resizes a circular cursor so that it only contains one object. A 3D extension of the bubble-cursor, which uses a sphere instead of a circle, was first presented by Vanacken et al. (2007). Various area cursor techniques have been designed however these techniques may actually perform worse in cluttered environments since even small movements will cause the cursor to constantly resize to select new targets (Figure 3b). It is also difficult to select very small targets. Argelaguet and Andujar (2008) proposed a technique by dynamically scaling potential targets and by using depth-sorting to remove occluding potential targets for selection (Figure 3c). However, the two presented embodiments do not provide a significant advantage in selection time. When targets are more closely packed together, the benefit of these techniques tends to degrade and can even be detrimental to selection accuracy. Rosa and Nagel (2010) proposed three bubble cursor variations: double, bound and depth. The double-bound cursor has two semitransparent spheres, called depth sphere, which varies in size in accordance with the depth at which the cursor is located. The Bound bubble cursor has a 2D ring, which varies in size according to the depth of the cursor (Figure 3e). In both of these techniques, the object in contact with the inner sphere is the one selected. In the depth bubble cursor the object closest to the center of the cursor, among all the objects in contact with the sphere is selected. However, these techniques have not been do not perform well in highly dense VEs and also for distant targets. The motion-pointing technique by Fekete et al. (2009) allows users to select individual objects without pointing at them, by assigning different elliptical motions to each object. It also reduces the number of selectable objects by selecting the top four motion matches and distributing them in a pie menu for direct selection. This technique, however, may not be suited for interfaces with dense objects, as it demands high precision. The click-and-cross cursor and cross-and-cross technique by Findlater et al. (2010) allows users to select an area of the screen and expand the items contained in that area into arcs that can be selected by clicking and crossing rather than pointing. However, if the user accidentally crosses the wrong arc, smoothly returning the mouse to the inside of the circle prevents a selection (Figure 3d). These techniques offer higher completion time for small targets however were negatively impacted in dense environments. Similar work by Ariza et al. (2018) included target sphere diameters ranging from 2.9 to 7.5cm, discs size ranging from 8.5mm to 612mm (Tu et al., 2019), and three target widths of "1◦, 2◦ and 3◦ of visual angle" (Kopper et al., 2010). Delamare et al. (2022) proposed a new variant of the 3D Bubble Cursor named Multi-Finger-Bubble. With Multi Finger Bubble, users aim at the cluster of targets of interest with their palms. Users can then control a semi-transparent sphere in the cluster with a 1:1 mapping between the sphere and users' hand positions. MultiFingerBubble includes up to four targets, and each target is then associated with a specific finger (Figure 3f). Users can then validate the selection of the desired target by flexing the corresponding finger.

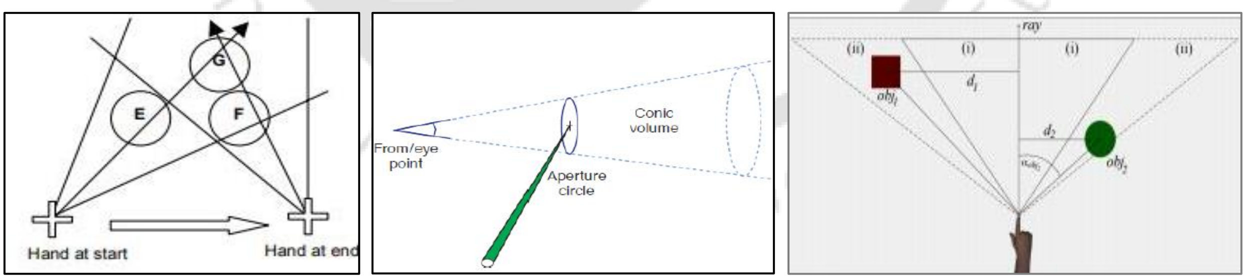


(a) (b) (c)

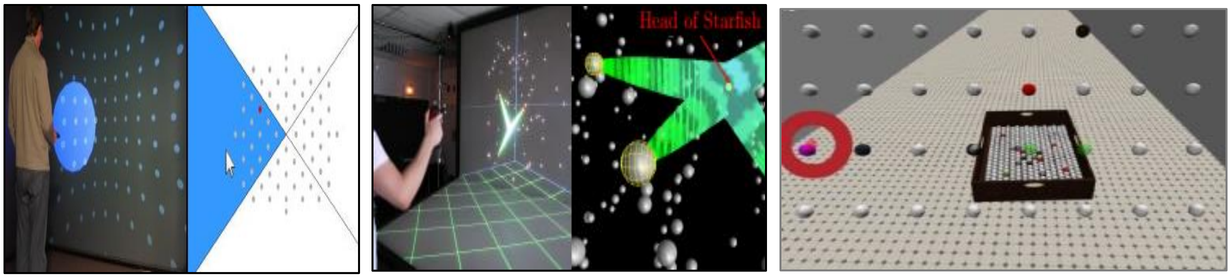
(d) (e) (f)

Figure 3: Object selection techniques for selection of small objects (a) The Bubble cursor (Grossman and Balakrishnan, 2005) improves upon area cursors by dynamically resizing its activation area (b) Dynamic scaling and Forced Disocclusion (Argelaguet and Andujar, 2008) dynamically scale potential targets and uses depth-sorting to remove occluding potential targets for selection. (c) 3D extension of the bubble-cursor (Vanacken et al., 2007) (d) The click-and-cross cursor (Findlater et al., 2010) allows the user to select by crossing the arc instead of pointing (e) **Bound Bubble cursor (Rosa and Nagel 2010) has a 2D ring, which varies in size according to the depth of the cursor. (f) Multi-Finger-Bubble (Delamare et al. 2022) proposed a multi-finger 3D Bubble Cursor**.

For selection in dense VEs, the ray-based techniques were improvised into cone-shaped volume techniques (Liang and Green, 1994). Cone-based techniques require the definition of a vector and a volume to determine users' selection intention. The axis of the cone can be increased and decreased depending on the number of objects to be intersected. The first cone-based technique to be implemented was the flashlight technique by Liang and Green (1994). It replaces the virtual ray with a conic selection volume, with the apex of the cone at the input device (Figure 4a). Objects that fall within this selection cone can be selected using the disambiguation mechanism. The technique, therefore, allows the selection of small objects even when located at a distance from the user. A problem with the flashlight technique is the disambiguation of the desired object when more than one object falls into the cone. Two basic rules are usually used for disambiguation (Liang and Green 1994). First, if two objects fall into the selection volume, then the object that is closer to the centerline of the selection cone is selected. Second, if the angle between the object and the centerline of the selection cone is the same for both objects, then the object closer to the device is selected. However, the task completion time using these techniques will be higher as multiple objects will be selected and there will be issues of fatigue while a selection of small objects in dense VE. The Aperture Selection by Forsberg et al. (1996) improves upon the Flashlight by allowing the user to control the spread of the selection volume (Figure 4b). The user can interactively control the spread angle of the selection volume simply by bringing the hand sensor closer or moving it farther away. The aperture technique thus improves the flashlight technique by providing an efficient interactive mechanism of object disambiguation by interactive control of the selection volume. However, in the aperture technique, a disambiguation metric is incorporated to choose among multiple objects and leads to increased task completion time and fatigue. Also, because the cone has an infinite extent, it is not clear what actions are appropriate when very distant objects are selected. Steinicke et al. (2006) introduced region examination and sticky ray interaction metaphor for object selection in VEs (Figure 4c). Objects within the cone-casted region are considered and depending on the distance from the center of the ray the nearest object is selected. If no object falls within the region an enlargement process of the target is performed and repeated until an intersection is found. In the sticky ray technique, the first object to be hit by the ray becomes the active object and remains active until the virtual ray hits another selectable object. However, in highly dense environments these techniques require users to interact very carefully to accomplish selection, and may actually result in worse performance than standard raycasting in some situations. Improvements on selection performance depend on the size of targets and the density of the VE, and again improvements are more apparent in sparse environments where the targets are not very small. To address these challenges, selection methods that use a progressive refinement of the set of selectable objects have been proposed. SQUAD by Kopper et al. (2011) uses a modified version of ray-casting that casts a sphere onto the surface to determine the selectable objects (Figure 4d). The indicated objects are split into four groups through a QUAD menu user interface. The user performs repeated selections until the selection contains a single object. SQUAD technique uses progressive refinement to iteratively select an object within a group of interest and is designed for selection in cluttered environments. This technique enables the accurate selection of a single target however it is time taking and induces fatigue. Also, these techniques have not been evaluated for the selection of small-size objects placed at a distance in dense VE. The Starfish technique (Wonner et al., 2012) employs a starfish-shaped closed surface for selection in dense VEs. Each branch ends exactly on preselected near targets (Figure 4e). When the desired target is captured by one of the branches, the user can lock the shape and select the desired target. However, the shape of the starfish is constantly rebuilt until the user finalizes the desired target. Adjusting the Starfish position to capture the correct target does require much effort, time-consuming and leads to fatigue. Cashion et al. (2012) developed the Zoom and Expand techniques for selection in dense VEs. The Zoom technique is an extension to raycasting that helps select small or partially occluded objects by first zooming in on the region of potential targets (Figure 4 f). The area inside the cursor is zoomed in to provide ample area for selection. The Expand technique is a variation of Zoom and SQUAD that selects with progressive refinement. After zooming in, the selected target objects are presented in a grid for the user to choose from depending on the number of targets selected. However, the disadvantage of using these techniques is that due to zooming the environment the original context is lost. These techniques are also time-consuming and induce fatigue for multiple object selection.



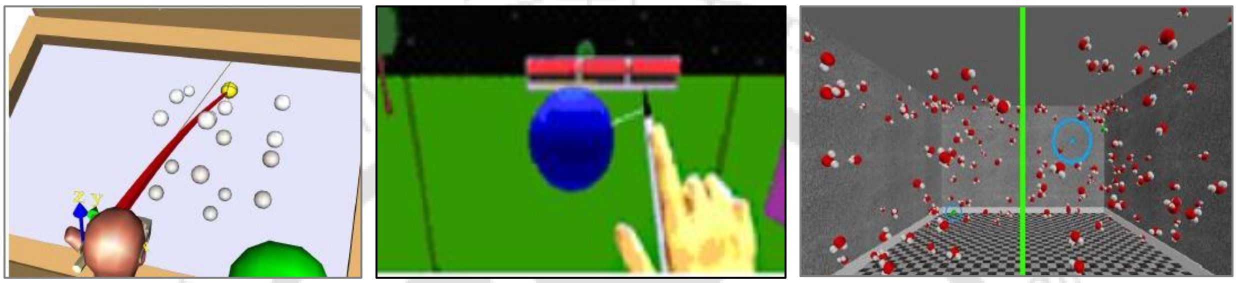
(a) **(b)** **(c)**



(d) **(e)** **(f)**

Figure 4: Object selection techniques for selection in dense VEs (a) Flashlight technique (Liang and Green 1994) replaces the virtual ray with a conic selection volume. (b) The Aperture Selection (Forsberg et al. 1996) improves upon the Flashlight by allowing the user to control the spread of the selection volume (c) Region examination (Steinicke et al., 2006) interaction metaphor for object selection in VEs. (d) SQUAD (Kopper et al., 2011) casts a Sphere onto the surface and divides the selectable objects in the form of a QUAD menu. (e) Starfish selection employs a starfish-shaped closed surface for selection in dense VEs Wonner et al. 2012 (f) Expand Technique uses zoom and SQUAD to select objects.

Some other technique that use heuristics and scoring-based functions for selection in dense VEs is the PRISM technique (Precise and Rapid Interaction through Scaled Manipulation) presented by Frees and Kessler (2005). PRISM is a behavioral approach to selection that dynamically adjusts the control-display ratio (Figure 5b). It scales up hand movement for distant selection and manipulation and scales the hand movement down to increase precision. While these techniques can achieve high levels of precision, they cause a significant mismatch of the physical pointing direction and pointing position, and the mapping is nonlinear. IntenSelect, by Haan et al. (2006), is another selection technique that dynamically assists the user in the selection of 3D objects in VEs. A scoring function is employed to calculate the score of objects, which fall within a conic selection volume (Figure 5a). By accumulating these scores for the objects, a dynamic, time-dependent, object ranking is obtained. The highest-ranking object is indicated by bending the selection ray towards it and the selection is made. The Smart ray by Grossman and Balakrishnan (2006) employs target selection based on target weights to determine which target should be selected when multiple targets are intersected. Target weights are continuously updated based on their proximity to the ray cursor. This technique, however, may not be suited for interfaces with many objects as the required precision would increase. The scope is another technique developed by Cashion and LaViola, (2014), in which the activation area of the cursor is altered in relation to the velocity of the handheld controller and the environment density. The hook technique is another heuristics-based method developed by Ortega, (2013) that uses time for computing the scores of each target during the selection (Figure 5c). These scores change during the pointing task, and their value increases or decreases depending on the distance between the cursor and each target. By sorting the targets by score, the system provides a list of the closest targets over time. A target that has been regularly proximal to the cursor will have a high score. The system considers the target with the highest score as “hooked” and provides visual feedback to inform the user that the target is selectable.



(a) **(b)** **(c)**

Figure 5: Controller-based object selection techniques that use heuristics and scoring-based functions (a) IntenSelect employs a score of objects within a conic selection volume (Haan el al., 2005) (b) PRISM technique (Precise and Rapid Interaction through Scaled Manipulation) presented by Frees and Kessler, 2005. (c) The scope technique developed by Cashion and LaViola, 2014 alters the activation area in relation to the velocity of the handheld controller and the environment density.

In addition to research experiments, a range of commercial HMD-based VR platforms enable controller-based object selection. HTC Vive uses a platform dependent on two motion-tracked remote controls to select and manipulate virtual objects. The remote controls have a button each which casts a ray in the VE and further gets selected when pressed. Similar to HTC Vive, Oculus Rift also uses a platform-dependent Oculus touch hand controller that cast a ray to hover the virtual objects. The objects are selected when a button on any of the Oculus touch is pressed. Google Daydream viewer and Samsung VR platform use a controller supported through ray casting which gets selected when a button is pressed on the controller. Although all commercially available platform performs object selection through ray casting, no evidence is present for exploring these controllers for small object selection in a dense VE for targets positioned at varying distances.

2.3.1 Research Gap in Controller-based Object Selection Method

Exploration of suitable objection selection techniques in HMD VR interfaces is of utmost importance due to (a) its increased usage and application in recent years and (b) different challenges than desktop and projection-based VR interfaces. Overall, a wide variety of controller-based input techniques has been explored for object selection in 3D UI and VR interfaces. Although in the past decade, the approaches have advanced from simple ray-casting-based methods and its variations to disambiguation and heuristic-based input methods, they still possess many challenges that impact effective object selection in VR interfaces, especially in selecting objects that are small, presented in a dense environment and have varied distances from the user. The current research on controller-based technique for object selection primarily targets mid to large-size object selection which is ineffective for small distant objects due to issues to occlusion, hand jitter and incorrect selection. While some techniques have attempted to investigate small distance targets, there has been less exploration of how these techniques are affected by the environment density i.e. dense VE. Moreover, most of these techniques are explored in desktop and projection-based VR, but are limited in HMD VR interfaces. At last, which in our opinion is the most important one - all these techniques require an active controller for object selection. This demands a user to hold an active object which is not suitable in conditions of physical and virtual multi-tasking (e.g. object selection in a VE while performing physical activities in the real-world) as the user needs to keep shifting between physical objects and the controllers. Moreover, it is unnatural, and demands learning to hold different devices and input interactions designed for different technology platforms (e.g. hand gloves, remote control, oculus touch etc.) which often increases the cognitive load and learning curve among the targeted users. This demands for new approaches that overcome the challenges of consistently learning to hold a physical object, learn to use new input interactions and take factors of different object density, object distance, and object size in a VE.

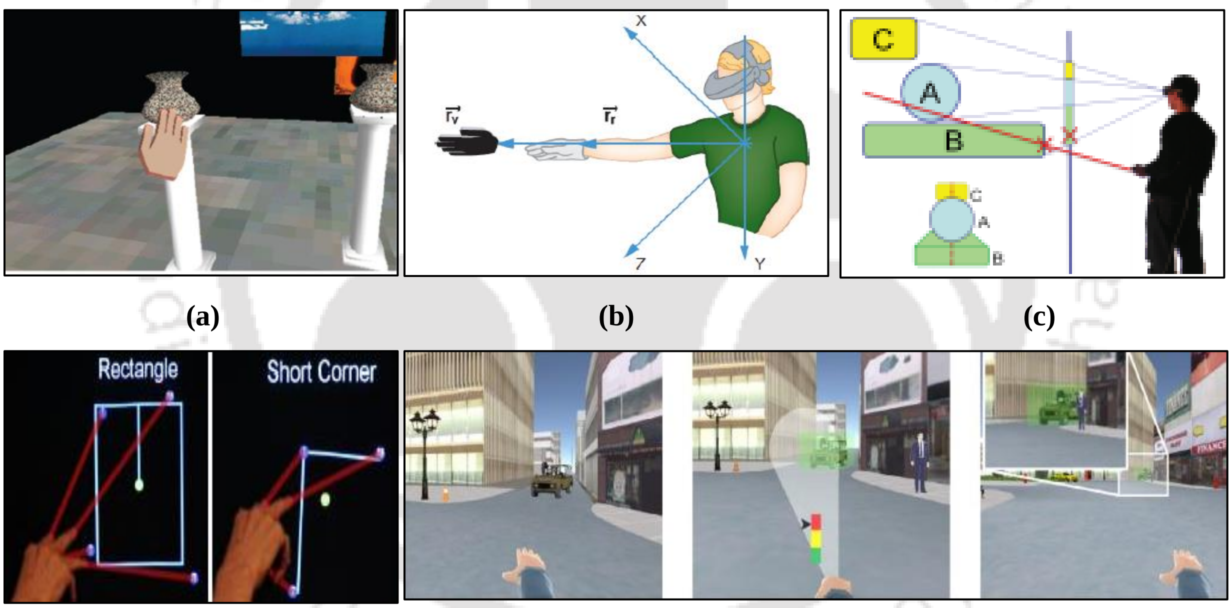
The following section presents the literature on the controller-less object selection method, its advantages and the scope of future research.

2.4 Controller less Object Selection in HMD-based Interfaces

In recent years, researchers have extensively explored controller-less input interaction methods primarily due to the challenges associated with physical controllers including the need to constantly hold the controller despite its inactive usage, learning new input techniques for each VR device, and the fear of breaking the controllers. One of the commonly observed approaches is to explore body-gestures due to their naturalness, acceptability and adoption (Ren and O'Neill, 2013, Seixas et al., 2015). This section covers the literature on controller-less input methods, advantages and research gaps in object selection.

In controller-less object selection, users can use hand gestures to select, manipulate, and interact with objects in the VE instead of traditional input devices like a mouse, joystick or handheld controllers. Mine et al. (1995) developed the virtual hand metaphor as an alternative to controller-based raycasting. In this approach, the user's hand position is tracked in real space which is used to control a hand cursor in 3D (Figure 6a). A one-to-one mapping is established between the device and the virtual hand. These techniques are deemed intuitive because they simulate our interaction with everyday objects and effectively select nearby objects. However, the virtual hand technique is difficult to select small objects as the hands might occlude the target and it could be time taking to precisely select a small target. Issues of lag and tracker jitter also decrease the selection performance. An alternative to virtual hand interaction is the hand extension metaphor to select distant targets. In this approach, the user's hand position is tracked in real space, typically using cameras, and the hand position is used to control a cursor or a virtual hand in 3D. The area of influence is the volume defined by the virtual hand, the control-display ratio is 1:1. An extension of the hand extension metaphor is presented through the method of arm-length. Arm-length refers to techniques where the length of the user's arm limits the reach of the virtual objects. The "Go-Go" technique by Poupyrev et al. (1998) attempts to improve the simple virtual hand by providing an unobtrusive technique that allows the user to interactively change the length of the virtual arm. When the user's real hand is close to the user, Go-Go uses a one-to-one mapping however, when the user extends their hand beyond a predefined distance threshold, the mapping becomes nonlinear and the virtual arm can be extended to select distant targets (Figure 6 b). The Go-Go technique provides direct, seamless, 6-DOF object manipulation both close to the user and at a distance. It allows users to both bring faraway objects near and select objects farther away. However, there has not been many experiments conducted to understand the maximum afforded reaching distance. Furthermore, as the distance increases, the technique maps small movements of the user's hand into large movements of the virtual hand, which complicates precise positioning at a distance. The Go-Go technique is usually less effective than ray-casting as it requires 3-DOF control as opposed to 2-DOF. The Stretch Go-Go by Bowman and Hodges (1997) technique improves on Go-Go by being able to extend the virtual arm until infinite. Wilkes and Bowman (2008) developed the Scaled HOMER (Hand-Centered Object Manipulation Extending Ray-casting) technique in which the user selects the object with the ray. Still, instead of the object becoming attached to the light ray, the virtual hand moves to the object position, and the object is attached to the hand. When the object is dropped, the hand returns to its natural position. This allows simple grabbing and manipulation. The Scaled HOMER improves upon the Go-Go technique by increasing precision when interacting with objects at a distance.

Ray casting methods have also been explored for controller-less input interaction for distant object selection. Nickel and Stiefelhagen (2003) proposed raycasting using the orientation of the head referred to as Head Ray Cast (HRC). They further investigated elbow-rooted techniques by using the ray between the elbow and the hand (forearm ray cast (FRC)). Wyss et al. (2006) developed a technique called iSith which uses two ray pointers, and a 3D point to select a target when the distance of these two rays falls below a defined threshold**.** However, this technique, it occupies both hands while selecting the virtual objects. The use of both hands may not be suitable for application that requires multi-tasking (e.g., selection & manipulation of virtual objects). These selection techniques have also not been implemented for the selection of small targets in dense VEs. Argelaguet et al. (2008) distinguish between ray-casting by eye-rooted and hand-rooted techniques. Using the eye orientation as a ray cast is referred to as gaze ray casting and is implemented similarly to pointing tasks using eye-tracking (Figure 6c). However, eye orientation ray casting requires special equipment and extra eye calibration. Mayer et al., (2015) evaluated three ray-casting techniques Index Finger Ray Cast (IFRC), Eye Finger Ray Cast (EFRC) and Forearm Ray Cast (FRC) to investigate pointing accuracy for targets at a distance of 2m and 3m. They found the IFRC was the most accurate however had errors while selecting distant targets. Mayer et al. (2018) further investigate the use of correction models to investigate the impact of visual feedback on humans' pointing performance for different ray-casting methods. Similarly, Matulic and Vogel (2018) explored Multi-ray (multi-finger ray-casting) where each finger directly and independently emits an unmodified ray from a distance on large-scale displays (Figure 6 d). Rays were cast from the entire finger and also from the distal phalanx. The proposed techniques, created by hand postures, form 2D geometric shapes to trigger actions and perform direct manipulations that extend the single-point selection. However, these techniques have not been sufficiently explored for dense VEs and small objects. Also, fatigue is another crucial aspect that needs to be taken into consideration. Mendes et al. (2017) proposed PRECIOUS, a mid-air technique for selecting out-of-reach objects featuring iterative refinement in VR interfaces PRECIOUS (Progressive REfinement using cone-casting in Immersive virtual environments for Out-of-reach object Selection) uses cone-casting from the hand to select multiple objects and moves the user closer to the objects in each refinement step, to allow accurate selection of the desired target (Figure 6 e). It offers infinite reach, using an egocentric virtual pointer metaphor. While pointing, users can also make the cone aperture wider or smaller, and change the cone's reach. While this technique is suitable for selecting objects at a distant, it might be difficult to manipulate the selection cone in such a way that it only intersects a single object when two objects are very close to each in cluttered environments.

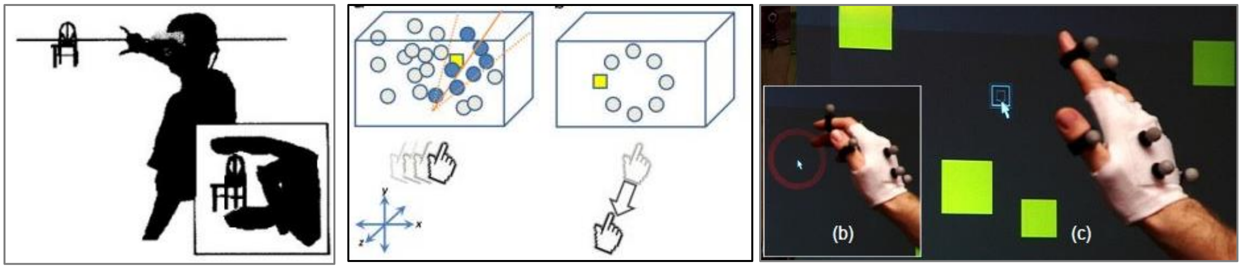


(a) (b) (c)

(d) (e)

Figure 6: Virtual Hand extension (Mine et al. 1997) developed as an alternative to controller-based raycasting (b) The Go-Go technique (Poupyrev et al. 1996) attempts to improve on the simple virtual hand by providing users to interactively change the length of the virtual arm. (c) Eye-rooted and hand-rooted techniques (Argelaguet et al. 2008) (d) Multiray (Matulic and Vogel 2018) uses multi-finger raycasting to independently emit unmodified rays from a distance on large-scale displays (e) PRECIOUS, Mendes et al., 2017, uses cone-casting for selecting out-of-reach objects featuring iterative refinement in VR interfaces.

Various other selection techniques using hand gestures was used in the Head crusher technique developed by Pierce et al. (1997). In the Head crusher technique of the image plane, the user positions his thumb and forefinger around the desired object in the 2D image. The object is selected by casting a ray into the scene from the user's eye-point between the user's forefinger and thumb (Figure 7a). The sticky Finger technique by Pierce et al. (1997) provides an easier gesture when picking large or close objects by using a single outstretched finger to select objects on the user's image plane. However, holding your finger six inches in front of your eyes and alternately closing each eye demonstrates the importance of using the technique. However, this technique has also not been investigated for small object selection and might not perform well in dense VEs. Inspired by Bowman and Wingrave (2001) and Bowman et al. (2002) work on Menu selection using Pinch gestures, Ni et al. (2011) introduced the rapMenu technique, which allows menu selection by controlling wrist tilt and employing multiple pinch gestures (Figure 7b). It takes advantage of the multiple discrete gesture inputs to reduce the required precision of the user hand movements for object selection. Vogel and Balakrishnan (2005) also explored different alternatives for triggering selection for free-hand pointing in large displays. They proposed two different hand gestures to perform the selection, AirTap and Thumb Trigger in a combination of visual and auditory feedback (Figure 7c). The AirTap technique is similar to how the index finger clicks a mouse button or taps a touch screen. For the thumb trigger technique, the thumb is moved in and out towards the index finger side of the hand providing kinesthetic feedback. Target sizes used were 144mm, 48mm, and 16mm and participants stood 4m away from targets. Results found no significant differences in completion time between the two techniques. Also, it was difficult to select small (16mm) targets while standing 4m away from the display using these techniques. Mendes et al. (2014) evaluated a set of techniques using a pinch (index finger and thumb) and grab gestures to directly select and manipulate objects in mid-air using a stereoscopic table top. Lin et al. (2016) implemented three ways of interacting with objects: innate pinching, magnetic force, and a physical button attached to the index finger**.** Innate pinching requires grasping an object with a pinch gesture and in magnetic force, a finger magnetically attracts an object within the grabbing object distance to grab an object. This implementation solves the accuracy problem other pinching implementation presents. However, it also requires more concentration to select and deselect objects and also to select small objects. It is also often hard to correctly select objects which are close together.



(a) **(b)** **(c)**

Figure 7: Controller-based object selection using body gestures (a) The Head-Crusher technique (Pierce et al. 1997) in which the object is between the user's finger and thumb in its image plane. (b) The rapMenu technique (Ni et al., 2011 allow menu selection by controlling wrist tilt, point and pull gestures (c)AirTap and Thumb Trigger (Vogel and Balakrishnan, 2005) performs finger clicks and thumb trigger taps on the side of the index finger providing kinesthetic feedback.

Researchers have also explored disambiguation methods for controller-less object selection to overcome the challenges of dense VEs. Ren and O’Neill (2013) investigated the design of freehand 3D gestural interaction that could be used in everyday applications without complex configuration or attached devices or markers. The study investigates target selection techniques in dense and occluded 3D environments. The selection is accomplished by using a selection cone whose apex is kept fixed while the center of the cone base can be moved in a vertical plane by the movement of the user’s hand. The objects that are intersecting the cone can be selected, and a menu-based selection disambiguation method similar to SQUAD is employed. In this method, the user can perform a hand pull gesture in order to display a 2D menu that lists the objects being intersected by the selection cone. Here, the selection is accomplished by picking one of the listed items. While such menu-based disambiguation methods are accurate, they remove the environmental context from the selection procedure and reduce the user’s sense of presence in the VE. Moreover, these techniques do not investigate VE with small object selection positioned at varying distances. Seixas et al., 2015 compared two selection gestures: screen tap, and hand grab. The screen tap gesture consists in moving the pointing finger towards the screen and returning to the original position. The hand grab gesture requires two hands to point and select: the dominant hand is used for controlling the position of the pointer on the screen; the auxiliary hand is used to perform the selection by closing and opening the hand (i.e., making a fist). The target size used for the study was 13mm. Results indicate that the hand grab gesture that uses two hands improves the performance of the selection task was 12 % faster when compared with the screen tap gesture.

Vosinakis and Koutsabasis (2018) evaluated grasping (using all fingers) with bare hands. However, their study was focused on providing different visual feedback techniques. Mutasim et al., (2019), compared the performance of pinch with click and dwell selection techniques for eye-gaze-based selection in VR. Results revealed that dwell as a selection technique made the least errors. However, it had the worst performance in terms of selection time and throughput. Pinch was sometimes frustrating due to recognition errors and seems to have induced more physical and mental fatigue however, authors recommend pinch when controller-less selection technique. In their ISO-based pointing tasks, target sizes used were (1.5, 2.5, and 3.5 cm) and target distances were (6°, 7.5°, 9°, 10.5°, or 12°).

IDS method by Periverzov and Ilies (2015) offers hand placement fault tolerance according to the level of confidence shown by users with respect to the position in space of their hands. A proximity sphere is placed around the simplified hand model of the user such that the fully extended fingers of the hand touch the interior surface of the sphere. The proximity sphere is swept along the path by the motion of the hand, and the objects that are intersected by it are considered to be candidate objects for selection. The size of the proximity sphere is adjusted according to the users' level of confidence about the position of their hand. These studies offer ways to select. In addition to hand, arm and finger-based gestures, researchers have investigated head movement for object selection in VR interfaces. Ramcharitar and Teather, (2018), presented EZ cursor VR which is a 2D head-coupled cursor fixed in the screen plane for selection in HMD VR. The control cursor was fixed in the center of the field of view, and thus can be controlled by the user's head gaze as well as by external input devices. However, the head gaze did not perform well in terms of selection accuracy and task completion time as compared to external input devices. The selection of small distant targets was also difficult for selection.

Commercially available google cardboard or similar low-cost cardboard platform uses a ray casting method that explores the VE using the head movements and activated using dwell gaze. Oculus quest devices incorporate pinch gestures. However, the cone-based visualization and the cursor-to-point make it difficult to select small targets in dense VE where targets are at a distance. Although researchers have investigated the use of body-gestures combined with voice or other activation methods, they still face challenges of different languages, dialects, privacy issues (especially in a public environment) and accuracy issues in a dense VE setup.

2.4.1 Research Gap in Controller-less Methods for Object Selection

The literature suggests extensive use of body-gestures for controller-less object selection in VR interfaces. This includes the use of ray casting including multi-finger and two-hand ray casting, virtual hand-based object selection, and arm and finger-based gestures other than pointing and disambiguation techniques. Variations in disambiguation and heuristic-based methods have also been incorporated. They still possess many challenges that impact effective object selection in VR interfaces, especially in selecting objects that are small, presented in a dense environment and have varied distances from the user. Most of the techniques are studied in desktop and projection-based VEs, and limited work is found for object selection in HMD-based VR interfaces. Moreover, very few techniques have been designed for small object selection, small object selection in dense target VEs, or the selection of objects which are occluded from the user's viewpoint. Further, the issues of object distance, object proximity, and object size are yet untouched in the literature of object selection using the controller-less input method.

2.5 Chapter Summary

Overall, this chapter presents a comprehensive overview of object selection and object selection techniques. We present a literature review of the two focus areas for object selection using controller-based and controller-less object selection techniques. Within these two domains, a selection of small objects positioned at arm's reach or distant objects in dense VE is presented. Through in-depth reporting and analysis of existing literature, the lack of research for small object selection the challenges of selecting small objects positioned within arms' reach or at a distance in dense environments using controller-based techniques are presented. It also presents limitations of controller-based object selection such as demands user to hold the device, learn different input interactions for different platforms and hassle in a continuous shift from physical objects and controllers. It also establishes the fact that most of these techniques are explored in desktop and projection-based VR, but are limited in HMD VR interfaces. The literature suggests the use of controller-less techniques, such as body gestures for object selection in VR interfaces. However, most of these body-centric object selection techniques are studied for desktop and projection-based VR, and there is limited research on HMD-VR interfaces. The current literature also presents the limitations and suitable interventions in investigating the effectiveness of body-gestures for object selection in dense VEs including varied object sizes, proximity, and distances.

In the next chapter, we present the methodology to elucidate suitable body gestures for small object selection for two environments (i) Dense VE where targets are small and placed at arm's length and (ii) Dense VE, where targets are small and placed at a distance. The gesture elicitation study is divided into two stages: (i) study to generate natural and intuitive object selection gestures for two VEs and (ii) study to evaluate the finalized gestures of study 1 for the chosen VEs. These two studies are strongly interrelated to each other and create relevant gesture designs for formalizing the final gesture-based experiment for this research.

Chapter 3

3. User-Generated Gesture Elicitation Study

The study of object selection in virtual environments has been an ongoing topic of research since the inception of virtual reality. The recent availability and popularity of commodity 3D tracking hardware and its expanding application areas, ranging from home entertainment to engineering and medical systems, have renewed interest in gesture-based object selection in virtual reality. Touch-sensitive interactive devices have been developed that do not require the user to hold or wear any special equipment or markers, thereby enhancing the fluidity and immediacy of freehand interaction in virtual reality. These sensing devices accurately recognize finger, hand, and body-based gestures and are now becoming available at low costs, such as the Microsoft Kinect and Asus Xtion.

Body gestures have been widely used in Human-Computer Interaction (HCI) to enable intuitive and easy interaction between users and devices (Norman, 2010; Tian, 2017). Body gesture interaction is more natural, intuitive, and comfortable for users than traditional interaction methods (Nacenta et al., 2013; Pereira et al., 2015). Despite being a natural means to interact, gestures can be difficult to use, learn and memorize (e.g., Norman and Nielsen, 2010; Zhao et al., 2014), especially in the context of object selection. This difficulty arises in large part because designers have defined the gestures without fully considering the compatibility between the gestures and the task. Therefore, it is important to consider users' mental models when developing body gestures for gestural interaction (Wickens et al., 2021), especially for object selection in dense virtual environments where targets may be placed at varying distances.

Gesture elicitation studies have emerged from the field of participatory design and have been widely applied to help designers select the most appropriate gesture set for a given function and application context. The main approach is to first define the function and then ask end-users to propose gestures that they find suitable for the function. Finally, the gesture vocabulary is extracted after analyzing the collected data. User-defined gestures are typically more intuitive and natural, compared to arbitrary designs that are technology or designer-centric. Earlier research by Nielsen et al. (2004) has used elicitation to distill ergonomic designs for gestural interfaces. Besides being more intuitive and natural, user-defined gestures are easier to remember and are preferred by non-technical users. Some researchers have reported that user-elicitation can actually help develop more complete sets of gestures than those defined only by experts or designers. This approach has been used for eliciting user-defined gestures for various devices and applications, including mobile devices (Ruiz et al., 2011), augmented reality (Piumsomboon et al., 2013), tabletop systems (Zaiţi et al., 2015; Vatavu, 2012), and cross-device interaction (Wu, 2016).

In this chapter, we aim to investigate RQ1 (RQ 1.1 and RQ1.2). The objective of this study is to propose natural and intuitive gestures for the selection of small objects in two types of dense VEs: (i) VE1: one where the objects are within arms' reach, and (ii) VE2: where they are positioned at a distance. To achieve this, we conduct a two-stage gesture elicitation study. In the first stage, we gather gestures from 40 participants for object selection in the two types of dense virtual environments. These gestures are then categorized and analyzed based on posture and overall score, resulting in three unique gestures for the VE1 and two for the VE2. In the second stage, we evaluate the extracted gestures in terms of ease of performance, body-part suitability, gesture appropriateness, user preference, and effort required to select a single gesture for each VE.

This chapter presents the methodology for identifying suitable body gestures for small object selection in dense HMD-based VR, taking into account varying distances. We commence with a comprehensive review of elicitation methodologies for gesture design, as discussed in section 3.1. We then proceed with the methodology for our first study, aimed at generating natural and intuitive object selection gestures for VE1 (section 3.2). The objective of the study, VE design, participants' details, study set-up and apparatus, and study procedure are presented in sections 3.2.1 to 3.2.5, respectively. The results and gestures collected during the study are reported in section 3.2.5. The final gesture list and visual representations for VE1 and VE2 are presented in section 3.2.7. Section 3.2.8 provides a discussion of the findings of VE1 and VE2. The second study evaluates the subjective ratings of gestures for VE1 and VE2 in section 3.3. The details of the participants' information and methodology, as presented in section 3.3. The results of the study are presented in section 3.4, followed by results on each of the environments VE1 and VE2 are presented in 3.3.4.1 and 3.3.4.2. Section 3.3.5 discussion of Results of Study 2. This is followed by the classification of gestures from Study 1 in section 3.4. Section 3.4.1 presents the detailed gesture classification for VE 1 for Study 2 and section 3.4.2 presents the gesture classification for VE 2 for Study 2. Gesture Taxonomy in VR is presented in 3.5. An introduction to gesture taxonomy is presented in 3.5.1. A proposal for gesture taxonomy is presented in section 3.5.2. The chapter is summarized in section 3.6.

3.1 Overview of Gesture Elicitation Studies

Nielsen et al. (2004) proposed a method for deriving a usable gesture set by collecting user-generated gestures and conceptualizing the functions of the proposed system across users. This approach involves extracting gestures based on the semantic representation of associated functions. Henze et al. (2010) extended this approach by validating the outcome of each step to derive a gesture set. They employed usage functions combined with participatory design to define a gesture set, which was further evaluated and improved. Wobbrock et al. (2009) also proposed a participatory design approach to derive basic gestures for surface computing. They used the think-aloud protocol and video analysis to obtain qualitative data and logged quantitative measures such as gesture timing, activity, and preferences to portray a set of user-defined gestures.

Overall, the above methods involve users in the design process and expound gestures from users' input. Since then, similar studies have been conducted to design user-centered gesture commands on mobile devices (Ruiz et al., 2011), augmented reality interfaces (Lee et al., 2015; Piumsomboon et al., 2013), controlling home appliances (Vatavu, 2012), smart TVs (Zhang, 2016) and multi-display environments (Vatavu et al., 2015). In the context of HMD-VR interfaces, a recent study by Ortega et al. (2019) employed eliciting gestures for object manipulation to improve whole-body interactions in VR. They found that for common referents (e.g., rotate, swipe object), users generate different gestures depending on the object's size. This study justified the assumptions on the effect of size on users' interaction in a VE. (Wu and Wang, 2002) designed a two-stage gesture elicitation study for object manipulation in VR. The collected gestures were validated for suitable match, memorability, learnability and comfort. The findings reveal a preference for metaphorical and one-handed gestures. Participants preferred large-scale dynamic gestures (e.g., limb-based stretch) when selecting distant objects on large displays. Wu et al., 2002 conducted a similar study for standard tasks (e.g., select, rotate, shrink or put on clothes) for VR shopping applications. Traditional WIMP interfaces and touch interfaces inspired the gestures collected. (Nanjappan et al. 2006) identified natural and intuitive interactions for 3D manipulation using dual-hand controllers. The results suggest a preference for shoulder motions (e.g., shoulder abduction and horizontal abduction) and elbow flexion movements. Besides, users preferred one-handed interactions that allowed them to alternate between their hands. Several elicitation studies have been conducted in varied contexts for object selection and manipulation. However, to our knowledge, there have been no elicitation studies designed specifically for object selection in a dense and occluded VE with varying object distances and small-size objects. For our study, we employed a user-generated gesture design method proposed by Wobbrock et al. (2009) to design a gesture set suitable for the context and users.

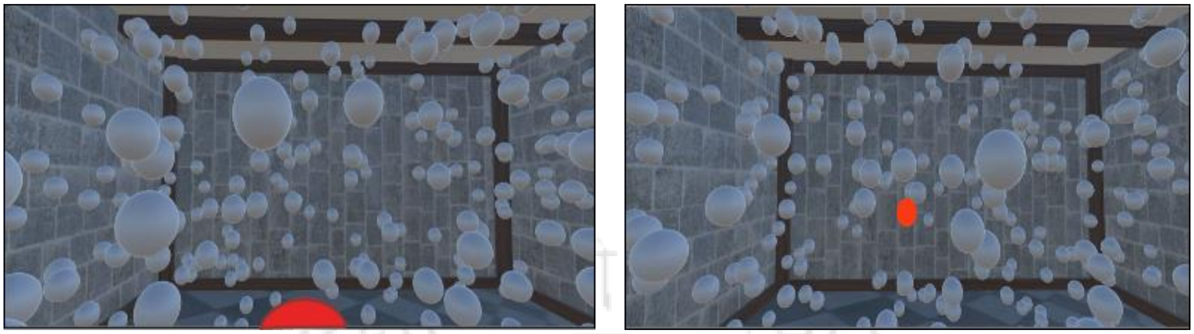
3.2 Study 1: Gesture Elicitation Study: Methodology

3.2.1 Objective

The aim of this study was to propose natural and intuitive gesture commands that incorporate full-body movements for small object selection in HMD VR. Specifically, we solicited participants to suggest the most natural and intuitive gestures for selecting small-size objects, in two distinct VE conditions – VE1: dense VE with small objects placed within arm's reach (50cm) and VE2: dense VE small objects placed at a distance (5.2m).

3.2.2 Design of VE

For this study, we utilized the Unity 3D game engine to develop two dense virtual environments (VEs), each containing 924 static spheres, including a red-colored target object and grey distractors. The small-sized spheres were randomly distributed in the VE and placed such that the participant's gesturing hand could aim and point at individual targets. The VEs were room-scale and measured 20 x 15 x 10 ft. The tracking area allowed participants to navigate through real walking. The nearest object was placed at a distance of 0.4 m, while the farthest virtual object was placed at a distance of 5.2 m, which was deemed plausible for room-sized scenarios. We designed two VEs: (i) VE1, which contained the target object placed within arm's length (50 cm motor space), and (ii) VE2, which contained the target object placed at a scaled distance of 5.2 m. The target object changed randomly after each gesture proposal in all VEs, and the task and VE type were displayed in the VE. To aid the participants, a virtual hand, tracked via the Leap Motion device, was presented in both the VEs. We calibrated the device when full-hand gestures were performed outside the Leap Motion device's range, resulting in poor tracking. Figure 8 provides an expanded view of the VEs and targets in the environment.



(a) **(b)**

Figure 8: Expanded view of the designed VEs – VE1 and VE 2 and the targets (in red color).

3.2.3 Study Participants

The study recruited 40 participants (27 males, 13 females) aged between 18 and 35 (Mean= 25.42, SD= 4.07) who were university students. Only right-handed participants were chosen based on the findings of Plaumann et al. (2017), which showed that acknowledging users' handedness can significantly improve selection accuracy. All participants considered their right hand to be their dominant hand and had prior experience using HMD-VR platforms, such as Oculus Rift and HTC Vive, for a minimum of 10 hours in the last six months. They were also familiar with gestural interfaces due to prior experience using a Nintendo Wii remote or Microsoft Kinect for at least 10 hours in the last six months.

3.2.4 Study Setup and Apparatus

The study utilized the Oculus Rift HMD-VR device, which was connected to a VR-ready computer with an i7-8700 quad-core processor, Nvidia Geforce 1060 GPU, 8GB RAM, and a Microsoft Windows 10 operating system. A Leap Motion device was mounted on the HMD-VR to show participants' hand position inside the VE. The Leap Motion Controller uses two monochromatic IR cameras and three infrared LEDs and supports hand and finger motions without requiring hand contact or touching. Participants performed the gestures while standing, and their actions were captured by two video cameras placed diagonally and horizontally to their position. Figure 9 provides details of the study setup and apparatus.

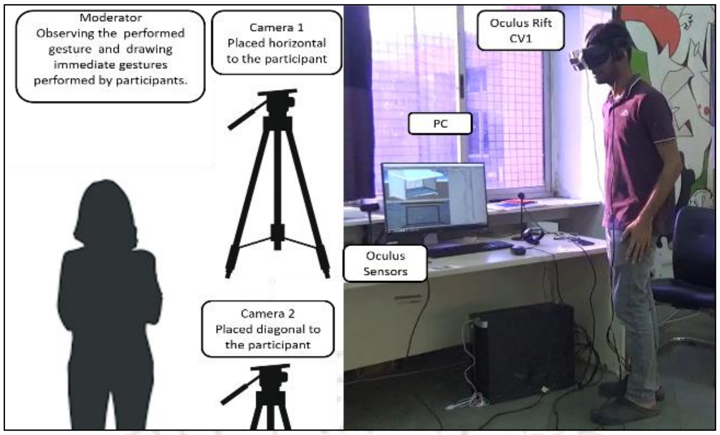


Figure 9: The study setup used for the elicitation study. The moderator observes the participant while the cameras placed diagonally to the participant record the gestures for further analysis.

3.2.5 Study Procedure

The study was conducted in a laboratory at a university. Participants were provided with a verbal introduction to the procedure by the moderator. Following this, they were given a 10-15 minute training session to familiarize themselves with the VR environment. During this time, participants were shown the VR environment and allowed to contemplate potential gestures for the upcoming task. The experiment required participants to be in a standing position, with arms kept at their sides. Participants were then instructed to propose up to 5 gestures that they believed would be the most natural and intuitive for selecting a target object. The target's location changed randomly after each gesture proposal in both VEs, and five different target locations were presented to participants for each VE. Participants were advised that they could use full-body gestures, rather than being limited to hand-based gestures, and were also informed to perform gestures without considering technical feasibility. After performing the body gestures, participants notified the moderator, and a Wizard of Oz setup was used to evaluate the gestures systematically. Participants were asked to explain the choice of a particular gesture and to talk aloud while performing the task to elucidate the motivation behind their selection. Following the completion of the task, participants ranked their proposed gestures from best to worst. Participants were allowed to rest at any time if they were tired during the experiment, which was recorded via video and took approximately 45 minutes per participant. Prior consent was obtained from each participant.

3.2.6 Data Collection Method

Following the performance of gestures, the moderator carefully examined and visually documented the beginning, end, and trajectory of each gesture in a notebook. To ensure the accuracy of data collection in a Wizard of Oz setup (Dahlback et al., 1994), the performed gestures were also confirmed with the participant and recorded via video. To determine the most suitable gestures, we applied a two-parameter approach, consisting of gesture frequency (equal to or greater than 10) and gesture score. The gesture score was calculated by combining two variables: preferences (1st and 2nd preference) and posture. This approach is similar to that used by Pareire et al. (2015). Posture referred to the final postures of the fingers, wrists, forearm, and legs, and assigned a biomechanical risk score based on the most extreme joint postures created during the gesture. For example, a full extension of finger joints was given a score of 3, while a partial extension was given a score of 1. The angle of extension/flexion of the wrist from a neutral position of 0° to 14° was given a score of 1, while an angle between 15° to 45° was given a score of 2, and an angle greater than 45° was given a score of 3. A higher score indicated a greater risk of fatigue. The posture scores were obtained from Pareire et al. (2015). To obtain the overall posture score for a gesture, the scores were normalized to a mean value of 10 and a standard deviation of 1 before being summed. The overall score was calculated by summing the normalized values of the two variables.

3.2.7 Results

The study involved 40 participants who generated a total of 390 gestures for two VEs. Upon analyzing the frequency of gestures proposed, 196 and 194 gestures were proposed for VE 1 and VE 2, respectively, resulting in a total of 52 (23 and 29) unique gestures, of which 47 were dynamic, and 5 were static gestures. Gestures were categorized based on the specific body parts used to perform them. For example, various pointing gestures performed with different fingers, thumbs, or palms were grouped under pointing.

After considering the participants' first two preferences, 7 unique gestures were finalized for each VE. Figures 10 and 11 provide a visual representation of the extracted gestures for both VEs, while tables 1 and 2 present details of the 23 and 29 unique gestures for VE1 and VE2, respectively. Tables 3 and 4 provide a comprehensive overview of the finalized unique gestures, including their frequency, preferences, posture, and overall score for VE1 and VE2. The variation section of the table includes gestures performed by participants using different hands or feet to achieve the same action, such as pointing with one finger, two fingers, three fingers, or the palm in a vertical position. In the study, participants used gestures for two purposes: direct selection and confirmation. Direct selection refers to the use of a gesture to select a target, such as a grab and rotate. On the other hand, confirmation refers to the need for a separate gesture to confirm the final selection, such as pointing with one finger and tapping.

Table 1: List of the Unique Gesture Details of VE 1: Dense VE, Object Placed at Arms' Length.

	Unique gestures	Variation	Direct selection	Confirmation	Final gesture	Freq
1		One hand	One hand	rotating the wrist clockwise and coming back to position	*Grab and wrist rotate*	20
2		one hand	one hand	nil	*Grab with one hand*	5
3		both hand	both hand	nil	*Grab with both hands*	6
4	Pointing	one finger (index finger)	index finger	diectic pointing	*Point with index finger*	6
5			index finger	emblem drawing a circle around the target	*Point with one finger and draw a circle around target*	5
6			index finger	pull hand backwards	*Point and pull hand closer backwards*	4
7			index finger	tap (once)	*Point and tap*	26
8		palm	vertically placed	thrust forward	*Palm pointing and thrust*	4
9		2 finger (index finger and middle finger)	two finger (horizontally placed)	tap (once)	*Point with two fingers and tap*	6
10		3 finger	three finger (horizontally placed)	long tap (3 seconds)	*Point and long tap*	6
11 12				bring it closer backwards	*Point and bring hand closer backwards*	15
13	Knuckle			tap (once)	*Point with Knuckle and tap*	5
14	Walk and point with index finger		Point		*Walk and point with index finger*	6
15	Kick			Feet touch	*kick (feet touch)*	11
16	Lean and grab	One hand			*Lean and grab*	6
17	Gaze		gaze	nil	*Gaze*	19
18				dwell gaze	*Dwell gaze*	7
19				leg tap	*Gaze and leg tap*	6
20				nod	*Gaze and nod*	4

21	Walk and grab with one hand		one hand		*Walk and grab with one hand*	11
22	Swipe hand uptill the target		one hand		*Swipe hand uptill the target*	8
23	Body touch				*Body touch*	9
						196

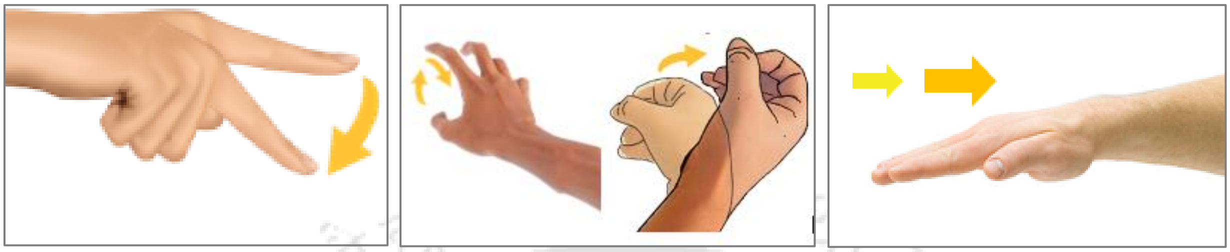
Table 2: Unique gesture details of VE 2: Dense VE, object placed at arms' length

	Unique gestures	Variation	Direct selection	Confirmation	Final Gesture	Freq
1	Point	thumb press on side for continuous zoom in	point	*tap (once)*	*Point and thumb press and point and tap*	5
2		Two hand index finger simultaneously		*tap (once) simultaneously*	*Point with two hands index finger and tap*	6
3		Knuckle and open knuckle to select	Point	*nil*	*Point using knuckle and open knuckle to select*	3
4		two fingers apart to cone cast	Point	*nil*	*point using two fingers to cast cone and point*	5
5	Pinch-out the VE and point & tap	Using index fingers		*tap (once)*	*Pinch-out the VE and point & tap*	29
		Using two hands join and apart	point	*tap with both hands*	*Pinch-out the VE using hands join and part and point & tap*	6
		all fingers joined	point	*tap with hand*	*Pinch-out the VE using all fingers and point & tap*	2
6	Extending hand and pull	One hand		*pull backwards*	*Extending hand and pull*	23
7	Palms joined together and separate to see miniature VE inside the palm of hands	Two hand	point	*nil*	*Palms joined together and separate to see miniature VE inside the palm of hands and point*	5
8	Gaze and head movement to remove outlier and gaze			*nil*	*Gaze and head movement to remove outlier and gaze*	2
9	Walk towards the target and point and tap	Legs and fingers	point	*Tap (once)*	*Walk towards the target and point & tap*	9
10	WIP to zoom VE	Two legs		*leg tap (once)*	*WIP to zoom and leg tap*	2
11	Raise hand and call gesture	One hand		*nil*	*Raise hand and call gesture*	2
12	Dwell gaze			*nil*	*Dwell gaze*	7

13	Tap with two finger to enact walking to zoom in VE and point	Fingers	point	*nil*	*Tap with two finger to enact walking to zoom in VE and point*	4
14	Walk towards the target and grab	Legs and one hand	grab	*nil*	*Walk towards the target and grab*	5
15	Bring two hands closer and closer vertically and horizontally to minimize the VE and pull hand to bring near and point	Hands and fingers	point	*nil*	*Bring two hands closer and closer vertically and horizontally to minimize the VE and pull hand to bring near and point*	4
16	Leg touch another sphere and kick towards the target	One leg		*nil*	*Leg touch another sphere and kick towards the target*	9
17	Grab and bring VE near (multiple times) and grab	both hands	grab	*nil*	*Grab and bring VE near (multiple times) and grab*	4
18	Hand shaped like O alphabet to zoom in specific area of VE and point	One hand and fingers	point	*nil*	*Hand shaped like O alphabet to zoom in specific area of VE and point*	6
19	Lean to zoom-in VE and point	One hand	point	*nil*	*Lean to zoom-in VE and point*	4
20	Walk and leg touch	Leg			*Walk and leg touch*	6
21	Lift leg and hand and tap	Leg and hand		*tap (once)*	*Lift leg and hand and tap*	8
22	Tap with one leg to zoom in VE	One Leg			*Tap with one leg to zoom in VE*	6
23	Swipe hand to remove outliers and tap	One hand		*tap (once)*	*Swipe hand to remove outliers and tap*	9
24	Point nearest outlier and hand swipe	Finger and hand		*nil*	*Point nearest outlier and hand swipe*	4
25	Join hands and open to view miniature VE between the hands and point	Two hands	point	*nil*	*Join hands and open to view miniature VE between the hands and point*	4
26	Leg swipe drag to bring VE near and point	Leg and hand	point	*nil*	*Leg swipe drag to bring VE near and point*	2
27	Make a fist to squeeze VE and point	Hands and fingers	point	*nil*	*Make a fist to squeeze VE and point*	2
28	Gaze and pinch	One hand	pinch	*nil*	*Gaze and pinch*	5
29	Scroll with the index finger to zoom and point	One hand	point		Scroll with the index finger to zoom and point	1

						194

3.2.7.1 Results of Gesture Extraction for VE1: Dense VE with Objects Placed within Arm's Reach



(a) (b) (c)

Figure 10: Visual representation of the extracted gestures for VE1: (a) Point and tap (b) Grab and rotate (c) Point & bring hand closer

The study identified 23 unique gestures for VE1, of which 7 were selected based on 1st and 2nd preferences. Three gestures, namely *point and tap*, *grab and rotate the wrist*, and *point and bring the hand closer to the body*, were preferred due to their higher frequency and scores of 26, 20, and 15 and the scores 33.56, 33.25 and 30.02 respectively. *Point and tap* gestures were inspired by traditional WIMP interfaces, and participants used their index and middle fingers to mimic a mouse click. One participant said, *"This feels similar to a mouse click and is easy to remember."* This finding is in line with Wu et al. (2019) in which traditional WIMP interfaces inspired the gestures collected for a VR shopping application. A few participants associated the number of fingers used with the size of the object for accurate selection. Similar findings were identified and reported by Wobbrock et al. (2009). A participant stated that, "*I will use two fingers if the target is bigger. Use of 2 fingers (to perform point and tap) enables faster selection and ensures accuracy."* Another participant said he used more fingers "*to ensure accurate more reliable selection."*

Grab and rotate the wrist gesture required more effort but provided a sense of control and was similar to a door-knob opening action. Thus, making it easy to remember. The degree of rotation varied from 60-150 degrees for the participants. The selection of the target in point and bringing the hand closer to the body gesture involved pointing towards the object, followed by bringing the hand closer to confirm the selection. The act of bringing the hand closer was seen as a metaphorical representation of bringing the object closer to the participant. Similar to the *point* *and tap* gesture, participants associated the size of the target with the number of fingers used for pointing. They used more fingers to point at larger objects, ensuring greater accuracy in selection.

The remaining gestures, including *grab*, *point*, *palm point & thrust*, and *dwell gaze*, were found to have a lower frequency and overall score. Participants expressed that *grab and point* gestures resulted in erroneous selections as they mistakenly selected other objects within their reach. Participants also felt the need to confirm the selection for smaller objects due to the difficulty in accurate selection. *Dwell gaze* was less preferred due to the problem of the Midas touch (Jacob, 1990), and a few participants felt uncomfortable performing this gesture as it felt like staring at something without any specific reason. Additionally, *palm point and thrust* gestures were found to be less natural than the extracted gestures.

Table 3: Gesture details and Overall details of VE1 considering 1st and 2nd preferences: Dense VE, object placed at arm's length.

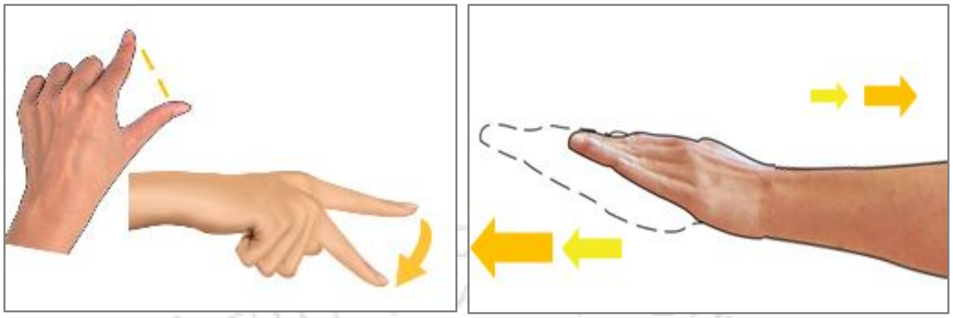
Sr no.	Unique gestures	Selection	Confirmation	F	1st pref.	2nd pref.	Posture	Overall score
1	*Point and tap**	Point	Tap	26	11.59	11.80	10.16	33.56
2	*Grab and rotate the wrist**	Grab	Clockwise wrist rotation	20	11.01	10.94	11.29	33.25
3	*Point and bring hand closer to the body**	Point	Bring hand closer to the body	15	10.05	9.79	10.16	30.02
4	*Palm point and thrust*	Palm point	Thrust	6	9.09	9.51	10.16	28.77
5	*Dwell gaze*	Gaze	Gaze for 3 sec	5	10.05	9.51	10.16	28.53
6	*Grab*	Grab	-	5	9.09	9.51	7.92	26.52
7	*Point*	Point	-	3	9.09	8.93	10.16	28.20

**Gestures chosen*

3.2.7.2 Results of Gesture Extraction for VE2: Dense VE with Objects Placed at Scaled Length

The study identified 29 unique body gestures for VE2, out of which 7 were selected based on 1st and 2nd preferences. *Pinch-out VE and point & tap* (Figure 11a) and *extending the hand*

and pull (Figure 11b) were chosen based on their frequency of 29 and 23 and an overall score of 34.74 and 31.47, respectively.



(a) **(b)**

Figure 11: Visual representation of the extracted gestures for VE2: (a) Pinch-out VE and point and tap (b) Extending a hand and pull hand closer to body.

Pinch-out VE and point and tap involved the use of two hands, with the non-dominant hand performing the pinch-out gesture to zoom in on the VE and bring the target closer. This was followed by *point and tap* with the dominant hand, which selected and confirmed the target. Participants found this gesture useful as it increased target visibility and allowed them to accurately select small and distant objects. A participant stated that "*The objects are small and dense. Pinching out (VE) allows me to expand the VE so that I can view the object clearly for selection"*. On the other hand, *extending the hand and pull* involved participants visualizing a virtual hand that stretched towards the target as they leaned forward, similar to the 'go-go technique' proposed by Poupyrev et al., 1996. Once the virtual hand touched the object, participants performed a pulling hand backward gesture to confirm the selection, thus enabling accurate object selection in a dense VE.

While participants also used walking gestures to reach the target, such as *Walk towards the target and point & tap*, *Walk towards the target and grab*, and *Walk-in place and point & tap*, they were less preferred due to the physical effort required to reach the target. Some participants were also concerned about safety while walking in a virtual environment. A participant stated, *"Although we may not place any physical object in the tracking area, what happens if I hit a passing human being. I prefer gestures with limited movement as they are safe". Pinching-out the VE and grabbing* were found to be less preferred as the grabbing gesture did not allow for confirming the selection. Participants believed that, despite zooming-in the VE, confirming the selection was necessary to avoid incorrect object selection. Similar to VE1, *dwell gaze* was perceived as uncomfortable and less preferred due to its longer duration.

Table 4: Gesture Details and Overall Score of VE2 Considering 1st and 2nd Preferences: Dense VE, Object Placed at Scaled Length.

Sr no	Unique gestures	Target reach	Selection	Confirm-ation	F	1st pref.	2nd pref.	Posture	Overall score
1	*Pinch-out the VE and point & tap**	Pinch-out using index finger and thumb	Point	Tap	29	11.70	11.68	11.35	34.74
2	*Extending hand and pull**	Extend hand (leaning forward)	Touch the target	Pull hand	23	11.15	11.07	9.25	31.47
3	*Walk towards the target and point & tap*	Walk till target	Point	Tap	9	9.50	10.05	10.30	29.86
4	*Dwell gaze*	-	Gaze	Gaze for 3 sec	7	9.68	9.44	6.09	25.23
5	*Walk towards the target and grab*	Walk till target	Grab	-	5	9.50	9.24	9.25	28.01
6	*Pinch-out the VE and grab*	Pinch-out using index finger and thumb	Grab	-	4	9.32	9.24	9.25	27.81
7	*Walk-in place and point & tap*	Walk-in place	Point	Tap	3	9.13	9.24	11.35	29.74

** Gestures chosen*

3.2.8 Discussion on the Findings of VE1 and VE2

The study's findings suggest that users prefer gestures that provide a sense of control, are easy to remember, and resemble familiar actions. In a total of 52 unique gestures, and 81 variations 39 gestures used the "confirmation" step (VE 1 had 45 variations and 28 confirmation gestures, and VE 2 had 36 variations and 11 confirmation gestures). In the case of gestures chosen as 1st and 2nd preferences, a total of 10 gestures incorporated the confirmation technique out of a total of 14 unique gestures for both VEs.

The study's findings suggest that users prefer gestures that provide a sense of control. We observed participants' preference to "confirm" before the target is selected. The "confirmation" step increased object selection accuracy and reduced errors, hence increasing participants' confidence. A participant stated, *"While it may increase selection time, confirmation helps accurately select a target which is located in a complex (dense) VE (in VE2). It increases my* *confidence to interact with such complex VR systems".* It is also observed in lower preferences of *grab and point* gestures that highlight the importance of confirmation gestures for accurate selection, especially when objects are small in size. In both VE1 and VE2, users preferred gestures that were easy to remember and resemble familiar actions.

In VE1, three gestures - *point and tap, grab and rotate the wrist,* and *point and bring the hand closer to the body* were preferred. These gestures were inspired by traditional WIMP interfaces, such as mouse clicks and knob rotations, and were found to be intuitive and easy to remember by the participants. Participants could easily recall the gesture and quickly perform it, resulting in a faster selection process. This is in line with previous research that has suggested that simple and easy-to-remember gestures are more effective and efficient in virtual environments (Bowman et al., 2001).

Similarly, in VE2, participants preferred the *pinch-out VE and point & tap* gesture as it was easy to remember and perform. The pinch-out gesture was already familiar to participants as it is commonly used in mobile devices for zooming in and out. Therefore, combining this with the *point and tap* gesture made it intuitive and easy to remember. The use of two hands for this gesture was not perceived as a barrier to its usability, as it increased the perceived accuracy and confidence in selecting distant objects among participants. In both VE1 and VE2, we found that participants preferred to use upper body gestures, particularly hand-based gestures, which accounted for 41 out of 52 unique gestures. The results are in line with the findings of Schulze et at., 2016 and Lin et al., 2019.

Of the remaining 11 unique gestures, three were leg-based (e.g., *lift the leg in the air and tap feet to select, walk and touch an object with the knee,* etc.), and eight used a combination of *leg and upper body gestures* (e.g., *walk and grab/point*). Leg-based gestures were predominantly used for selecting objects placed below the torso, with participants finding it easier than bending. A participant stated, "*The object is placed below the torso. Using my leg to select them is easier than bending"*. *Pointing* was the most commonly used gesture for object selection, with 14 out of 25 gestures selected as first or second preferences. This preference for pointing aligns with previous studies (Eriksson, 2016; Argelaguet and Andujar, 2013), which have found it to be a natural and intuitive gesture for users when interacting with virtual environments. As one participant put it, *"Pointing is one of the most natural ways to select an object, even in the real world. Hence, selecting an object in the virtual world using pointing comes very spontaneously."*

In VE 2, participants were observed using both hands for object selection, primarily for zooming in on the environment and bringing targets within reach. The non-dominant hand played the role of a "modifier," bringing targets closer and swiping to remove outliers, while the dominant hand (typically the right hand for right-handed participants) was used to select and confirm the target. This finding aligns with the principles outlined in Guiard, (1987) study on the asymmetric division of labor between hands during gestural tasks. Furthermore, participants in VE 2 reported visualizing a miniature version of the VE within the space created by joining their palms and spreading their hands apart. This technique provided a closer and clearer view of the dense environment, allowing for more effective target selection. Overall, the use of both hands in VE2 allowed for a more nuanced and precise interaction with the virtual environment.

3.3 Study 2: Evaluation of Extracted Finalized Gestures from VE1 and VE2

In order to further refine the gesture selection for VE1 and VE2, we conducted a second study. In this study, we focused on the three gestures that received the highest overall scores for VE1 and the two gestures that received the highest scores for VE2 in the first study. The goal of this study was to identify the most suitable gesture for each VE based on factors of ease of performance, body-part suitability, gesture appropriateness, user preference, and overall effort required. The details of the study, participants, procedure, data collection method, and results are explained in the following sections.

3.3.1 Participants

For this study, a new group of 33 participants was selected, consisting of 22 males and 11 females, all between the ages of 18-35 (mean=23.48, SD=4.40). All participants were right-handed university students with prior experience using VR for playing games or watching movies for at least 10 hours in the past six months, as well as experience using a Wii remote or Microsoft Kinect for playing games for at least 10 hours in the past six months. None of the participants had previously taken part in Study 1.

3.3.2 Study Procedure

The study began with a verbal introduction of the study objectives to each participant. Following this, a 10-15 minute familiarization session was conducted where the participants were shown the VEs using HMD-VR. After the familiarization session, the elucidated gestures from

Study 1 were introduced via a recorded video and a demonstration given by the moderator. The participants were then given a kinaesthetic priming session for further clarification. The sequence of VEs and gestures presented to each participant was counterbalanced to avoid any order effects. As in Study 1, participants were required to select a target colored red, placed among 924 distractors, using the given gesture. Once the target was selected, its color changed to green, and we used a Wizard of Oz technique to confirm their gesture performance and indicate their selection. Each participant was asked to perform each elucidated gesture five times on cue, after which they rated the ease of performing, body-part suitability, and gesture appropriateness using a 7-point Likert scale. They also ranked their preferences and physical effort for each gesture on a separate paper. The participants' decision-making process was further explored through discussions on each gesture, variable, and preference. The study was approximately 45 minutes long for each participant and was video recorded with prior consent for further analysis.

3.3.3 Data Collection Method

The participants were asked to rate the ease of performance, suitability, and appropriateness of each gesture on a 7-point Likert scale (-3 = low to 3 = high) after performing the gestures. They were also asked to rank their preference and effort for each gesture. The rating and ranking data were recorded on a printed sheet provided to the participants. Ease of performance, body-part suitability, and gesture appropriateness were stated as: *"the gesture is easy to perform," "the body part used is suitable for its intended purpose," and "the gesture is appropriate for its intended purpose*". Gesture appropriateness is defined as a gesture that is a "better match" or "fit for purpose" for its intended use. The participants ranked the gestures based on their preference and effort using the statements "rank the gesture according to your highest preference" and "rank the gesture that requires the least amount of effort." With the participants' consent, the study was video-recorded for further analysis and presentation of the findings.

3.3.4 Results

To test for normality, we applied the Shapiro-Wilk test to the data, which showed that the data followed a normal distribution. A one-way analysis of variance (ANOVA) was conducted, followed by post hoc t-tests with Bonferroni correction (corrected significance level = significance level * 3) and paired-sample t-tests to determine if there were significant differences between the gestures.

3.3.4.1 Results of VE 1: Dense VE with Objects Placed within Arms' Reach

Ease of performing: For ease of performing, we conducted a one-way ANOVA to test for statistical significance among the three gestures in VE1. The results showed a significant difference ($F(2,96)=9.75$, $p=0.0001$). A post-hoc t-test with Bonferroni adjustment revealed that the point and tap gesture (Mean=2.92, SD=0.90) was significantly easier to perform than the point and bring hand closer gesture (Mean= 1.48, SD=0.93, $p=0.0001$) and the grab and rotate the wrist gesture (Mean= 2.18, SD=0.84, $p=0.02$).

Body-part suitability: We conducted a one-way ANOVA to assess the statistical significance of body-part suitability across the three gestures in VE1 ($F(2,96)=11.68$, $p=0.0001$). Post-hoc t-tests indicated that the point and tap gesture (Mean=2.42, SD=0.90) was significantly more suitable than the point and bring hand closer gesture (Mean= 1.48, SD=0.93, $p=0.0001$) and the grab and rotate gesture (Mean= 2.18, SD=0.84, $p=0.002$).

Gesture appropriateness: A one-way ANOVA was performed to determine the statistical significance in gesture appropriateness for all three gestures, which revealed significant differences between them ($F(2,96)=5.32$, $p=0.0006$). Post-hoc t-tests showed that the point and tap gesture (Mean=2.52, SD=0.96) was significantly more appropriate than the point and bring hand closer gesture (Mean=1.27, SD=1.23, $p=0.02$) and the grab and rotate gesture (Mean=1.87, SD=1.05, $p=0.002$).

Preference and Effort: Among the 33 participants, 21 preferred the point and tap gesture, while 9 and 3 participants preferred the grab and rotate and point and bring hand closer gestures, respectively. Further, 24 out of 33 participants rated the point and tap gesture as requiring the least effort, while 7 and 2 participants rated the grab and rotate and point and bring hand closer gestures, respectively. These results are presented in Figure 12, displays the outcomes of VE1 in the second study.

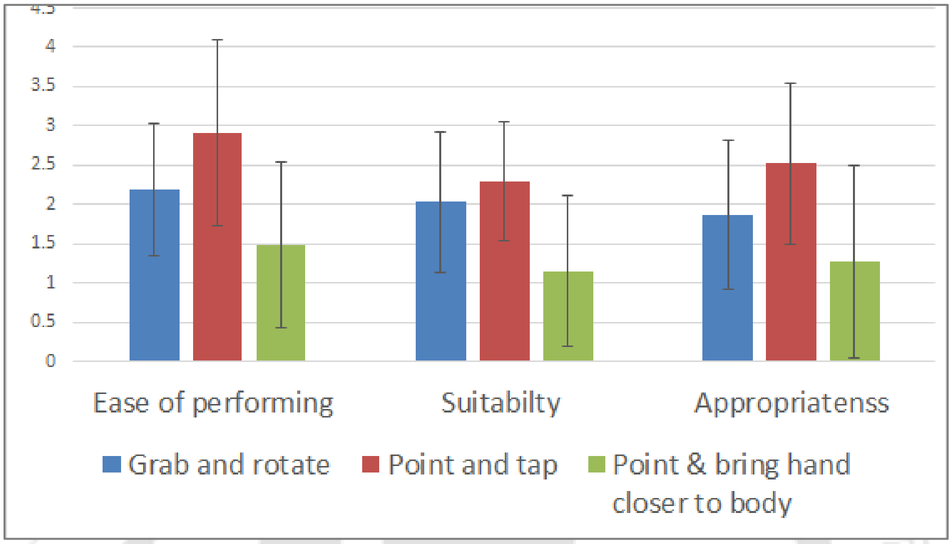


Figure 12: The mean and SD values and error bars for ease of performing, body-part suitability and gesture appropriateness for finalized gestures of VE1.

3.3.4.2 Results of VE 2: Dense VE with Objects Placed at Scaled Length

Ease of Performing. A paired-sample t-test was conducted to examine the ease of performing for the two gestures in VE2. The results revealed a statistically significant difference (p=0.002) between the mean ease of performing for the pinch-out the VE and point and tap gesture (Mean= 1.93, SD=0.93) and the extending the hand and pull gesture (Mean= 1.48, SD=0.93). Specifically, participants found the pinch-out VE and point and tap gesture significantly easier to perform than the extending the hand and pull gesture.

Body-part Suitability. A paired-sample t-test was conducted to determine the statistical significance of the body-part suitability between the two gestures. Results showed that the pinch-out the VE and point and tap gesture (Mean= 1.84, SD=0.79) was significantly more suitable (p=0.003) compared to the extending the hand and pull gesture (Mean= 1.15, SD=1.03).

Gesture Appropriateness. Based on a paired-sample t-test, a significant difference was found in the appropriateness of the two gestures. Specifically, extending the hand and pull gesture (Mean= 1.27, SD=1.2) was found to be less appropriate compared to pinch-out the VE and point and tap gesture (Mean=1.9, SD=0.87) with a p-value of 0.018.

Preference and Effort: In the second study, 21 out of 33 participants preferred the pinch-out the VE and point and tap gesture over the extending the hand and pull gesture, which was preferred by 12 participants. However, participants found extending the hand and pull gesture less effortful (18 participants) compared to pinch-out the VE and point and tap gesture (15 participants), primarily due to the involvement of two hands in performing the latter gesture. Figure 13 presents the results of ease of performing, body-part suitability, and gesture appropriateness for VE2.

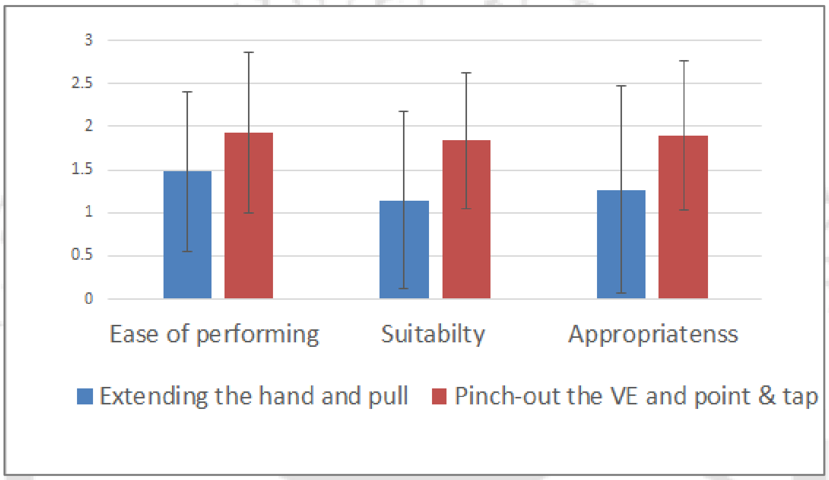


Figure 13: The mean and SD values and error bars for ease of performing, body-part suitability and gesture appropriateness for finalized gestures of VE2.

3.3.5 Discussion of Results of Study 2

Our study on VE1 showed that the *point and tap* gesture was significantly easier to perform, more suitable for body parts, and more appropriate compared to the other two gestures. Additionally, it was the most preferred and least effortful gesture among the participants. The simplicity and familiarity of the point and tap gesture, which resembles the point-and-click interaction in WIMP interfaces, made it easy for participants to remember and adopt. One participant noted that this gesture was easy to remember as it felt like performing point and click using a mouse on a desktop computer. This finding is consistent with prior research, such as Vogel and Balakrishnan (2005), Wu et al. (2013), and Wu et al. (2019), which showed that participants tend to propose gestures similar to those used in WIMP and touch-based interfaces due to their prior experience and preference.

Participants had concerns about the appropriateness and ease of performing *grab and rotate* and *point and bring hand closer* gestures. For the *grab and rotate* gesture, participants were uncertain about the appropriate amount of force and rotation needed for selection and expressed concerns about potential wrist fatigue if selecting multiple objects. A participant said, *"It feels that if I rotate slowly the selection will not confirm. It feels that I have to rotate very quickly and with lot of force. This might also lead to wrist fatigue if I select multiple objects."* Similarly, for *point and bring hand closer* gesture, participants were uncertain about the appropriate distance to bring their hand for selection, and how quickly to bring their hand for final confirmation. Participants also reported potential arm fatigue from bringing their hands close to their body multiple times. Consequently, these two gestures received lower scores for ease of performing, gesture suitability, and appropriateness. These findings are consistent with prior research indicating that complex or ambiguous gestures may be less effective than more straightforward and intuitive gestures for selection tasks.

In VE2, we found that *pinch-out the VE and point and tap* gesture was significantly easier to perform, more suitable, and more appropriate than extending the hand and pull gesture. Despite participants experiencing increased physical effort, they still preferred *pinch-out the VE and point and tap* gesture. This was because this gesture brought the target object within the participants' arm's reach, providing an unobstructed and clear view and enabling accurate and error-free object selection. One participant explained, *"It brings the target in my area and provides me a clear view. This makes my task easy and accurate, and I will make fewer mistakes".* This finding is different from the study conducted by Wu and Wang (2013), where participants preferred to bring the target to their arms' reach for distant objects. They felt that it allowed easy access to the target and provided accurate object selection. For example, one participant stated that, *"I would want to first bring the objects within my (hand's) access, so I know which object I am selecting. This helps me select the object accurately and confidently".*

In line with the findings of study 1, participants in study 2 also showed a preference for gestures that align with established mental models such as touch and WIMP interfaces. Specifically, participants preferred the *pinch-out gesture* to reach distant objects due to its ease of performing and relevance to touch interfaces. This finding is consistent with the idea of leveraging prior experience with technology to improve interaction in novel interfaces. Participants also mentioned that the *pinch-out* gesture provided an appropriate scaling factor to magnify the target object. Similarly, participants preferred the *point and tap* gesture due to its similarity to mouse clicks on desktop interfaces. One participant noted, *"I use pinch-in/out to zoom in/out images on my mobile phone. Hence, it is easy for me to remember pinch-in/out."* These observations underscore the importance of leveraging users' prior experience and familiarity with established interaction models to design intuitive and user-friendly gesture-based interfaces. Participants recommended the use of varying finger numbers for performing point and tap gestures in proportion to the size of the target object. This recommendation stemmed from the participants' belief that using multiple fingers would increase the accuracy of selection for bigger targets. Additionally, participants suggested that VR systems should provide users with the flexibility to modify their gestures based on the size of the target object. One participant commented, *"It would be more intuitive to use two or three fingers if the target is bigger. This would also result in more accurate selections compared to just using the index finger. VR systems should offer such variations to their users."* Similarly, participants suggested using multiple fingers to increase the magnification levels while performing pinch-out gestures. This finding is consistent with the results of a study by Ortega (2019), which observed that users generate different gestures based on object sizes. The participants' recommendations for gesture variations in relation to object size highlight the importance of providing users with customizable gesture options.

In our study, we found that participants prioritized accuracy over effort, particularly for selecting small objects in dense virtual environments (VEs). Despite requiring higher physical effort, participants preferred using *pinch-out the VE and point and tap* gestures over other techniques because they believed these gestures improved selection accuracy and reduced errors. Furthermore, participants preferred manipulating the VE to decrease the distance to the target and provide a clear view, in contrast to techniques that emphasized screen visualization of the selection tool rather than VE manipulation. Interestingly, six participants suggested variations of the *pinch-out* gesture that involved using multiple fingers for easy and quick magnification of the VE. Participants found that using more fingers while zooming the VE increased the speed of the process. For example, using three or five fingers to perform pinch-out magnified the VE three times or five times, respectively, making it easier to reach the target. Participants found the finger-based VE magnification process particularly easy for dense VEs, such as those used for visualizing and analyzing molecular structures or 3D microscopy.

Overall, the results of Study 2 showed that participants preferred using the *point and tap* gesture over the *grab and rotate* and *point and bring hand closer* gestures for object selection in virtual environments (VEs). Similarly, they preferred the *pinch-out VE* and *point and tap gesture* over *extending the hand and pull* gesture. Participants prioritized accuracy over effort and believed these gestures improved selection accuracy and reduced errors. Participants also recommended using varying finger numbers for performing point and tap gestures and using multiple fingers for easy and quick magnification of the VE. The findings highlight the importance of leveraging users' prior experience and familiarity with established interaction models to design intuitive and user-friendly gesture-based interfaces and provide users with customizable gesture options.

3.4 Classification of Gestures From Study 1

In our study, we conducted a gesture classification of all the identified gestures from Study 1. This classification is crucial in understanding the reasons behind the gestures performed, the cognitive processes involved, the types of gestures identified, and how they are differentiated or grouped. It is also valuable in the field of human-computer interaction (HCI) research as it assists researchers in determining appropriate gestures based on user behavior, mental models, and constraints. Moreover, a well-established classification of gestures can aid HCI researchers in gaining a better understanding of different perspectives on gestural interaction, as emphasized by Zhang et al. (2019).

To classify the gestures, we analyzed the 390 gestures performed in Study 1, which included 52 unique gestures with 23 and 29 unique gestures for VE1 and VE2, respectively. We relied on the taxonomies proposed by Karam and Schraefel (2005) due to their comprehensive coverage of taxonomies that encompass most of the taxonomies presented in the literature, particularly those that focus on full-body and in-air gestures.

Our analysis showed that there was a higher preference for upper-body gestures (87%) compared to lower-body (12%) and full-body gestures. Among the upper-body gestures, hand-based gestures (76%) were the most common for both VEs. Further, the gesture classification revealed that manipulative and deictic gestures (14.3%) were dominant, followed by pantomimic (10.6%), deictic (10.6%), deictic and pantomimic (9.5%), and semaphoric (8.3%) gestures. We provide more details on the VE-specific classifications in the subsequent sections

3.4.1 Gestures Classification for VE 1

Our analysis of gesture preference for VE1 revealed a higher preference for *pantomimic gestures* (12 out of 23 unique gestures), including the combination of *deictic and pantomimic* gestures. *Pantomimic gestures* allowed participants to perform a static pose or a dynamic movement, or a combination thereof, for object selection. These included actions such as *grab* (to identify) and *rotate* (to confirm), *grab with curved fingers and open palm*, *make a fist* (static pose), punch (to identify) and *wrist rotate* (to confirm), and *swipe a hand over the target*. The larger target size may explain the higher preference for pantomimic gestures, as participants matched their hand size with the target size, and could shrink and enlarge the curved fingers according to the target size (i.e., curving the finger to reduce the grabbing area for the grab and rotate gesture). One participant noted, *"I can map the target size to the size of my curved fingers. It gives confidence that I have held the target properly for selection."* Additionally, we observed the use of pantomimic gestures in combination with gaze-based interactions, such as gaze and grab, gaze and nod, and gaze and blink.

In addition, we noticed that participants often combined pantomimic gestures with deictic gestures, such as pointing and bringing the hand closer to the body, palm pointing near the body and thrusting, and pointing and drawing a circle around the target. These combined gestures ensured that the selection was verified by using a confirmation gesture to confirm the correct selection. Interestingly, we observed that participants tended to use a two-stage strategy when performing these deictic and pantomimic gestures. Specifically, participants used deictic gestures to indicate the target, followed by the use of pantomimic gestures to confirm the object selection.

The classification of gestures revealed that *semaphoric gestures* (3 out of 23) predominantly consisted of dynamic semaphores, such as the call gesture using a vertical palm, knuckle tap, flip hand, and knee touch and tap with a leg. While *semaphoric gestures* are not typically considered natural (according to Karam and Schraefel, (2005) taxonomy), we speculate that they were employed primarily due to the eyes-free nature of the interactions.

The minority of gestures identified in our study did not fall into the categories proposed by Karam and Schraefel (2005) and Zhang (2019). These included gestures such as "dwell gaze" and "blowing the object." Although no clear pattern was observed among these gestures, some were found to be derived from low-cost virtual reality interactions, such as the "gaze" gesture.

3.4.2 Gestures Classification for VE 2

In VE 2, our observations revealed the use of several gesture categories, including manipulative and deictic (5 out of 29), pantomimic (3 out of 29), deictic and semaphoric (4 out of

29), and deictic (2 out of 29) gestures. We also observed the use of combined gesture categories, such as the combination of pantomimic and deictic gestures, manipulative, pantomimic, deictic and semaphoric gestures, others and manipulative gestures, and pantomimic, deictic, and semaphoric gestures.

In our study, we observed a combination of various gesture categories used by participants in a virtual environment that presented challenges related to distant objects placed in a crowded space. The need to bring target objects into visible space to ensure accurate and error-free object selection was the primary reason behind this combination of gestures. *Manipulative and deictic* gestures were commonly performed as participants preferred to manipulate the environment to bring objects closer. Although manipulation required increased physical effort, participants preferred this approach for precise selection. Examples of such *manipulative and deictic* gestures include *grab and bring environment near and point*; *scroll with the index finger to zoom and point*; *pinch-out VE and point and tap*, among others. This approach was also observed for gestures performed under the *deictic and semaphoric*, *pantomimic*, *deictic and semaphoric*, and *manipulative*; *pantomimic*, *deictic,* and *semaphoric* categories.

3.5 Proposal of Gesture Taxonomy in VR

In our study, we have developed a gesture taxonomy based on the analysis of the data collected in Study 1. Our taxonomy is not only based on existing gesture taxonomies but also introduces two new categories. The new categories are based on *hand dominance,* namely *dominant hand only* and *non-dominant hand first,* and *multi-body part movement gestures,* which include both *simultaneous* and *sequential* movements. We anticipate that this proposed classification and taxonomy will aid designers and researchers in creating more efficient and effective full-body gestural interfaces, particularly for VR interfaces.

3.5.1 Introduction to Gesture Taxonomy

Gestures are a popular modality in Human-Computer Interaction (HCI), but their diversity and complexity require a well-structured classification (Aigner et al., 2012, Vafaei et al., 2013). Such a classification can aid gesture designers, practitioners, and researchers in selecting appropriate gestures that suit their design purposes while considering user behavior (Wobbrock et al., 2009), mental models (Poggi, 2001; Zhang, 2019), and constraints. Furthermore, a well-established taxonomy can enhance the understanding of different perspectives in gestural interaction research. Gesture classification and information of its user behavior and mental model have supported designing surface computing and motion gestures (Ruiz et al., 2011, Wobbrock et al., 2009). Existing gesture classification research in HCI covers a broad range of dimensionalities, but does not accommodate gesture classification studies conducted for VR platforms or full-body gestures in VR.

The field of HCI has seen several recent studies proposing classifications of hand gestures, such as the taxonomy presented by Karam and Schraefel, (2005). Another example is the work of Wobbrock et al. in 2009, who focused on hand gestures for surface computing. Ruiz et al. in 2011 proposed a taxonomy with two broad dimensions, gesture mapping and physical characteristics. To accommodate the dimensionalities of gestures in Augmented Reality (AR), Piumsomboon in 2013 developed two additional dimensions, symmetry and locale. However, while these classifications cover a broad range of gesture taxonomies, they do not account for classifications of gestures specific to virtual reality (VR) platforms, nor do they consider full-body gesture classification, including VR. Therefore, we present a new taxonomy that includes these dimensions to assist designers and researchers in creating effective full-body gestural interfaces for VR applications.

3.5.2 Proposal of Gesture Taxonomy

In this study, we introduce gesture taxonomies that focus on whole-body gestures performed in VR interfaces. We classify 52 unique gestures identified from an elicitation study and propose two new taxonomies. Our work has two main contributions: (i) the analysis and classification of a user-defined set of gestures, and (ii) the proposal of two novel gesture taxonomies. This research attempts to fills a gap in the existing literature by providing a comprehensive classification of gestures specifically for VR platforms.

We present two gesture taxonomies based on 52 unique gestures identified from study 1. Gestures were classified based on hand dominance and motion performed by hands, as well as posture and sequence of gestures. The proposed taxonomies include *hand dominance* and *multiple body-part movement* gestures, each with sub-categories outlined in table 5 and described below. Our work is focused on full-body gestures, and therefore, facial and micro expressions are not included in the taxonomy.

We developed the first dimension, called *hand dominance*, to classify gestures based on whether they were performed using the dominant or non-dominant hand. We categorized the gestures into three sub-categories: *dominant hand only*, *non-dominant hand first*, and *equal hand dominance.*

Table 5: Taxonomy of Controller-less Gestures in Virtual Reality

				Definition
Hand Dominance	Dominant hand only	Single movement gestures	Static hand pose	A gesture performed by the dominant hand and has a single movement. Hand pose is held static. i.e., *Extend hand and point*
			Dynamic hand pose	The gesture performed by the dominant hand and having multiple movements. Hand pose is dynamic. i.e., *Point and bring hand closer to body*
		Multi-movement gestures		The gesture performed by a dominant hand that has multiple movements i.e., *pinch-out VE and point and tap*
	Non-dominant hand first			A gesture performed by a non-dominant hand (to perform a secondary task) followed by a dominant hand (to perform a primary task). i.e., *pinch-out and rotate the VE (non-dominant hand) and point and tap (dominant hand)*
	Equal hand dominance			A gesture that involves simultaneous movement of both hands. No specific preference is given to any hand for the task. i.e., *point (with the index finger of both hands) and tap*
Multi-body part	Simultaneous			Gestures that occur simultaneously to complete a task. i.e., *lift leg and hand together and tap*

				Definition
movement gestures	Sequential			Gestures that involve a definite sequence to complete a task. i.e., *walk (to the target object) and rotate the environment and point and tap*

The category of *dominant hand-only* gestures comprises 40 out of 52 unique gestures, which accounts for 36% of all gestures observed in both VEs. This category can be further divided into *single-movement* and *multi-movement* gestures, which require one or multiple hand movements by the dominant hand, respectively. Single-movement dominant hand-only gestures can have a *static* (e.g., extend hand and point) or *dynamic* (e.g., point and bring hand closer to the body) final hand pose. This category constitutes 39% (16 out of 40 gestures) of dominant hand-only gestures. *Multi-movement gestures* involve more than one movement performed by the dominant hand (e.g., *point and bring hand closer to the body*, *palm point and thrust*). Although there is no set limit to the number of hand movements, our observations suggest that participants tend to perform a maximum of 3-4 dominant hand movements, as additional movements can increase physical effort. The *dominant hand-only* gestures are useful in environments where the non-dominant hand is responsible for performing secondary tasks, such as virtual navigation with the non-dominant hand while using the dominant hand to pick up virtual shopping items in a VR shopping mall.

We introduce a taxonomy of gestures performed using the *non-dominant hand first*, as part of the hand dominance category. In this taxonomy, the gesture is initiated with the non-dominant hand and completed with the dominant hand. The non-dominant hand can be used for secondary or supporting activities to accomplish the task. For instance, the *pinch and rotate gesture* is performed to enhance visibility and access to the target object by pinching the VE with the non-dominant hand and rotating the VE with dominant hand. *Point and tap* is performed to select the target object accurately and performed by pointing with dominant hand and tapping with dominant hand.

The third taxonomy in our proposed *hand dominance* category is equal hand dominance. The taxonomy involves giving equal importance to both hands in performing the primary task. For example, *pointing with the index finger of both hands and tap* to select uses both hands to perform the object selection independent of hand preference and dominance.

We present the taxonomy of multi-body part movement gestures, which involve the use of more than one body part to perform the gesture. This taxonomy is further divided into two categories: *simultaneous* and *sequential multi-body part movements*, based on whether the body parts move simultaneously or sequentially to achieve the task. For instance, some gestures such as "*walk towards the object, rotate the environment and point and tap"* require multiple body movements (i.e., legs and hands) in which walking, rotation and pointing that are performed sequentially. On the other hand, *simultaneous multi-body part gestures* are performed in unison to confirm the selection of the target object, such as *lifting the leg and hand together and tapping*, *gazing and pointing*, or *nodding and tapping.*

3.6 Chapter Summary

In this chapter, we presented a gesture elicitation study aimed at identifying natural and intuitive gestures for small object selection in dense virtual environments (VEs), a task that has not been thoroughly investigated in the context of HMD VR. This study consisted of two stages, where the objective was to collect the most suitable and user-friendly gestures for object selection in two types of dense VEs: (i) VE where targets are small and within arm's reach, and (ii) VE where targets are small and kept at a distance. In Study 1, we collected 160 gestures and their frequencies, of which 23 and 29 unique gestures for VE1 and VE2, respectively, were analyzed based on an overall score. We finalized 3 and 2 unique gestures for VE1 and VE2, respectively. Study 2 evaluated the finalized gestures based on ease of performing, suitability, appropriateness, effort, and user preference, and one gesture was finalized for each VE. The *point and tap* gesture was selected for VE1, while *pinch in/out the VE and point and tap* gesture were finalized for VE2.

We also presented a categorization of all collected gestures, which revealed a preference for upper body gestures, mainly finger and arm-based, and a focus on confirmation. The findings also underscore the importance of leveraging users' prior experience and familiarity with established interaction models such as touch and WIMP interfaces, a preference for bringing targets within hand reach for accurate selection, precise selection over physical effort, and large-scale manipulation of the VE. We also observed the use of varying finger numbers for performing gestures in proportion to the size of the target object.

Furthermore, we proposed two new gesture taxonomies based on hand dominance and the involvement of body parts in performing the gestures. These taxonomies were identified after analyzing 52 unique gestures performed during the elicitation study. We proposed three subcategories of the dominant hand-only taxonomy, namely single and multiple hand movements to indicate one and multiple hand movements in accomplishing a task. We also proposed multi-body part movement gestures, where a combination of different body parts performed sequentially or simultaneously to accomplish a task. These new taxonomies contribute to the existing literature on gesture classifications, which is essential in understanding user behaviors, mental models, and user performances for different tasks and technology platforms.

However, some limitations of the study should be acknowledged. First, the study only included young adults, and the elicited gestures may not be applicable to users of different age groups with no prior VR experience. Second, the study was limited to right-handed users who consider their right hand as dominant hand. The elucidated gestures may differ if hand dependency is eliminated. Third, the proposed gestures were not aided by a functional gesture recognition system, which may have restricted the gestures that participants could propose.

The next step is to evaluate the efficiency of the finalized gestures compared to existing object selection techniques, which will be presented in the subsequent chapter.

Chapter 4

4. Design and Evaluation of Locked Dwell Time-Based Point and Tap for Small Object Selection Within Arm's Length in Dense Environment

Chapter 3 reports on a user-generated gesture elicitation study that identified natural and intuitive gestures for small object selection in two types of dense VEs. The study produced unique gestures that were subjectively evaluated to finalize a single gesture for each VE. The *point and tap* gesture were selected as the most suitable and appropriate gesture for small object selection within arm's reach in dense (VE1).

This chapter presents a comparative study of the finalized point and tap gesture with existing gestures identified in the literature. A new technique, Locked Dwell Time-Based Point and Tap (*LDTPT*), was designed to select small objects within arm's reach in a dense VE. We first present the design rationale for *LDTPT* technique in section 4.1. We then present a walkthrough of the *LDTPT* technique in section 4.2. Section 4.3 presents a user study to investigate the efficiency of technique with techniques in literature. The objective of the study, baseline techniques, design of VE, participants, study procedure, study tasks and data collection method are presented in section 4.3.1-4.3.8. Section 4.4 presents the results of the study in terms of task completion time (section 4.4.1), error rate (section 4.4.2), easy to use (section 4.4.3), easy to learn (section 4.4.4), naturalness (section 4.4.5), preference and effort (section 4.4.6). The findings and discussions of the study is elaborated in section 4.5. The chapter is summarized in section 4.6.

4.1 Design Rationale: Point and Tap Gesture

The design of the *Locked Dwell Time-based Point and Tap (LDTPT)* gesture was motivated by the challenges encountered in selecting small objects within a dense virtual environment (VE1) using the *point and tap* gesture. Hand jitter often leads to accidental selection errors, causing fatigue and extended acquisition time, thereby negatively impacting user performance and efficiency (Cockburn and Firth, 2004). To address these issues, we redesigned the point and tap gesture by introducing a dwell time of 600 ms to acquire the target object, based on previous studies (Tomfelde, 2007) that suggest feedback delay times of 500-1000 ms for dwell-based pointing actions. This delay ensures that intentional binding between action and effect is maintained, thus supporting the user's sense of control over the environment. Additionally, a lock time threshold of 1000ms is provided to enable users to confirm their selection by tapping anywhere within this period, which can be extended if the user keeps pointing at the target. The lock time threshold also mitigates the "Midas Touch effect" (Penkar et al., 2012) commonly observed in dense environments. During pointing, a blue sphere visualization is created near the fingertip to aid in target selection. At the same time, the padding area of the index finger is increased or decreased depending on the number of fingers used for selection (ranging from 1cm to 4 cm). By combining pointing, pointing with multiple fingers, visualization aide, and locking the target before tapping, we believe the *LDTPT* gesture will effectively reduce accidental selection errors, reduce acquisition time, and enable precise selection of small targets within dense VEs.

4.2 Walkthrough: Locked Dwell Time-Based Point and Tap (LDTPT)

The *LDTPT* technique comprises four distinct phases users use to select a target object. In the first phase, users position themselves and their hand to point at the object they want to select. As the user points, the target object changes color from red to yellow to indicate it is being pointed at. In the second phase, users remain still for 600 ms, known as the Dwell Time (DT), while pointing at the target. During this phase, the target object changes color from yellow to green, and audio feedback confirms that the selection is correct. The third phase, which is the focus of this work, is the Locked Dwell Time (LDT) phase that lasts for 1 second. During this phase, the target remains locked and the user can confirm their selection by tapping anywhere on the interface within the given time frame. Audio feedback confirms the selection, and the target object changes color from green to blue. If the user continues to point at the target without tapping, the LDT phase is extended to allow more time for the user to confirm their selection. Figure 14 provides a clear depiction of the temporal structure of the *LDTPT* gesture. The significance of this technique lies in its ability to reduce accidental selection errors, improve target acquisition time, and enhance the precision of selection for small targets in highly dense virtual environments.

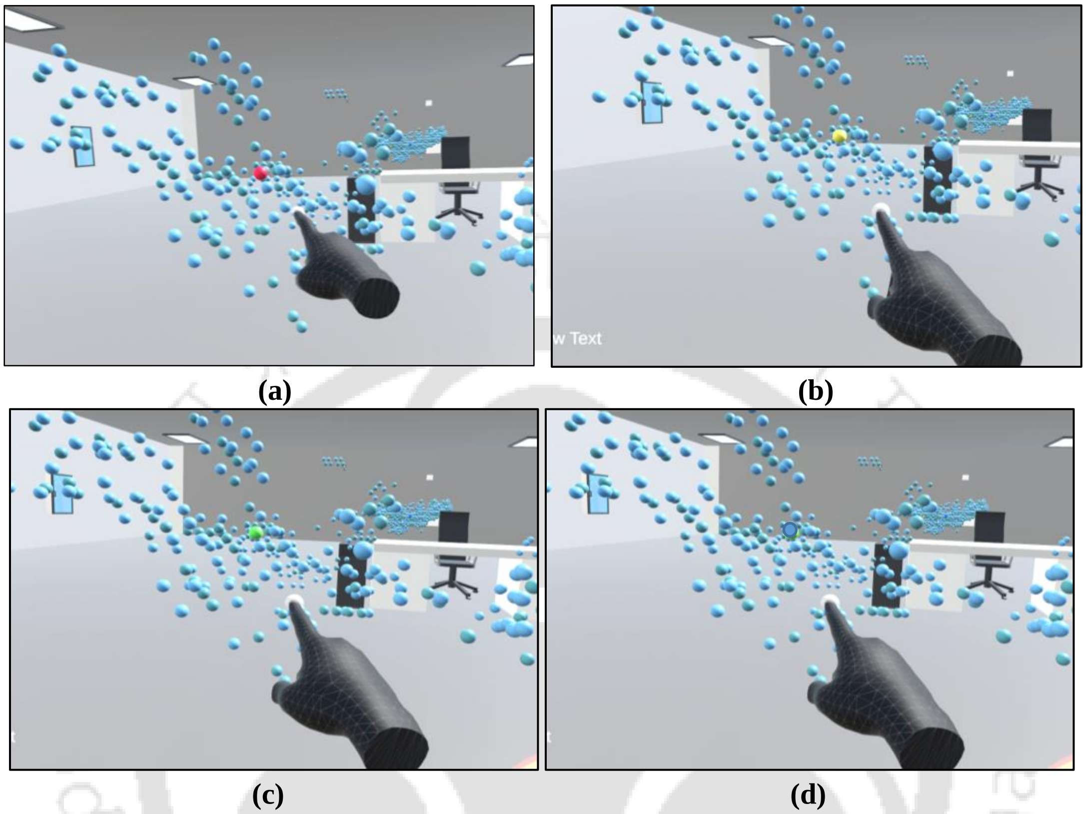


Figure 14: Selection of small object using *LDTPT* technique. (a) Selection of small object using one finger pad area. (b) The target object changes from red to yellow. (c) The target object changes its color from yellow to green and is in a locked phase (d) User taps anywhere on the environment to confirm the selection. The target object changes its colour from green to blue.

4.3 Evaluation of LDTPT

The objective of the study is to investigate errors, accuracy, and target acquisition time of object selection in dense VE for small objects placed within users' arm's reach.

4.3.1 Research Hypothesis

For this experiment, we formed four hypotheses:

H1: It is hypothesized that the LDTPT technique will result in faster selection times than magnetic grasp, raycasting and pinch techniques when selecting small and distant targets in a dense virtual environment.

H2: It is hypothesized that the LDTPT technique will result in a lower error rate than magnetic grasp, raycasting and pinch techniques when selecting small and distant targets in a dense virtual environment.

H3: It is hypothesized that the LDTPT technique will be perceived as easier to learn and use compared to magnetic grasp, raycasting and pinch techniques when selecting small and distant targets in a dense virtual environment.

H4: It is hypothesized that the LDTPT technique will be perceived as more natural compared to magnetic grasp, raycasting and pinch techniques when selecting small and distant targets in a dense virtual environment.

4.3.2 Baseline techniques:

In our study, we conducted a comparison between the Locked Dwell Time-based Point and Tap (*LDTPT*) technique and three direct selection techniques identified from existing literature: Magnetic grasp (Lin et al., 2016), Innate Pinch (Lin et al., 2016), and Raycasting (Mine et al., 1997). We selected these baseline techniques based on their direct selection interaction within arms' reach, naturalness, and ability to accurately select nearby objects. The first baseline technique, innate pinch, involves the user grasping an object with a pinch gesture using their index finger and thumb. Pinch gesture detection is performed by finding the closest distance between the tip of the index finger and the tip of the thumb, with a threshold of 1.5 cm. The target object must be positioned within the index finger and thumb for successful pinch selection, indicated by highlighting the object in yellow. Release from the pinch gesture is determined when the closest distance between the thumb and index finger is greater than the release trigger distance. The second baseline technique, Magnetic grasp, that treats each finger as a magnet to attract objects within a grabbing object distance of 10 cm. The nearest object within the grabbing object distance is highlighted to indicate selection. The third baseline technique, Raycasting, involves the user casting a blue ray and directing their hand toward the target object for accurate selection.

4.3.3 Design of VE

For our experiment, we designed a VE that consisted of 924 spheres randomly positioned within the VE. All of the distractor spheres were grey and measured 4 mm in size. The distance between the participant and the nearest object was set to 0.4 m. The target sphere was colored red, while the distractors were colored blue. Visual feedback was provided to the participants by changing the color of the selected sphere from red to yellow and then from yellow to green, depending on the selection technique they used. After selecting the target sphere, its position within the VE was randomly changed. A virtual hand representation was presented to the participants, and finger and arm-based gestures were detected using a Leap Motion device. We created the VE using the Unity 3D game engine.

4.3.4 Experimental Setup

For the experiment, we utilized an Oculus Rift HMD VR system, which was connected to a computer with the following specifications: an i7-8700 quad-core processor, an Nvidia Geforce 1060 GPU, 8GB RAM, and the Microsoft Windows 10 operating system. We used a Leap Motion device that was mounted on the Oculus Rift and faced upwards to detect finger-based gestures. Before beginning the experiment, we instructed the participants to stand with their hands kept straight at their sides. Additionally, we placed a horizontal video camera to capture the participants' movements during the experiment. Figure 15 provides a visual representation of the experimental setup

Figure 15: Experimental setup used for the study. The user is wearing Oculus Rift headset with a mounted Leap motion device. Camera placed diagonal to user took the photograph.

4.3.5 Study Participants

For this study, we recruited a sample of 40 participants (21 male and 19 female) between the ages of 18 and 30 (Mean = 24.2, SD = 3.02) who were enrolled as students at a university. To ensure consistency in the experiment, we only included right-handed participants. In terms of gaming experience, we selected participants who reported having at least 10 hours of experience using a Wii remote or Microsoft Kinect for gaming purposes within the last six months. Additionally, all participants reported having an average of 20 hours of experience playing games in VR, such as Robocop or Space Pirate, within the past year. All participants in this study were unpaid volunteers who provided informed consent before participation.

4.3.6 Study Procedure

Prior to conducting the experiment, we provided an explanation of each of the techniques to the participants. Subsequently, we provided training to allow them to become acclimated to the environment and the technique being tested. The training environment mimicked the actual environment, and the participants could select an infinite number of targets. Audio feedback signaling correct and erroneous selections were provided to them during the training, and visual feedback for each technique was also explained. The training was performed under easy conditions to aid in comprehending the technique without making mistakes. The training session lasted for approximately 15-20 minutes. Following this, the participants were directed to perform the task described in the subsequent section. After finishing the tasks for each technique and with varying sizes, the participants were asked to complete a questionnaire. The order of techniques was randomized for each participant to avoid biased results. Subsequently, the participants expressed their preferences and rated their efforts using the selection method. The user evaluation sessions lasted approximately 45 minutes for each participant.

4.3.7 Study Tasks

The study included five object selection tasks that participants were required to complete. Upon wearing the HMD headset, participants tapped on an OK button (red color) in the VE to initiate each task. The VE was filled with distractor objects. This procedure was employed to maintain a consistent angular amplitude of the movement by having the participants click on a predetermined spot before starting each task. Each task commenced with five randomly placed objects in the VE, including a red target sphere that needed to be selected by the participant. After the participant selected the target, its location was changed, and the participant had to locate and select it again. This process was repeated until the fifth target was selected, after which a screen would appear to indicate the completion of the experiment.

4.3.8 Study Variables

In our study, we manipulated two independent variables: Technique (*LDTPT*, magnetic grasp, pinch, raycasting) and target size (4 mm). The dependent variables included task completion time, error rate, ease of use, ease of learning, and perceived naturalness of the selection method. 5 targets were positioned within the comfortable zone in the VE. The order of the conditions and factors was determined using Latin squares. As a result, our experiment had a total of $4 \times 5 = 20$ target spheres.

4.3.9 Data Collection Method

The data collection method for collecting data on error rate and task completion time was automated and recorded by the VR system, as these are objective measures of performance. Participants evaluated the performance of the selection techniques using a 7-point Likert scale, ranging from 1 (low) to 7 (high), on five criteria: easy to use, ease of learning, naturalness, preference, and effort. Specifically, participants rated their easy to use, ease of learning, and naturalness based on the statements: "It is easy to maneuver the method and understand what was happening", "*It is easy to learn the technique*" and "*It is natural to use the technique in the specific environment*". The easy to use, ease of learning, naturalness, preference, and effort ratings were collected through self-report measures using a 7-point Likert scale. Participants would rate each of these factors for each of the techniques tested in the study. Afterward, participants ranked the gestures according to their preferences and effort based on the statement "rank the gesture according to your highest preference" and "the gesture that requires the least amount of effort." The data was collected through questionnaires administered after the end of the experiment.

4.4 Results

The normality of data was assessed using a Shapiro-Wilk test ($p>0.05$), as well as by visually examining histograms, normal Q-Q plots, and box plots. A one-way ANOVA test with a Bonferroni correction was conducted to find significant differences in normally distributed data. Additionally, we used the Friedman non-parametric test with Dunn-Bonferroni posthoc tests for

Likert-scale data. In the subsequent section, we report the results of task completion time, error rate, easy to use, ease of learning, naturalness, user preference, and effort scores for all the techniques.

4.4.1 Task Completion Time

The differences in the task completion time between the four techniques were analyzed using ANOVA with a significance level of 0.05. We found statistical significance in the task completion time for all four techniques ($F(3,39)=10.411$, $p=0.001$), revealing a statistically significant difference. The effective size R was found to be 0.62. This represents a medium effect size. Post-hoc tests with Bonferroni correction indicated that the *LDTPT* method (avg = 24.66s, $Z = -2.430$) was significantly faster than all other techniques, including raycasting (avg = 29.24s, $Z = -2.730$, $p = .045$), pinch (avg = 27.27s, $Z = -3.479$) and magnetic grasp (avg = 40.77s, $Z = -4.430$). Significant differences were found between magnetic grasp and pinch ($p= 0.001$), and between magnetic grasp and raycasting ($p=0.01$). No other significant differences were observed between the conditions. These results suggest that *LDTPT* performed significantly better than the other three techniques in task completion time, followed by pinch, raycasting, and magnetic grasp, respectively. Figure 16 presents the mean and SD values for task completion time for all four techniques.

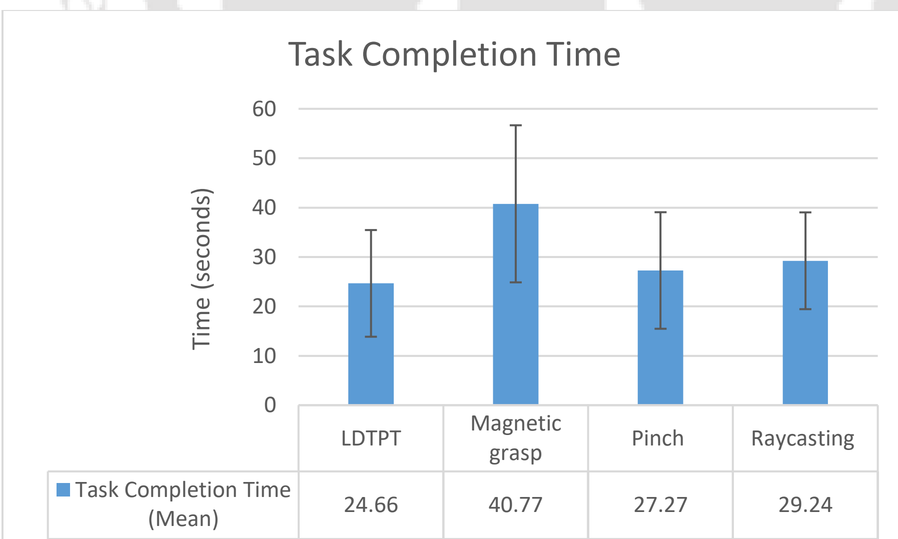


Figure 16: The mean and SD values for task completion time for techniques *LDTPT*, Magnetic grasp, Pinch, and Raycasting respectively.

4.4.2 Error Rate

The differences in the error rate between the four techniques were analyzed using ANOVA with a significance level of 0.05. We found statistical significance in the task completion time for all four techniques ($F(1,35) = 33.388$, $p < 0.001$. Post-hoc tests with Bonferroni correction indicated that the *LDTPT* (M=0.11, SD=0.1) made significantly lower errors than all other techniques, raycasting (M=0.49, SD=0.2), pinch (M=3.19, SD=2.6) and magnetic grasp (M= 0.74, SD=0.6). Significant differences were found between magnetic grasp and pinch (p= 0.001), and between pinch and raycasting (p=0.01). No other significant differences were observed between the conditions. These results suggest that *LDTPT* made significantly less errors than the other three techniques raycasting, pinch and magnetic grasp. Figure 17 presents the mean and SD values for error rate for all four techniques.

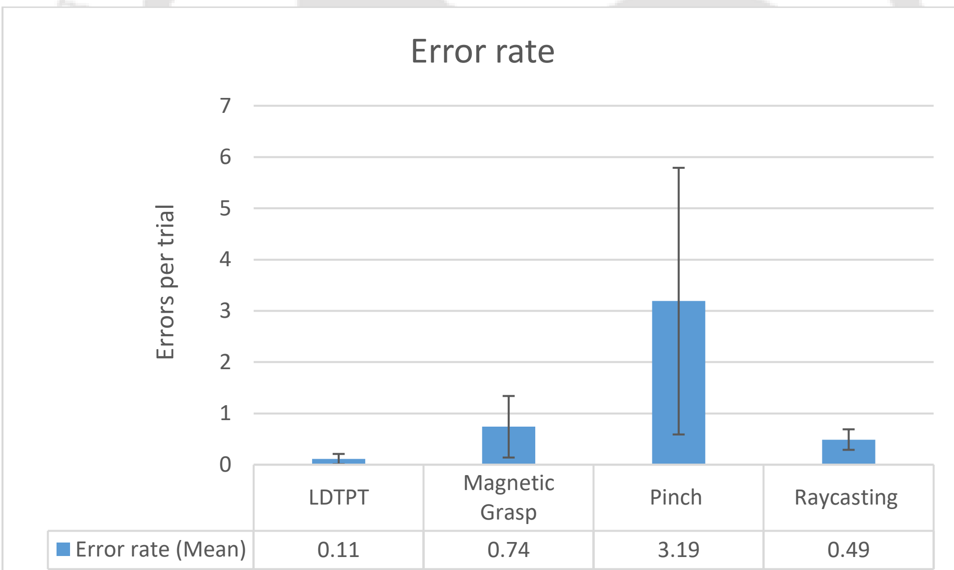


Figure 17: The mean and SD values for Error rate for techniques *LDTPT*, magnetic grasp, pinch and raycasting respectively.

4.4.3 Easy to Use

A Friedman test was carried out to compare easy to use of all four techniques. A significant difference was found between all the techniques, $\chi^2(3)=32.44$, p=0.001. Dunn-Bonferroni post hoc tests were carried out, and *raycasting* was significantly easier to use than the other techniques.

There were significant differences between *LDTPT* and *magnetic grasp* (Z= -3.68, p=0.0001), *pinch* and *raycasting* (Z=-2.97, p=0.001), *pinch* and *magnetic grasp* (Z=-4.003, P=0.001) and between *magnetic grasp* and *raycasting* (Z= -4.748, p= 0.001). There were no significant differences between any other combinations. Figure 18 presents the mean and SD values for ease of use for all four techniques.

4.4.4 Ease of Learning

A Friedman test was carried out to compare the ease of learning of all four techniques. A significant difference was found between all the techniques, $\chi^2(3)=21$, p=0.001. Dunn-Bonferroni post hoc tests were carried out and *raycasting* was significantly easier to learn than the other techniques. There were significant differences between *raycasting* and *LDTPT* (Z=-2.28, p=0.0001), *pinch* and *raycasting* (Z=-2.27, p=0.001), and between *magnetic grasp* and *raycasting* (Z= -3.48, p= 0.001). There were no significant differences between any other combinations. Figure 18 presents the mean and SD values for ease of learning for all four techniques.

4.4.5 Naturalness

A Friedman test was carried out to compare the naturalness of all four techniques. A significant difference was found between all the techniques, $\chi^2(3)= 21.24$, p=0.001. Dunn-Bonferroni post hoc tests were carried out and *LDTPT* was significantly more natural than the other techniques. There were significant differences between *LDTPT* and *raycasting* (Z=-3.17, p=0.0001), *magnetic grasp* and *raycasting* (Z= -2.18, p= 0.001), and *pinch* and *raycasting* (Z=-2.76, p=0.001), and between. There were no significant differences between any other combinations. Figure 18 presents the mean and SD values for naturalness for all four techniques.

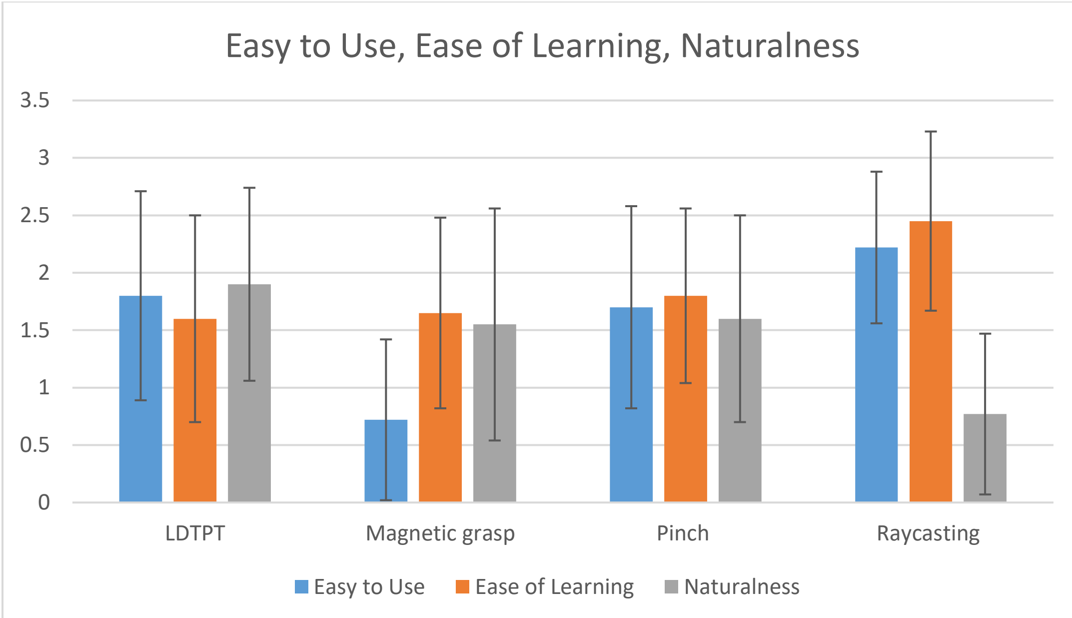


Figure 18: The means and SD values for ease of use, easy to learn and perceived naturalness for techniques *LDTPT*, magnetic grasp, pinch and raycasting respectively.

4.4.6 User Preference and Effort

Participants' preferences for the selection techniques were assessed by asking them to rank the techniques based on their selection task. The results showed that 29 out of 40 participants (72%) ranked the *LDTPT* gesture as their first preference. In contrast, only 10 participants rated raycasting as their second preference. However, it is noteworthy that raycasting was rated as requiring the least effort (30 out of 40) among all techniques. This was followed by *LDTPT* (19 out of 40) and pinch (12 out of 40). Magnetic grasp was the most effortful considering all the techniques.

4.5 Discussion

For H1, the results indicate that the LDTPT technique outperformed all other techniques in terms of task completion time. Hence, we accept H1 in which we hypothesized that LDTPT will be faster than all other techniques. Participants found the locking phase of LDTPT helpful in situations where errors were made, as it provided an additional step and time to recover from the error. A participant said, *"The second step (lock phase) could be helpful in situations where we need 'error recovery' because it gives additional step (time) to come out of any error selection."* Participants also found that using two fingers to select targets with the *LDTPT* technique provided more stability and control over the selection process, particularly for small targets in dense virtual environments. By increasing the finger pad area, participants were able to select small targets more accurately, which is particularly important in virtual environments where the accuracy of target selection can affect task performance. A participant said, *"I could increase and decrease my finger pad area according to the different target sizes in a dense environment. It is also very intuitive and easy to remember".* Participants found the visualization function of the *LDTPT* technique helpful. It allowed participants to adjust the finger pad area according to different target sizes, which made it easier to select nearby and far objects in dense virtual environments. One participant added, *"This visualization function could be helpful to select nearby and far object selection."*

We found that target size significantly affected the selection time for magnetic grasp and pinch techniques. With the magnetic grasp technique, the virtual hand often obstructed the view of the target despite its reduced opacity, making it difficult for the user to determine if the object was indicated for selection. In dense VEs, this technique required precision to select objects and resulted in longer selection times and higher error rates. Participants also struggled to correctly select small targets due to their position in the environment and commented that they were unable to see if the targets were highlighted. A participant commented*, "I cannot see if the small targets are highlighted. Also, the target position (target depth) makes it difficult to select it correctly."* For the pinch gesture, participants found it challenging to determine how much to pinch, and the size of the smallest objects often could not accommodate the user's index finger and thumb during the final selection. One participant suggested that visualizing a line indicating the distance to pinch could be helpful. The Pinch technique also depends on accurate finger tracking, which cannot always be provided by an optical tracking system like the Leap Motion, which views the hands from only one direction. Overall, as a result of Leap Motion's hardware limitations, hand tracking is not always accurate, which resulted in the loss of tracking of the hand while using the techniques.

For H2, in terms of error rate, the *LDTPT* technique made low errors (1%) as compared to other techniques, pinch, magnetic grasp, and raycasting techniques. Thus we accept H2. This could be attributed to the two-step process of *LDTPT*, which allowed participants to lock onto the target and confirm the selection before executing the final selection. This process reduced the number of accidental selections and helped the participants recover from errors. In contrast, the raycasting technique exhibited a higher error rate than other techniques, which could be attributed to hand jitter. A participant said, *"As the targets are too small, using raycasting to select objects is difficult.*

I might mistakenly select the nearby target." Previous research studies have reported the impact of hand jitter on the accuracy of raycasting (Cockburn and Firth, 2004).

On the other hand, the pinch technique had a higher error rate due to the difficulty in determining the appropriate pinch distance and the small size of the smallest objects which made it hard to accommodate the index finger and thumb during the final selection. The magnetic grasp technique had a higher error rate because of the occlusion caused by the virtual hand, which made it difficult to determine if an object was indicated for selection, especially for small targets. Additionally, the grab trigger distance used could be higher to select objects from a distance, which led to accidental selection.

For H3, *LDTPT* technique was significantly more difficult to use and learn than the other three selection techniques. We hypothesized that LDTPT technique will be easy to use and learn. Thus we reject H3. The participants found it challenging to learn and use as participants had to remember the different phases of the technique to understand the audio and visual feedback provided during each phase during initial interactions. 3 participants commented on how feedback is provided. A participant commented, "*The visual feedback adds to cognitive load (in a dense environment); however, it provides the user the control to know what was happening and in which stage they were in*."

For Pinch technique it relies on accurate finger tracking, which the Leap Motion optical tracking system can only view hands from one direction, which can result in inaccurate finger tracking and unreliable pinch detection. Pinch could be difficult to use as it requires a certain level of dexterity and fine motor control to hold the pinch gesture to select small objects. The magnetic grasp technique was also difficult to use. One potential challenge is that it requires more concentration and precision to select and deselect objects compared to other techniques. This is because the magnetic force of each finger attract multiple objects within the grabbing distance, making it difficult to select a specific small object in a dense VE. This can be frustrating for users who need to re-grab objects multiple times. Raycasting was the most easy to use and learn technique. Participants had to only point the ray to select target objects which were within arms' reach.

For H4, we hypothesized that *LDTPT* will be more natural as compared to other techniques. Our results indicate that LDTPT was significantly more natural than the other three techniques. Thus we accept H4. We believe the *LDTPT* techniques' adoption from a user-generated gesture study resulted in increased familiarity and naturalness.

We observed multiple limitations in *LDTPT* technique. One of the limitations of *LDTPT* is the perceived time-consuming phases of LDTPT. While the technique performed well in task completion time and accuracy, participants still found the process of locking the target before final selection to be an additional step that created a perception of slowing them down. This suggests that some users may prefer a more streamlined selection process. Another limitation is the occlusion of the VE view by the virtual hand used in *LDTPT*. Although the confirmation step ensured accurate selection, participants felt that a less occluded view would allow them to navigate the VE more easily, particularly when objects are small and placed in a dense environment where accidental selection is more likely. This limitation may affect the overall user experience. A third limitation of *LDTPT* is poor mapping and disconnect between the size of the virtual hand and the objects. Participants found that the size of the virtual hand was disproportionately larger than the objects, particularly small objects in the VE. While participants felt that it does not directly impact selection efficiency, it adds visual clutter to the user interface, which may negatively affect user experience. Overall, these limitations suggest that further research is needed to explore potential modifications in *LDTPT* or alternative techniques that could address these limitations and improve the overall user experience.

4.6 Chapter Summary

Chapter 4 discusses the evaluation of a novel selection technique called *LDTPT*, which stands for *Lock Dwell Time Point and Tap*. The study compared *LDTPT* with three existing techniques: pinch, magnetic grasp, and raycasting. The study involved 40 participants who completed selection tasks for small objects placed within arm's reach in a dense VE. The results showed that *LDTPT* was the most efficient technique in terms of task completion time and had the lowest error rate compared to the other techniques. Participants also preferred *LDTPT* over the other techniques. *LDTPT* uses a two-step process where the user first locks onto the target and then confirms the selection. The lock step helps avoid accidental selection and provides an additional step for error recovery. The technique also allowed for different finger pad sizes depending on the target size, which proved helpful for selection of small objects in a dense VE. Magnetic grasp and pinch techniques showed limitations in small target selection. Using magnetic grasp caused occlusion of the target. Raycasting had higher error rates due to hand jitter, which was consistent with previous studies. Participants found it difficult to select small targets with raycasting in dense environments. Despite the advantages of *LDTPT*, it has few limitations. The virtual hand used in *LDTPT* occluded the view of the VE. Participants recommended a less occluded view to navigate the environment more easily, especially in dense VEs with small objects. Additionally, the size of the virtual hand was disproportionally larger than the objects, adding visual clutter to the user interface.

This study contributes to the existing literature on selection techniques in virtual environments. Previous studies have evaluated selection techniques based on various factors such as task completion time, accuracy, and user preference. *LDTPT* adds to the literature by introducing a new selection technique and comparing it to existing techniques in a dense VE for small object selection. The study provides insights into the strengths and limitations of *LDTPT* and other selection techniques. In conclusion, the study provides valuable insights for designers and researchers working on selection techniques in virtual environments. *LDTPT* showed promise as a selection technique, providing a more efficient and accurate selection process for small object selection in a dense VE. However, improvements are needed to address the limitations identified by participants. The study highlights the importance of considering user feedback and addressing limitations in the design of selection techniques.

In the next chapter, Chapter 5, we presents a new technique called Tiny hands, which aims to overcome the limitations of *LDTPT* while retaining its advantages. The chapter provides a detailed explanation of the Tiny hands technique.

Chapter 5

5. Design and Evaluation of Tiny Hands: An Object Selection Technique to Select Small Objects Within Arms' Reach in Dense VE

In chapter 4, we conducted a study in which we developed the LDTPT technique and evaluated its performance against other established techniques such as magnetic grasp, pinch, and ray casting techniques. Our findings indicated that the LDTPT technique outperformed the other techniques in terms of task completion time and exhibited lower error rates. Additionally, it was the most preferred technique due to its natural and low-effort operation. However, LDTPT technique was not easy to use and learn. We identified several limitations associated with the LDTPT technique, such as virtual hands occluding the VE, and poor mapping between the size of the virtual hands and objects.

In this chapter, we addressed the limitations identified in Chapter 4 by introducing a new technique, Tiny Hands, for small object selection in dense VEs. The technique involves an object selection process that allows users to scale the size of virtual hands to accurately select small objects placed within arm's reach in a dense VE. We conceptualized three gesture-based interaction techniques to trigger and scale tiny hands, (i) fisting and rotating the non-dominant hand, (ii) holding the palm downwards for two seconds and performing vertical movements upwards and downwards, and (iii) pinch-in and pinch-out of the non-dominant hand. In this chapter, we report on two studies focused on small object selection in dense VE using the tiny hands technique. We aim to investigate research question RQ 2 and RQ 2.1 in this chapter.

In the first study, we conducted a formative study to evaluate the three gesture-based interaction technique variations of the tiny hands object selection technique. Our initial findings suggest that the tiny hands technique is a versatile approach for selecting small objects in a dense VE. Participants found the technique natural, intuitive, easy to use, and accurate in selecting small targets in dense VEs. Among the three interaction techniques presented, participants preferred holding the palm downwards and performing vertical movement upwards and downwards to trigger tiny hands as it provided an easy-to-understand mapping of the vertical movement to the size of the tiny hands to select a small object in dense VE. In the second study, we conducted a comparative evaluation of the tiny hands technique with ray casting and pinch-to-select selection techniques. Our results show that the tiny hands technique outperformed the other techniques in terms of accuracy, task completion time, and user preference. Participants also reported that the tiny hands technique was more natural, and easy to use compared to the other techniques. The study findings indicate that the tiny hands technique is a promising approach for small object selection in dense VEs, providing users with greater control and freedom in navigating and accurately selecting small objects in a VE.

Chapter 5 introduces the tiny hands technique, which is aimed at addressing the challenges of small object selection in dense VEs. The chapter starts with the design rationale for the technique and presents the three interaction techniques used to trigger the tiny hands gesture in section 5.2. The three interaction technique are: fisting and rotating non-dominant hand, holding the palm downwards with non-dominant hand and performing vertical movement, pinching-in/out the non-dominant hand (section 5.2.1-5.2.3). Section 5.3 presents an informal evaluation of interaction techniques of tiny hands. The insights and findings of the study are presented in section 5.4. Section 5.6 presents the second study to evaluate *tiny hand* for small object selection within arm's reach in a dense VE. Section 5.7 presents the tiny hands technique using the palm downwards with non-dominant hand and performing vertical movement to trigger tiny hands. The user study is presented in section 5.8. The finding and results of the study are presented in section 5.9. The findings are elaborated and discussed in section 5.10. Finally, section 5.11 summarizes the chapter, emphasizing the effectiveness and versatility of the tiny hands technique for small object selection in dense VEs, and its potential to improve user experience in virtual reality applications.

5.1 Design Rationale-*Tiny hands*

In VR, using virtual hands to select objects is a popular and intuitive form of interaction, drawing on real-world analogies. However, when selecting an object with the virtual hand, it occludes the virtual object that leads to erroneous object selection (Bowman et. al., 2005; Witmer and Singer, 1998). Similar observations were revealed when evaluating LDTPT technique for selection of small objects within arm's reach for a dense VE. While the gesture of *point and tap*, adopted from the gesture elicitation study was found accurate, error-free and natural, it also found limitation in terms virtual hand occluding the environment. Further, we found a disconnect between the size of the virtual hand and the small objects presented in the dense VE. This resulted in difficulty in learning and using the LDTPT technique.

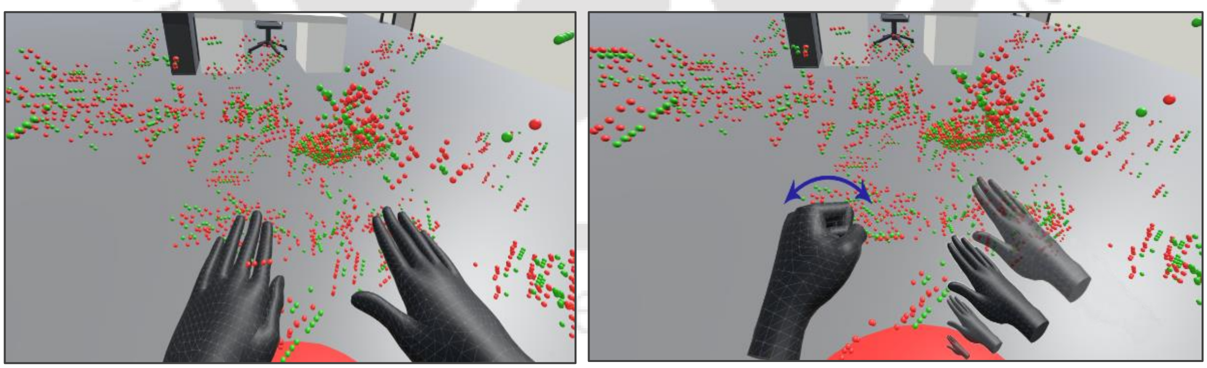
To overcome the challenges and limitations experienced in LDTPT technique, we propose tiny hands technique to select small objects within arm's reach in dense VEs. The objective is to explore the possibilities provided by VR technologies of using virtual hands and extend them in ways not possible in the real world to design a faster, accurate and efficient object selection method for small objects in a dense VE. Tiny hands aims to retain the pointing gesture elucidated from gesture elicitation study while ensure the limitations of LDTPT is eliminated. Tiny Hands technique involves scaling the size of the dominant virtual hand to facilitate navigation within the VE. Once the virtual hand is scaled to the desired size, the user can point at the desired object to select it. As the name suggests, the tiny hands technique reduces the size of the dominant virtual hand to allow for easy navigation and selection within dense VEs, while still enabling accurate selection through pointing. To activate tiny hands in a VE, the non-dominant hand performs a gesture (explained in coming sections based on each different interaction technique) specific to each interaction technique. The virtual hand's size is reduced by half (a gain factor of 0.5) from its default size. The maximum size of the virtual hand is full (default size 1), while the minimum size is 1/10th (a gain factor of 0.1). Users can modify the virtual hand's size by manipulating the gesture to increase/decrease the size of the dominant virtual hand. Users mirror the movements of their physical hand with the virtual hand in the VE. To confirm the selection of a desired object, the user points at it using their virtual dominant hand. When the virtual hand intersects with the desired object, the object is then considered as selected. We believe that the tiny hands technique has the potential to overcome challenges identified for LDTPT technique, associated with occlusion of virtual hands and poor mapping of the virtual hand to small object sizes.

5.2 Design and Evaluation of Interaction Techniques for Tiny Hands

We have developed and proposed three interaction techniques for the tiny hands concept. Although all three techniques involve scaling the virtual hand size, the methods for triggering and manipulating the tiny hands differ. The first technique involves the user making a fist gesture with their non-dominant hand to activate and rotate the fist to scale the dominant virtual hand for direct selection. In the second technique, the user places their non-dominant hand with palm facing downwards for two seconds, triggering the tiny hands on the dominant virtual hand. The user can then perform vertical movements with their non-dominant hand to adjust the size of the tiny hands for direct selection. The third technique involves the user performing a pinch gesture (index finger and thumb) using their non-dominant hand on top of their dominant hand. This triggers the tiny hands, which can then be adjusted using pinch-in/out movements for final selection. These techniques are explained in detail below.

5.2.1 Interaction Technique 1: Fisting and Rotating Non-Dominant Hand

This interaction technique involves fisting the non-dominant hand for 2 seconds to activate tiny hands. The fist must be made with the palm facing downwards to trigger tiny hands. By default, the size of the virtual hand reduces to half (a gain factor of 0.5) of the actual size. The user can increase or decrease the size of the virtual hand by rotating their wrist in a clockwise or anticlockwise direction. The fist can be released at any time during the user's interaction. To deactivate tiny hands, the user fists the non-dominant hand again for 2 seconds with the palm facing down. Figure 19 (a) and (b) visually demonstrates the fist-based interaction technique for tiny hands. Once the desired size of the virtual hand is achieved, the user can point and confirm the object selection.

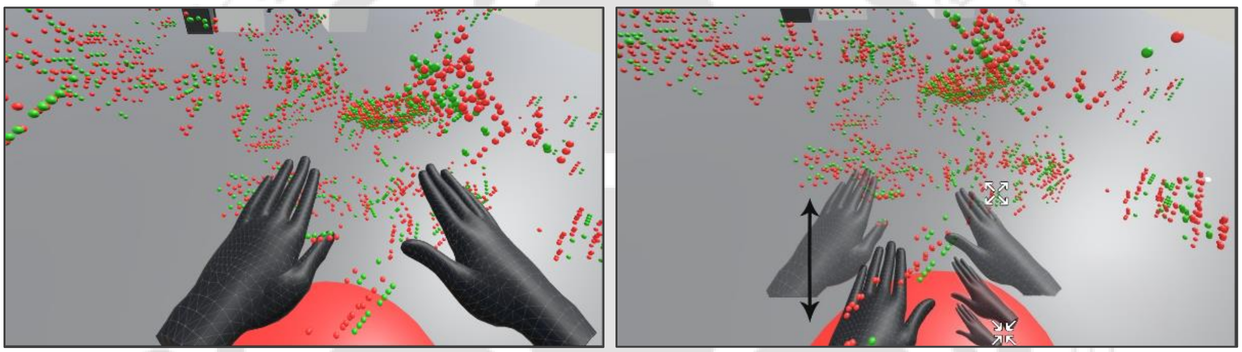


(a) (b)

Figure 19: Interaction technique 1: Fisting and rotating non-dominant hand (a) User brings the virtual hands in front of the user (b) the user makes a fist gesture on the non-dominant hand to trigger *tiny hands*. By rotating the fist clockwise/anticlockwise, user selects the hand size.

5.2.2 Interaction technique 2: Holding the Palm Downwards with Non-Dominant Hand and Performing Vertical Movement

This interaction technique involves the use of the non-dominant hand's palm facing downwards to trigger tiny hands, which is held for 2 seconds. The default size of the virtual hand is reduced to half (a gain factor of 0.5) as with the previous technique. The size of the virtual hand is then modified by performing vertical movements in the upwards and downwards direction. Once the desired size of the virtual hand is achieved, the user can point to select and confirm the object. The palm can be released at any time during the user's interaction. To deactivate the tiny hands, the user needs to place the non-dominant hand with palm facing downwards again for 2 seconds. Figure 20 (a) and (b) illustrates this technique.

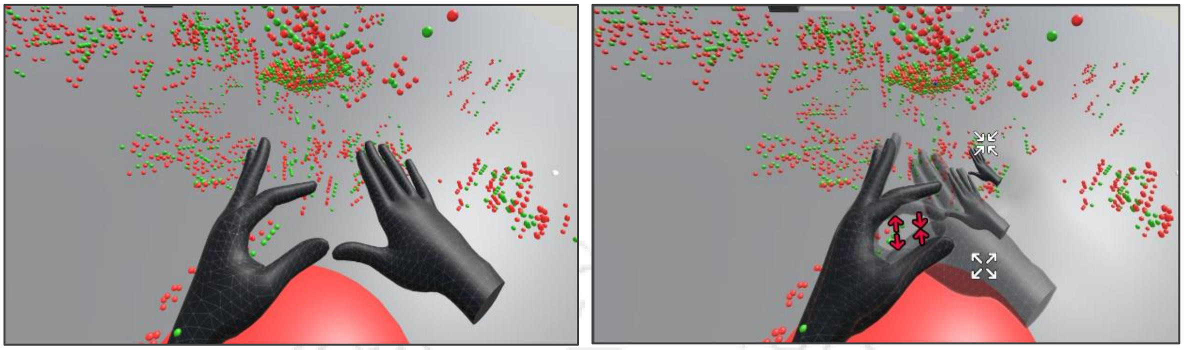


(a) (b)

Figure 20: Interaction technique 2: Holding the Palm downwards with non-dominant hand and performing vertical movement (a) Virtual hands are activated, user places non-dominant hand in the camera view static for 2 secs to trigger *tiny hands* (b) Vertical movement of the non-dominant hand up/down increases/decreases the virtual right-hand size.

5.2.3 Interaction Technique 3: Pinching-in/out the Non-Dominant Hand

The third interaction technique of the tiny hands concept involves pinching-in/out using the non-dominant hand. To trigger tiny hands, the user performs a pinch using the index finger and thumb of the non-dominant hand on the virtual dominant hand for 2 seconds. The size of the dominant virtual hand reduces to half (a gain factor of 0.5), similar to the other interaction techniques. Pinching in and out movements with the non-dominant hand are used to adjust the size of the virtual hand. The pinch-in/out interactions can be performed without placing the virtual non-dominant hand on the virtual dominant hand after activation. When the user achieves the desired virtual hand size, they can point to confirm the object selection. To deactivate tiny hands, the user repeats the pinch gesture on the virtual dominant hand using the non-dominant hand for 2 seconds. Figure 21 (a) and (b) provides a visual representation of this interaction technique.



(a) **(b)**

Figure 21: Interaction Technique 3: Pinching-in/out the non-dominant hand (a) Pinch in/out using index finger and thumb to decrease/increase the hand size. (b) Pinch-in using index finger and thumb to decrease the hand size.

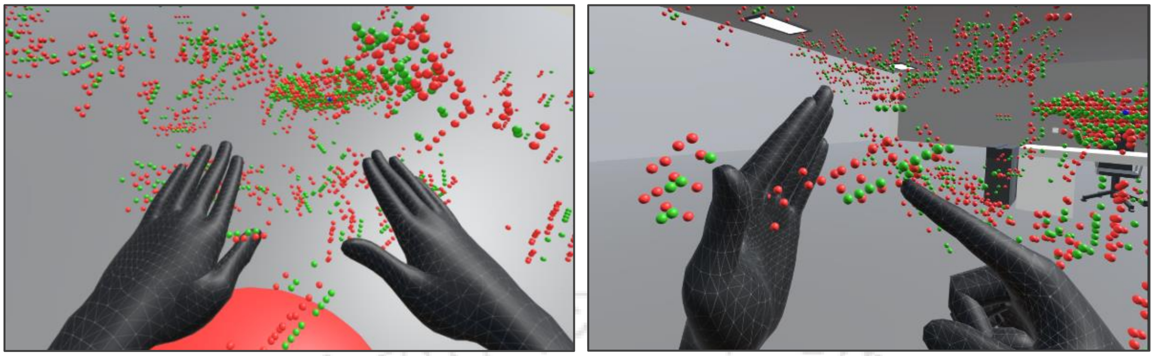
5.3 Study 1: Informal Evaluation of Interaction Techniques of Tiny Hands

Our study aimed to informally evaluate the usability of the "tiny hands" concept for selecting small-size objects in dense VEs using a think-aloud technique. We had two main objectives: (i) to learn about their likes, dislikes and preferences on the Tiny Hands concept and whether it is suitable for selecting small targets in a dense VE and (ii) to assess users' perception of the naturalness, ease of use, and accuracy of the three interaction techniques. To accomplish this, we implemented and evaluated the proposed interaction techniques using the Wizard-of-Oz technique.

The study included the design of the VE, recruitment of participants, task details, and study procedure. These are further elaborated in the following section.

5.3.1 Design of Virtual Environment

For the study, we designed a VE that featured a dense molecular structure consisting of 1192 spheres. The VE had dimensions of 20 x 15 x 10 feet and was modeled after a small-molecule compound commonly used in molecular modeling by biologists to study potential effects on humans. The spheres in the structure were colored red and served as distractors, while the target objects were colored blue. The spheres were static and measured 0.4 mm in diameter each. The molecular structure was positioned 0.4 meters away from the participant. Figure 22 (a) and (b) illustrates the VE utilized in the study.



(a) (b)

Figure 22: (a) A snapshot of the virtual environment and the virtual hands created for the study. (b) A comparison of the virtual hands size with the size of targets.

5.3.2 Study Participants

Ten participants (7 males, 3 females) aged between 18 and 35 (Mean= 27.22, SD= 5.12) were recruited for the study. All participants were right-handed and currently enrolled in graduate or PhD programs at the university. Participants had intermediate experience with VR technology, having used various VR platforms such as HTC Vive, Oculus Quest, and mobile phone VR for at least 2-3 hours per week over the last 6 months. Participants were not compensated for their participation.

5.3.3 Study Setup and Apparatus

The study was conducted using the Oculus Quest 1 and the Unity 3D game engine was utilized to develop the prototype. Visual stimulus was provided through the Oculus Quest HMD, which had a resolution of 1440 x 1600 per eye and a diagonal FOV of 115.52°. The Oculus Quest HMD was connected to a computer running on an i7 processor, integrated graphics 5500, 16 GB RAM, and a Microsoft Windows 10 operating system via the Oculus Link cable. During the experiment, the virtual hand (hand prefab provided by the Oculus Integration Kit) was scaled using the mouse scroll. A video camera was placed diagonally to the participant to capture visual data for further analysis of study results.

5.3.4 Study Task

The task assigned was to select the blue target that was placed at one location among the molecular structure. Each time the user selects the target, the target color changes to green to indicate selection and the next target location is indicated. This was repeated for 5 target selections.

5.3.5 Study Procedure

The study was conducted in a university laboratory, where we first introduced the concept of "tiny hands" to the participants. We then provided training on three interaction techniques for object selection. To minimize biases, we randomly assigned the techniques to the participants for selecting blue-colored target objects, which turned green after their successful selection. We used a think-aloud method (Ericsson and Simon, 1993) to learn about users' thought process during the interaction. Before the task, we showed the designed VE to the participants and encouraged them to verbalize their thoughts at each step of the evaluation. After completing the task, we conducted a semi-structured interview with the participants to understand their challenges and advantages of using tiny hands, as well as their perception of the ease of use, naturalness, and accuracy of the interaction techniques. We also discussed their likes and dislikes about the techniques. The study was video recorded for post-study analysis, and the interview findings and users' thought processes were documented via written notes.

5.4 Insights and Findings of the Tiny Hands as a Selection Technique:

In this study, we present two key sections outlining our insights and findings. The first section pertains to the evaluation of the tiny hands concept and whether it is suitable for selecting small targets in a dense VE. Additionally, we aimed to understand the users' preferences regarding the concept. Our findings suggest that the tiny hands technique was positively received by the participants for small object selection in a dense VE. The participants cited two primary reasons for the positive reception, namely (i) the mapping of reducing the size of virtual hands to select a small object, and (ii) the ease of navigation within a dense VE due to small virtual hands. Participants provided the following comments regarding the tiny hands concept: Participant 1 expressed positive comment, stating "*It is so cool to make the hand small. I can easily select the target*." Participant 2 noted the precision afforded by the technique, saying "*When the virtual hand size is small, the tip of index finger matches with the size of the target, which makes selection very easy and precise*." Three participants suggested that a visualization would be helpful in understanding the current size of the tiny hands, with one stating "*there could be a visualization (like a meter) showing the current size and how much to rotate the hand clockwise and* *anticlockwise for the desired size*" (Participant 4), and two others suggesting the inclusion of a widget (Participants 5 and 7), i.e., "*A visualization of a widget may be provided near the hand for users to understand the hand size currently in use and how much to rotate for increasing/decreasing to the next size*".

5.5 Insights and Findings on the Interaction Techniques

We conducted a think-aloud study to learn the users' perception of ease of use, naturalness, and accuracy of the 3 interaction techniques (i) Fisting and rotating non-dominant hand (ii) holding the palm downwards with non-dominant hand and performing vertical movement (iii) Pinching-in/out the non-dominant hand. In general, all three techniques received positive feedback in terms of being intuitive and natural for selecting small objects. A participant pointed that, *"It was easy to learn the technique, and I quickly became proficient in using it. Using tiny hands makes selection very easy and accurate."*

Regarding the first technique, where users rotated their fist to increase or decrease the tiny hand size, participants expressed that it was unintuitive and difficult to remember the gesture to trigger tiny hands and for an object selection task. A participant commented, *"The use of fist gesture with the palm facing downwards to activate tiny hands is not very intuitive or easy to remember for a selection task. I might need to be trained or reminded of the gesture repeatedly."* Rotating the fist could lead to reduced usability and fatigue if changes in tiny hands had to be made repeatedly. A participant articulated that, "*It was difficult to rotate the fist to increase or decrease the tiny hand size. It could also lead to fatigue if I had to make changes repeatedly. Also, the bent angle for fist for the smallest size makes the wrist a little fatiguing*." The gesture also presents difficulty in mapping to trigger tiny hands. A participant compared the rotation motion of the wrist to driving as used in many games, and stated that it felt frustrating. She added, "*I play many games, and making a fist and rotating left and right feels more like driving. It makes me feel that if I rotate left, the tiny hand should move toward the left and vice-versa. It feels frustrating, although I will get used to it*." Using the fist rotation gesture also presents with limited size adjustment. This may not provide enough flexibility for participants with different hand sizes or preferences, leading to discomfort or difficulty interacting with objects.

Regarding the second technique involving vertical movement for tiny hands, participants expressed their preference for this technique due to its appropriate mapping between the vertical hand movement to increase/decrease tiny hand size, making it easy to select the next desired size. A participant shared, "*Compared to other techniques, I found this one more intuitive as the vertical movement up/down to adjust the hand size was easy to remember while doing the task. It helped me select the desired size easily*." Participants also felt making larger movements of the arm and hand may be less fatiguing than smaller, more fine-tuned movements like wrist rotation for repeated selection tasks. A participant commented, *"I found that using vertical hand movements to trigger tiny hands was much less fatiguing than using smaller movements like pinching or rotating my wrist."* Vertical movements to increase decrease the tiny hand size also allow for a greater range of motion, which may result in more precise size adjustments. A participant commented, *"With vertical movements, I can easily adjust the size of the virtual hand. This maybe particularly useful when precise size adjustments are required. Vertical movements to adjust hand size is also more intuitive."*

Regarding the third technique, which involved pinch gesture to activate tiny hands and pinch in/out to modify the size of the dominant virtual hand, participants expressed that the activation of tiny hands using the pinch gesture on the virtual dominant hand is confusing and difficult to remember. One participant stated, *"I found it confusing to hold the pinch gesture on the top of the virtual hand and it was also difficult to perform the pinch-in/out gesture smoothly, especially when trying to adjust the virtual hand size quickly. It didn't feel as intuitive."* Few participants also felt limited range of motion for pinch in/out gesture to scale tiny hands. This could result in less flexibility and difficulty in adjusting the virtual hand size. A participant commented, "*I felt like I had limited range of motion when using the pinch in/out gesture to scale the virtual hand, to use it for fine adjustments. It felt difficult to get the virtual hand size exactly right."* In addition to usability issues, some participants also raised concerns about technical limitations with the pinch-in/out interaction technique for tiny hands. Depending on the device being used, it may not be able to accurately track the pinch-in/out movements on top of the dominant virtual hand, which could result in less reliable interactions. One participant commented on this issue, saying, *"I noticed that sometimes the hands disappear when I bring both hand too close to pinch-in/out. I had to repeat the gesture a few times before the virtual hand size would start to adjust. This made the interaction less smooth and more frustrating."*

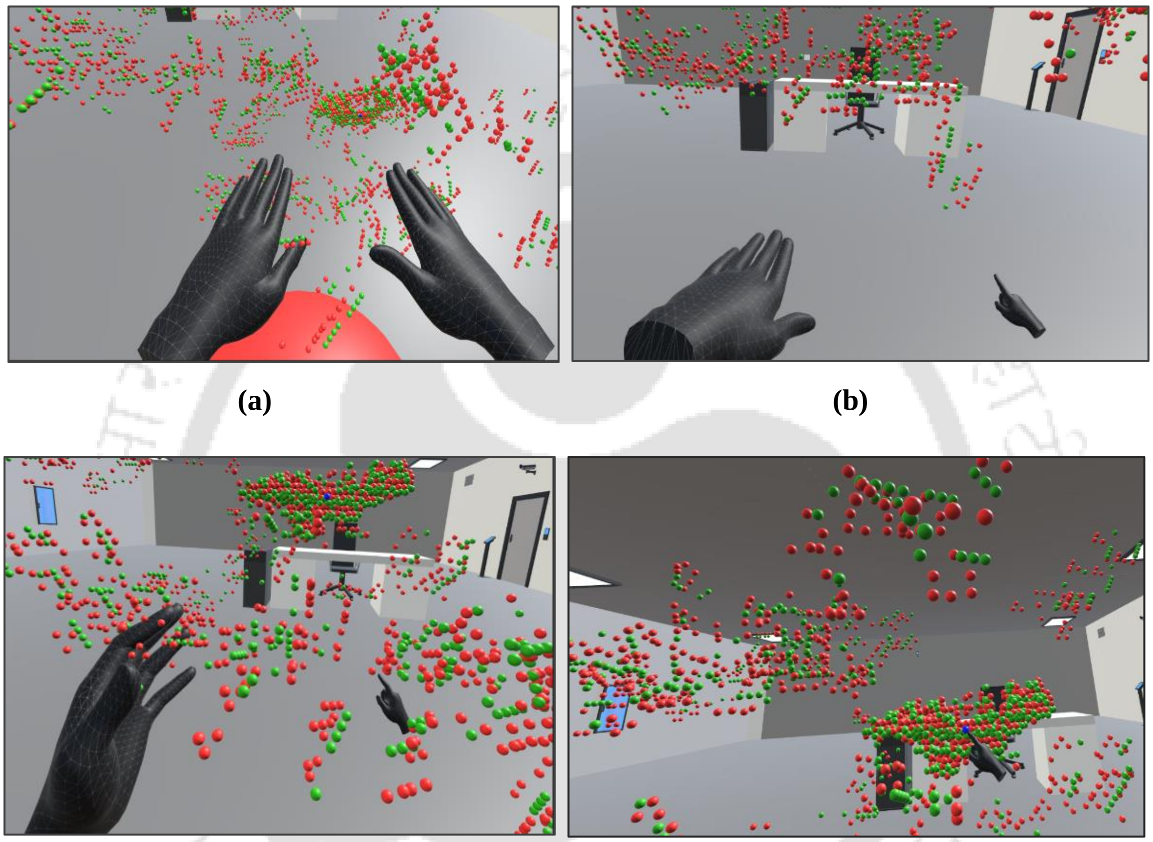
User evaluation and feedback indicated the *tiny hands '* technique activated through placing the hand with palm downwards and modified by vertically upwards/downward movement is perceived to be natural, and intuitive in selecting small targets in dense VEs. Vertically moving hands up and down also maps wells to users' mental models of making something small and large. Hence for the next study, vertically moving the hand up and down to increase/decrease the hand size was finalized as an interaction technique to activate *tiny hands*. In the next section, we present the evaluation by presenting comparative study of *tiny hands* performed by vertically moving hands up and down with existing techniques from literature, ray casting (Mine et al., 1997) and pinch-to-select (Oculus, 2022) to select small targets positioned within arms' reach in a dense VE.

5.6 Study 2: Evaluation of *tiny hands:* Object Selection of Small Objects Within Arm's Length for Dense VE

The objective of this research was to examine the effectiveness of the tiny hands technique for selecting small objects of arm's reach in a dense VE, specifically by analyzing task completion time, error rates, ease of use, ease of learning, and naturalness. In addition, user preference and effort for the tiny hands technique was also investigated. The study involves a comparison of the tiny hands technique to other techniques such as ray casting (Mine et al., 1997) and pinch-to-select (Oculus, 2022) methods. The tiny hands technique were activated through vertical downward and upward movements of the non-dominant hand. By conducting this study, we hope to gain insights into the benefits and limitations of the tiny hands technique for selecting small objects of arm's reach in dense VEs and determine its potential usefulness in various applications.

5.7 The *Tiny hands* Technique

The tiny hands technique is designed to facilitate the direct selection of small targets within arms' reach in dense VEs by scaling down the size of virtual hands. This technique involves reducing the size of the dominant virtual hand in immersive VR to navigate efficiently and accurately select the desired object. To activate tiny hands, the non-dominant hand palm is faced downwards and held in front of the user (in the camera view) for 2 seconds (Figure 23). The user then performs a vertical upward/downward movement to increase/decrease the size of tiny hands displayed on the dominant right hand. The size of the virtual hand can be scaled down to 0.1 from a gain factor of 1, which is the actual virtual hand size. The vertical movement in the upwards and downward directions can be used to increase or decrease the hand size, respectively. Once the desired virtual hand size is achieved, the user can point and intersect (or touch) with the desired object to confirm selection. Tiny hands are deactivated after the user holds the non-dominant hand in front of the user again for 2 seconds. We believe that the tiny hands technique allows for improved selection of small objects within arm's reach of dense VEs by reducing the size of the dominant virtual hand, making it easier for users to navigate and select objects with precision and accuracy.



(a) (b) (c) (d)

Figure 23: (a) User places virtual hands in front of the user for 2 seconds to trigger *tiny hands* (b) Vertical movement of the non-dominant hand downwards reduces the dominant hand size. (d) *Tiny hand* size is finalized (d) *Tiny hands* precisely selects the target.

5.8 User Evaluation of Tiny Hands

5.8.1 Research hypothesis

For this experiment, we formed five hypotheses:

H1: It is hypothesized that the Tiny hands technique will result in faster selection times than the Raycasting and Pinch-to-select techniques when selecting small and distant targets in a dense virtual environment.

H2: It is hypothesized that the Tiny hands technique will result in a lower error rate than Raycasting and Pinch-to-select techniques.

H3: It is hypothesized that the Tiny hands technique will be perceived as easier to learn and use compared to the Raycasting and Pinch-to-select techniques.

H4: It is hypothesized that the Tiny hands technique will be perceived as more natural than Raycasting and Pinch-to-select techniques.

H5: It is hypothesized that the Tiny hands technique will be perceived as the most preferred and low in effort than Raycasting and Pinch-to-select techniques.

5.8.2 Baseline Techniques

To evaluate the effectiveness of the tiny hands technique, we compared it against two commonly used techniques in VR: the ray casting technique (Mine et al., 1997) and the pinch-to-select technique. The ray casting technique involves pointing the index finger towards the target to cast a colored ray, which can be used to select objects within arm's reach. We adopted this technique as it is a widely used and effective method for selecting objects in VEs.

The pinch-to-select technique, on the other hand, is a popular technique used in current head-mounted display (HMD) devices, such as Oculus Quest (Oculus, 2021), Microsoft Hololens (Kress et al., 2017), etc. It uses a pinch metaphor for object selection, where the user can control the size of the cursor by pinching their thumb and forefinger together. The cursor is displayed as a circular cone that can be squeezed or expanded based on the pinch gesture. Once the fingers are pinched together, the cursor acts like a laser pointer that can be used to make selections on a distant screen. Additionally, this technique provides feedback on the direction in which the user is pointing.

5.8.3 Design of VE, Study Setup and Apparatus, and Task

The VE design, study setup and apparatus, and task for the user study are identical to the previous study. These details have been described in the previous sections of the paper (section 5.3.1-5.3.4)

To summarize the study details, the VE used in this study was the same as the one used in the previous study, featuring a dense molecular structure consisting of 1192 spheres colored red as distractors and blue as the target objects. The VE was modeled after a small-molecule compound used in molecular modeling by biologists to study potential effects on humans and had dimensions of 20 x 15 x 10 feet, positioned 0.4 meters away from the participant. The Oculus Quest 1 was used for the experiment with a resolution of 1440 x 1600 per eye and a diagonal FOV of 115.52°. The Unity 3D game engine was used to develop the prototype, and the Oculus Integration Kit provided the virtual hand prefab that was scaled using the mouse scroll.

A video camera was placed diagonally to the participants to capture visual data for further analysis of study results. The task assigned was to select the blue target that was placed at one location among the molecular structure. Each time the user selected the target, the target color changed to green to indicate selection, and the next target location was indicated. The study setup and apparatus remained the same as in the previous study, and the task assigned was also identical, with participants required to complete five target selections.

5.8.4 Participants

The study involved 22 participants, consisting of 12 males and 10 females, aged between 18 to 33 years (Mean = 27.42, SD = 4.37). All participants were Master's students who were studying interaction design. They had prior experience using HMD-VR platforms such as Oculus Rift and/or HTC Vive, and had completed a VR course in the previous semester of their program. No remuneration was provided to the participants for their involvement in the study.

5.8.5 Study Procedure

The study was conducted in a controlled laboratory environment at a university. Prior to the task, a moderator provided a verbal introduction to each of the 22 participants, who were all Master's students in the field of interaction design, and had prior experience with HMD-VR platforms. Participants were shown the VE and explained the three techniques being tested: the ray casting technique, the pinch-to-select technique, and the tiny hands technique. The moderator provided training to each participant until they felt confident in using each technique to complete the task. The training sessions lasted approximately 30 minutes per participant.

To begin the study, each participant was randomly assigned one of the three techniques to complete the task, which is to select a blue target object placed at 5 different location in the dense VE. Participants were permitted to adjust the size of their tiny hands at any point during the study, and any changes to the size were also noted. After completing the task, post-task interviews were conducted to gain insight into the participant's overall perception of the proposed techniques. The entire study was video recorded and lasted approximately 90 minutes per participant.

5.8.6 Data Collection Method

The experiment involved 22 participants, who completed a total of 330 trials (3x5x22) using three techniques (Tiny hands, Raycasting, Pinch-to-select) to select 5 targets. The application recorded the completion time and error rate for each trial. The task completion time was calculated after participants set the tiny hands size and selected the first target. An error was counted when the target selected by the user did not match the target specified during the task. The moderator noted the hand size selected by the participant. Additionally, the moderator administered a 7-point semantic differential scale to gather feedback on the ease of use, ease of learning, and naturalness of each technique. Participants were provided with a printed paper to collect their responses on ease of use, learning and naturalness. Participants ranked their preference for each technique at the end of the study, and the post-experiment interviews were recorded for further analysis.

5.9 Results

During the study, we collected data on user performance and preferences using system logs and questionnaires. The logs recorded data related to the completion time of the task and the number of incorrect selections made by the users. To assess the normality of the collected data, we employed the Shapiro-Wilk test. To determine significant differences in normally distributed data, we conducted a one-way ANOVA test with a Bonferroni correction. Furthermore, for the data collected through the Likert-scale questions, we used the Friedman non-parametric test, followed by Wilcoxon Signed Ranks post-hoc tests.

5.9.1 Task Completion Time

The study used ANOVA to analyze the differences in task completion time among the three techniques, with a significance level of 0.05. The results showed a statistically significant difference in task completion time between all three techniques ($F(2, 63)=8.86$, $p=.001$). Post hoc analysis using the Tukey test found that the mean task completion time for the tiny hands approach ($M=43.97$, $SD=12.90$) was significantly faster than both the pinch-to-select ($M=76.36$, $SD=34.93$) and ray casting ($M=64.51$, $SD=24.74$) approaches. There was no significant difference in task completion time between pinch-to-select and ray casting. Figure 24 displays the mean and standard deviation values for task completion time

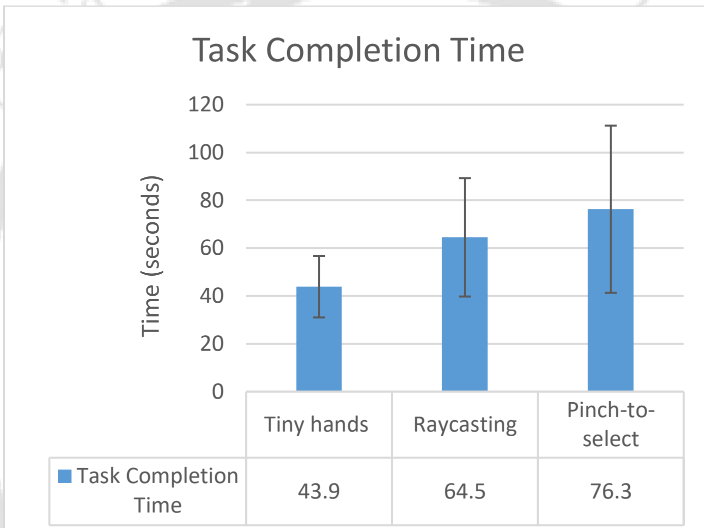


Figure 24: The average task completion time in seconds for the three techniques: *tiny hands*, ray casting and pinch-to-select

5.9.2 Error rate

To analyze the differences in error rate between the three techniques, we conducted ANOVA with a significance level of 0.05. The Tukey post hoc test indicated that there were significant differences in error rate between tiny hands ($M=0$, $SD=0$) and raycasting ($M=3.09$, $SD=2.1$) ($p<0.05$), as well as between tiny hands and pinch-to-select ($M=3.86$, $SD=1.96$). Notably, the tiny hands technique produced zero errors. Figure 25 displays the mean and standard deviation values for error rate for all three techniques.

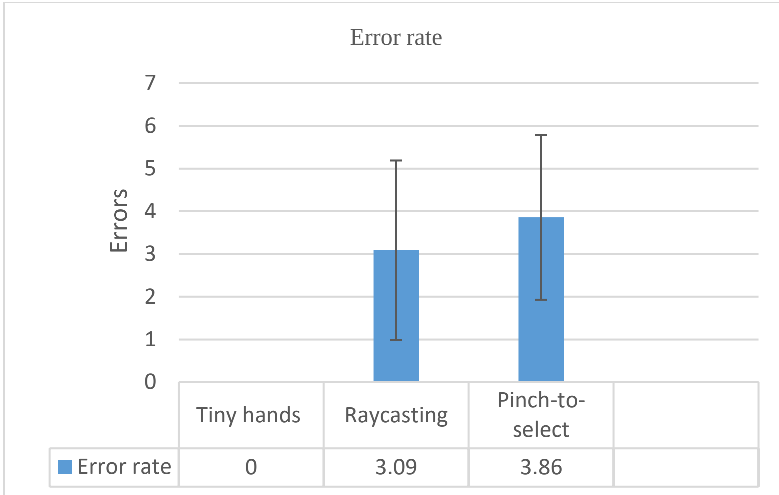


Figure 25: Error rate for the three techniques: *tiny hands*, raycasting and pinch. *Tiny hands* made zero errors.

5.9.3 Easy to Use

A non-parametric Friedman test was conducted to compare the ease of use of all three techniques. The test showed a significant difference between the techniques, $\chi2(2)=28.93$, $p<0.001$, indicating that at least one technique was significantly different from the others. Subsequently, post-hoc tests were performed using Wilcoxon Signed Ranks tests to identify the specific differences between the techniques. The results indicated that there were significant differences in the ease of use between tiny hands and Raycasting ($Z= -3.91$, $p=0.00$, as well as between tiny hands and Pinch-to-select ($Z=-4.01$, $p=0.003$). However, there was no significant difference between Pinch-to-select and Raycasting ($Z=-0.83$, $p=0.41$). Figure 26 shows the mean and standard deviation for easy to use for all the three techniques: tiny hands, ray casting and pinch-to-select.

5.9.4 Ease of Learning

A significance test was conducted to compare the ease of learning among the techniques, yielding a $\chi2(2)=19.90$ with $p<0.001$, indicating a statistically significant result at $p < .05$. Further analysis using the Wilcoxon Signed Ranks test revealed significant differences in the ease of learning between tiny hands and Pinch-to-select ($Z= -1.81$, $p<0.05$), and between ray casting and pinch-to-select ($Z=-1.14$, $p<0.05$). However, no significant difference was found between Tiny hands and ray casting ($Z=-0.13$, $p<0.05$). Figure 26 shows the mean and standard deviation for ease of learning for all the three techniques: tiny hands, ray casting and pinch-to-select.

5.9.5 Naturalness

The naturalness of the techniques was compared using a Friedman test, which yielded a significant difference between all the techniques ($\chi 2(2)=22.90$, $p<0.001$), indicating that there were significant differences in perceived naturalness among the techniques at a significance level of $p < .05$. Wilcoxon Signed Ranks post-hoc tests were conducted to examine the differences between pairs of techniques. The results showed significant differences in naturalness between tiny hands and ray casting ($Z= -2.18$, $p<0.05$), between tiny hands and pinch-to-select ($Z=-3.21$, $p<0.05$), and between ray casting and pinch-to-select ($Z=-1.64$, $p<0.05$). Figure 26 shows the mean and standard deviation for naturalness for all the three techniques: tiny hands, ray casting and pinch-to-select.

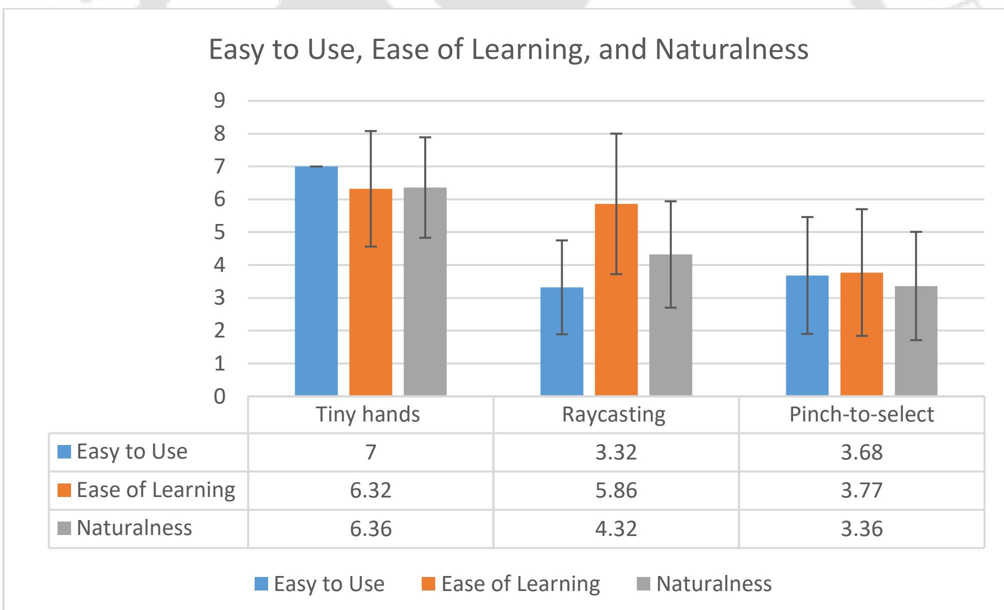


Figure 26: The mean and standard deviation for easy to use, ease of learning, naturalness and playfulness for the three techniques: *tiny hands*, ray casting and pinch-to-select

5.9.6 Preference and Effort

All 22 participants indicated a clear preference for the tiny hands technique as their first preference over the other two techniques. Ray casting was ranked as the second preference by 15 out of 22 participants. Only 5 participants ranked pinch-to-select as their second preference.

Participants perceived the tiny hands technique as requiring less effort compared to the other techniques (20 out of 22). Among the three techniques, the pinch-to-select technique was rated as more effortful by participants (18 out of 22).

5.10 Discussion

In this section, we present the results, hypothesis validation, and other findings of our study. Our investigation focused on comparing task completion times, error rates, ease of use, learning, naturalness, and user preference across three object selection techniques for small and distant objects within a dense VE. The three techniques we tested were Tiny hands, Raycasting, and Pinch-to-select. We will discuss the statistical results, validate our hypotheses, and provide a comprehensive overview of the subjective findings of the study.

H.1 predicted that tiny hands would result in faster task completion times compared to raycasting and pinch-to-select techniques. Statistical analysis revealed that Tiny hands was significantly faster to select small objects within arms' reach as compared to baseline techniques. Hence we accept H.1. The participants reported that the effectiveness of the technique could be attributed to two key factors: firstly, the ability to select small objects by making the virtual hands small and secondly, the ease of navigating within a dense virtual environment due to the small size of the virtual hands. Additionally, the small size of the virtual hands allows for a smaller selection volume, resulting in a higher level of accuracy and reduced errors when using the tiny hands technique. These findings suggest that the tiny hands technique could be a useful approach for selecting smaller objects in virtual environments, particularly when precision is required. Similar findings were reported by Piumsomboon et al., (2018), where a scaled down version of life-size avatar took significantly less time to complete a task as compared to baseline technique. In contrast Pinch-to-select look a longer time for selection as compared to both techniques. Participants found it difficult to adjust the cone formed between the index finger and thumb to select a small target which consumed most of the time. With ray casting, users had to precisely point towards an object, which often resulted in incorrect selection and inaccuracies due to hand jitter and difficulty in accurately pointing the ray as the VE was dense – a phenomenon commonly observed in ray casting technique (Batmaz and Stuerzlinger, 2019).

For H.2, our hypothesis was that tiny hands would result in lower error rates compared to Raycasting and Pinch-to-select technique. The results showed that tiny hands had no errors as compared to other techniques. Thus, we accept H.2. For tiny hands the target size matched with the size of tiny hands due to which tiny hand was able to accurately select the target resulting in error free selection. For raycasting the higher error rates was a result participants precisely pointing the ray to accurately select a small target. However, as the VE was dense, raycasting led to wrong selection of small objects. With pinch-to-select technique, we found that participants experienced difficulties in accurately selecting small objects. One of the reasons for this was the limited precision of the pinch gesture, as participants were unable to correctly measure the distance between their thumb and index finger while selecting the object. This led to incorrect object selection. Our findings are consistent with previous studies that have also reported limited precision with the pinch-to-select technique, which resulted in difficulties in accurately selecting small objects (Jiang, H., Weng, D., Dongye, X., & Liu, Y., 2022).

For H.3, we hypothesized that Tiny hands would be easy to use and learn as compared to Raycasting and Pinch-to-select technique. Our analysis also shows that participants found the tiny hands technique to be easy to learn, and use. Thus we accept H.3. This was attributed to its ability to customize the technique for a greater sense of control and confidence. The use of the non-dominant hand to trigger the tiny hands and further customize it by vertical movements was well-received by the participants, as it provided a sense of control and confidence among the participants. One participant stated, *"Using vertical movement to customize the size of the virtual hand felt natural and intuitive and gave me a sense of control and confidence that I could easily navigate and select the desired objects in the VE."* The difficulties encountered with the ray casting and pinch-to-select techniques were mainly due to their lack of customization options which led to poor sense of control and confidence (Lucas, 2005; Mutasim et al., 2021), hence prevented users from adapting them to their specific needs and the constraints of the VE. In contrast, the tiny hands technique allowed for greater customization and adaptability, which resulted in participants feeling more in control and confident while using it. This resulted in significantly higher ease of use and learning for tiny hands compared to ray casting and pinch-to-select technique.

For H.4, we hypothesised that Tiny hands would be natural as compared to Raycasting and Pinch-to-select technique. Our study found that participants perceived the tiny hands technique was significantly natural compared to the ray casting and pinch-to-select techniques. Thus we accept H.4. This was primarily due to two factors: First, the technique's mental model closely resembled traditional graphical user interfaces' mouse and cursor-based selection. This familiarity made it easy for participants to understand and use the technique. One participant stated, *"It feels like using a mouse and cursor that we have been using for years. The real dominant hand feels like the mouse, which is controlling the tiny hands that act as a cursor."* Second, the significantly better naturalness of tiny hands can be attributed to its direct selection approach for object selection and appropriate mapping of small virtual hands to select small virtual objects. It made the technique intuitive for interacting with dense VEs with small objects. This finding aligns with previous studies that have demonstrated the effectiveness of direct selection and manipulation techniques in increasing naturalness in VR (Steed and Parker, 2005). In contrast, participants found ray casting and pinch-to-select less natural as these techniques adopt indirect selection approach and adopted hand as a controller, which felt less intuitive and less in line with the natural movement of their hands (Bowman and Hodges, 1997; Kopper et al., 2006). Pinch-to-select, in particular, was challenging for participants as they often associated the pinch gesture with zooming or scaling objects, as is common on mobile devices. One participant commented, *"I found it difficult to select objects using pinch-to-select because my brain kept associating the pinch gesture with zooming or scaling, like I do with images on my phone."*

For H.5, we hypothesized that Tiny hands would be the most preferred and low in effort as compared to Raycasting and Pinch-to-select technique. Our results show that tiny hands was the most preferred and reported low in effort by participants as compared to other techniques. Thus we accept H.5.

During our study, we observed that participants did not typically set their tiny hand size to the smallest value available, which was 0.1. Instead, participants typically selected a range between 0.2 and 0.5 as their preferred tiny hand size. When asked about their reasoning, participants explained that setting the hand size too small made it difficult to control and perceive the distance between the hand and the objects in the VE. One participant mentioned, *"I feel like if it's too small, I might not be able to control it properly, and it seems like it's too far away from the objects."* Another participant mentioned that they preferred to compare the target size with the hand size to ensure accurate selection, but didn't feel it necessary to make the tiny hands extremely small. He commented, *"I am comparing the target size with the hand size and confirming the tiny hand size. I feel it need not be very small to select the object accurately."*

Our study also revealed some limitations of the tiny hands technique. Participants reported that using the tiny hands required more movement than other techniques. They also noted that moving the real hand away from the user made the tiny hands appear smaller, resulting in a feeling of distance. However, this was only due to depth perception-a phenomenon in which humans underestimate the distance of distant visual objects and overestimate the distance of near visual objects (El Jamiy and Marsh, 2019). A participant commented, *"It feels like I have to move more to move the tiny hands. Moving the real hand away from the user makes the tiny hands look smaller. Hence it feels like it's far away."* A few participants suggested the need for a visualization, such as a scale, to indicate how much to vertically move to decrease or increase virtual hand size. One participant commented, *"A visualization in the form of a scale could help me understand how much to decrease and increase the hand size. I also want to understand the current hand size that I am using while using tiny hands."*

5.11 Design Recommendations

Based on the results of our two studies (chapter 4 and 5), we have identified two recommendations for designing selection techniques for small targets within arms' reach in dense VEs.

Our first recommendation, R1, is to design object selection techniques that provide an option to lock the target before selection. This is because in such dense environments, small objects are difficult to access due to their small size, increased density, and low proximity between objects. As a result, users require higher cognitive effort to visually locate and confirm their selection, which negatively impacts user performance. Designing techniques that provide an option to lock the target can improve selection accuracy and efficiency of selecting small targets within arms' reach in dense virtual environments. For instance, the LDTPT technique provides the possibility to lock the target before selection. Additionally, in this technique the locked phase time can also be extended to allow more time for the user to confirm their selection. This can be particularly helpful for medical training applications for example, in a virtual surgery simulation, a user may need to select a small artery or vein to perform a procedure. Locking the target can help ensure that the correct structure is selected, reducing the risk of errors and also help in error recovery. This, in turn, can reduce cognitive load, increase user confidence, and improve selection accuracy. Designing similar techniques have reported improved selection time and reduced cognitive load (Grossman and Balakrishnan, 2006).

Our second recommendation, R2, is to design a user-controlled method that can scale down the size of the virtual hand to match the size of the small object, allowing for more precise and efficient interaction, leading to improved selection accuracy and performance. For instance Tiny hands technique scales down the size of virtual hands by one tenth of the actual size. This enables precise selection of small targets by mapping the target size with the virtual hand size. By scaling down the size of virtual hands, users can also reduce the amount of physical effort required to select and interact with objects, leading to improved user comfort and reduced fatigue. Additionally reducing the size of virtual hands also decreases visual clutter in a dense VE. Designing similar techniques has been recommended in earlier studies, and have reported improved selection time and reduced cognitive load (Piumsomboon, 2018).

5.12 Chapter Summary

VR technology offers an immersive experience to users by allowing them to interact with VEs in a more natural and intuitive way. However, selecting small objects within arm's reach in dense VEs has been a long-standing issue. This problem arises due to the limitations of current approaches to object selection in VR, which have not adequately addressed the challenge of selecting small objects in dense VEs. To address this challenge, we present the concept of tiny hands, which allows for the selection of small objects within arm's reach in dense VEs.

The tiny hands technique enables users to reduce the size of their virtual hand by performing a gesture using their non-dominant hand. Once the desired size of the virtual hand is achieved, a user can point to the desired object to select it. The object is selected when the point intersects with the desired virtual object. Three interaction techniques were conceptualized to trigger and customize tiny hands, including (i) fist formation and wrist rotation, (ii) palm downwards placement and vertically up/down movement, and (iii) pinch-in and pinch-out. To identify a suitable interaction technique to trigger and customize tiny hands, we conducted an informal study using a think-aloud method. Among the three interaction techniques, the gesture of non-dominant hand palm facing downwards for 2 seconds and its vertical upward/downward movement was finalized due to its actual mapping of the interaction technique to trigger tiny hands.

To evaluate the tiny hands technique, we conducted a second study and compared it with existing techniques from the literature, including (i) ray casting and (ii) pinch-to-select. The study found that participants preferred the tiny hands technique over the other two techniques. Participants reported that the tiny hands were easy to use, learn, and natural due to their mental model association with mouse and cursor-based selection and the mapping of small virtual hands to small virtual objects. They also appreciated the customization options available with the tiny hands technique, which allowed them to feel more in control and confident while using it. On the other hand, participants found ray casting and pinch-to-select less natural and less intuitive. This is because they involved indirect selection approach and adopted hands as controllers instead of direct selection. Participants also had difficulty associating the pinch-to-select gesture with object selection, as it is often used for scaling objects in traditional graphical user interfaces.

We believe that the tiny hands technique offers a promising solution to the problem of selecting small objects of arm's reach in dense VEs. It provides a natural and intuitive way of interacting with small objects within arm's reach in dense VEs. The tiny hands technique presents a potential to enhance the user's immersive experience in VR by enabling them to select small objects in dense VEs naturally, intuitive and efficiently. Additionally, we have extracted recommendations that can help the design of object selection techniques for small object selection within arms' reach in dens virtual environments.

In the next chapter, we present the design and evaluation of the gesture identified for VE 2. We present a novel technique, **AMAZE** (**A** **M**ulti-finger **A**pproach to **Z**oom in dense **E**nvironments) for selection of small objects placed at a distance (beyond arm's reach) in a dense VE.

Chapter 6

6. Design and Evaluation of AMAZE Technique to Select Small, Distant Objects in Dense Virtual Environments

Chapter 3 described a user-centered elicitation study that was conducted to investigate small object selection in two different virtual environments. The first environment was a Dense VE with small targets placed within arm's reach (VE1), and the second was a Dense VE with small targets placed at a distance (VE2). The final gesture chosen for VE1 was *point and tap* gesture and for VE2 was *pinch out the VE and point and tap gesture*. In chapter 4 we redesigned the gesture identified for VE1 considering the specific challenges of the environment and renamed it *Locked Dwell Time based Point and Tap (LDTPT)*. We performed a comparative study of the LDTPT technique with raycasting, magnetic grasp and innate pinch techniques identified from the literature. The results suggest that the LDTPT technique was the fastest and produced low errors for selecting small targets at arm's reach. It was also the most natural technique, most preferred among the participants and low in effort. However, the LDTPT technique was not easy to use and learn considering the feedback visualization provided and the projection of users hand for small object selection. We further improved upon LDTPT technique and explored the possibility of decreasing the hand size for the selection of small targets in dense VE where targets are positioned within arms' reach. In chapter 5, we came up with the concept of *Tiny hands* which reduces the size of virtual hands to select small targets. We compared tiny hands technique with raycasting and pinch-to-select techniques from the literature. Our results show that tiny hands was significantly faster and more accurate in selecting small targets as compared to other techniques. The tiny hands technique was also significantly easy to use, learn, natural and playful. It was the most preferred technique and exhibited less effort during the study. Chapters 4 and 5 focused on the design and evaluation of two new object selection techniques specifically tailored to VE1.

In this chapter, we present the design and evaluation of the gesture identified for VE2, Dense VE, where targets are small and placed at a distance. The finalized gesture for VE 2 was *pinch out the VE and point and tap*. We adopted the elucidated gesture to redesign it to suit the challenges of VE2. We present a novel technique, **AMAZE** (**A** **M**ulti-finger **A**pproach to **Z**oom in dense **E**nvironments) that offers zoom using multiple fingers in a dense VE where objects are small and placed at a distance (beyond arm's reach). The objective is to enable accurate, error-free and faster distant object selection of small objects in dense VE. First, we present the design rationale for the AMAZE technique in section 7. The design details of the AMAZE technique are elaborated in section 6.1. The design rationale for AMAZE is presented in 6.1.1. Design details of AMAZE is presented in 6.1.2. After this, we present two studies aimed at evaluating AMAZE to investigate its accuracy and task completion times with existing selection techniques. In the first study, we evaluate AMAZE with two techniques from the literature: Expand and Pinch-to-Select. The study details are presented in section 6.2. Section 6.2.1- 6.2.8 presents hypothesis, baseline techniques, VE design, participants, study set-up and apparatus, tasks, study procedure, Data-collection method. Section 6.3 presents the results. The findings are discussed in section 6.4. Our results indicate that the AMAZE technique was significantly accurate, natural and playful to use. However, it took higher selection time as compared to other techniques. With our analysis supported by participants' comments, we redesigned AMAZE and presented AMAZE-X. Section 6.5 presents the design of AMAZE-X technique. We improved upon AMAZE-X's zoom factor with the increased CD ratio of 1:5 and added subtle dash lines as visual elements during the corrective phase to improve distance to the user for faster selection Section 6.6 presents the evaluation of AMAZE-X with techniques from study 1. The results of the second study are presented in section 6.7. We found that AMAZE-X was significantly faster, easier to perform and easier learn than other techniques. We also found AMAZE-X to be the most preferred technique. The findings are discussed in section 6.8. We also provide a set of design recommendations elaborated in section 6.9. Finally, we summarize the chapter in section 6.10. The development of AMAZE and AMAZE-X is an important step toward improving the selection of small, distance targets in dense virtual environments. In addition, from the two experiments, we have extracted recommendations that can help the design of object selection techniques for 3D virtual environments.

6.1 Design of AMAZE: Object Selection Technique for Small and Distant Objects in a Dense VE.

6.1.1 AMAZE: Design Rationale

In order to address the challenge of selecting small objects located at a distance within a dense virtual environment (VE2), a user-centered elicitation study was conducted to identify an effective gesture for object selection. The chosen gesture, named "pinch out the VE and point, and tap," involved pinching out the virtual environment with the index finger and thumb to zoom in on the target. Once the target was close, the user would point to indicate selection and then tap to confirm. However, this gesture required multiple pinches and an additional tap for confirmation, which could be cumbersome and time-consuming. Therefore, we redesigned the gesture by upgrading the pinch options to include multiple fingers, allowing for faster zoom-in. We also removed the tap function for final confirmation as it was deemed unnecessary for objects within arm's reach.

This modified gesture, which we named **AMAZE** (**A** **M**ulti-finger **A**pproach to **Z**oom in Dense **E**nvironments), was designed to enable accurate and efficient selection of small objects in a dense VE2. The AMAZE technique is illustrated in figure 27. The proposed technique enables infinite zoom and uses a pinch gesture to bring small targets closer for selection. Users can employ their non-dominant hand's index finger and thumb to zoom in/out the target in the virtual environment (VE). Additionally, users can use one, two, three, or four fingers to move the selected object for precise placement and accurate selection (Figure 27b). Once zoomed in, the objects are always within the user's field of view (FOV) and maintain their relative positions to each other. This close-up view of the targets occupies a larger portion of the screen, allowing for faster and more accurate selection. The user can precisely bring the target closer using a 1-finger pinch gesture (Figure 27c), and once the target is in close view, the user can point to it using their dominant hand's (right hand’s) index finger for final selection (Figure 27d).

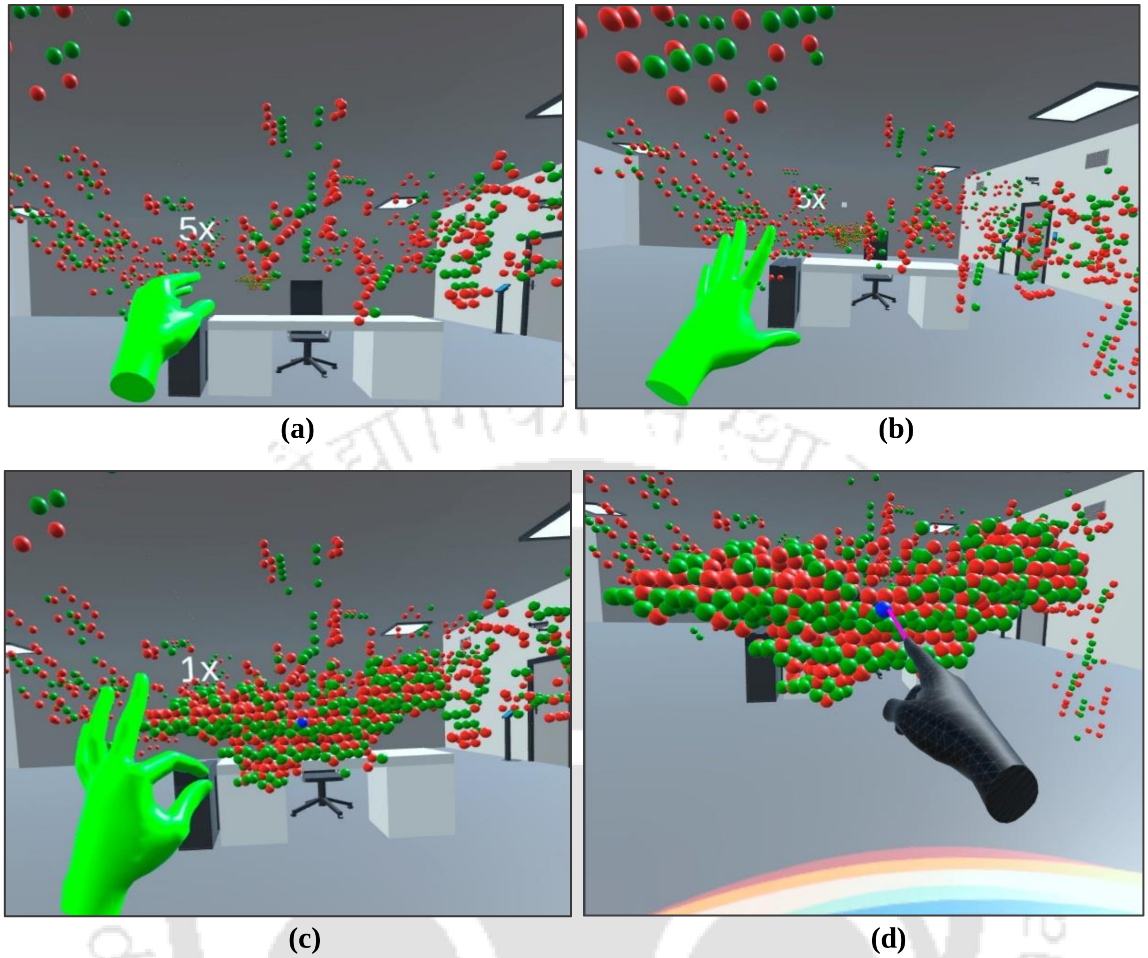
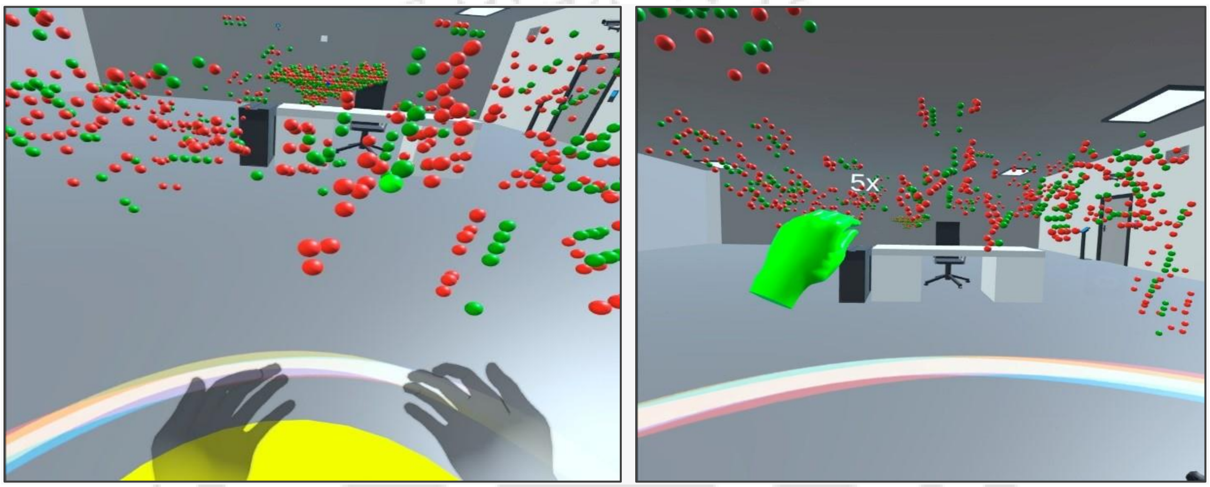


Figure 27: Selection using AMAZE technique: (a) User initiates the gestures (b) User performs pinch-out with five fingers which zoom the target at 5x zoom. This step is done multiple times to bring the target at a near distance from the user. (c) The user performs one-finger zoom (1x) to granularly place the target near the user. (d) User points using the right-hand index finger to select and confirm the target.

6.1.2 Amaze: Design Details

The AMAZE technique follows a two-stage selection process that incorporates (i) multi-finger pinch in/out and (ii) point-to-select actions. The primary objective is to utilize multi-finger pinch to bring objects within users' hand reach for faster and more accurate object selection. Pinch-in is used to bring objects within reach, while pinch-out zooms out the objects. After bringing the desired object within hand reach, the user can select it by pointing at it. The selection process is confirmed when the point intersects with the target object.

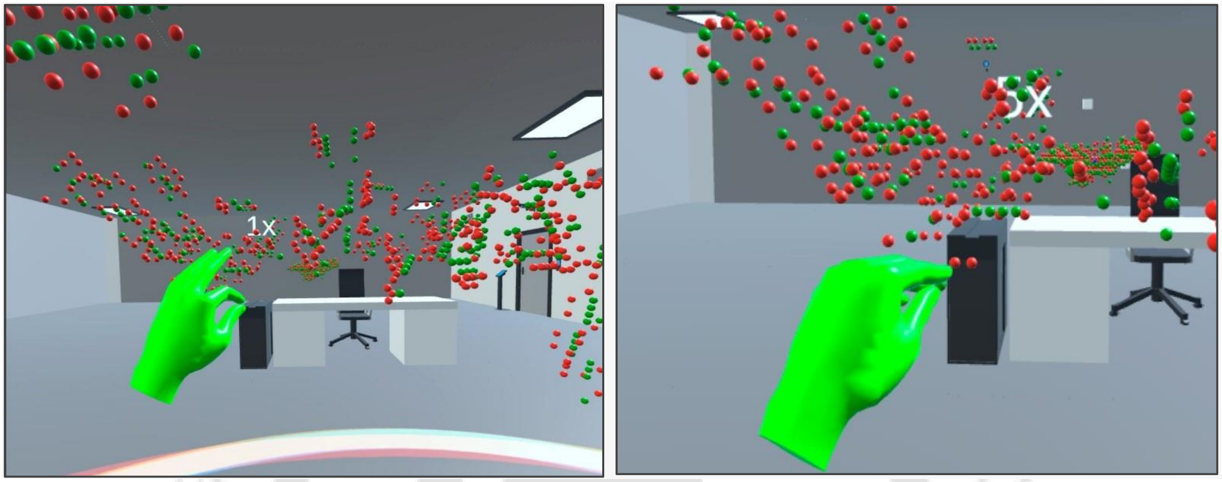
In our design, we created a spherical effective region with a white color that is 10 cm in diameter and centered at the user's hip region, as seen in Figure 28. The purpose of this region is to provide users with an unobstructed space to perform the pinch-in/out gestures. To initiate these gestures, the user must extend their hand outside of the effective region. When the user extends their hand outside of the effective region, the virtual hands change color from black to green to provide visual feedback and help the user better understand their hand position. Figure 28 (a) and (b) depict the virtual hands inside and outside of the effective region, represented in black and green color, respectively. Additionally, the effective region helps to avoid technical limitations in commercial devices such as Oculus VR and Leap Motion, which have limited capacity for tracking gestures within a 10cm radius of the user's position.



(a) **(b)**

Figure 28. (a) The effective region (white in colour) is designed to initiate the gesture and to avoid unintentional gestures. The virtual hand colour is black when hands are inside the effective region. (b) The virtual hand colour changes to green when user goes out of the effective region.

AMAZE uses multi-finger pinch-in/out gesture to control the zoom-in/out of objects in the VE. A continuous zoom metaphor is employed, where the number of fingers used in the pinch determines the amount of zoom-in/out performed. For instance, using the index finger and thumb results in 1x zoom, while using two fingers and the thumb results in 2x zoom, and so on up to 5x zoom when five fingers are used. To control the zoom speed, we designed two operating regions around the user, at distances of 30cm and 50cm from the user's hand reach, respectively. The first operating region provides a 20cm space to perform the multi-finger pinch-in/out gesture, and the zoom-in/out speed is set to 1m/s. In contrast, the second operating region is an additional 30cm away and allows for a faster zoom-in/out speed of 5m/s. The system provides visual feedback to the user regarding the number of fingers used and the target zoom level through a text-based message displayed on the screen (Figure 29 (a) and (b)). This design provides users with flexibility in selecting objects at faster speeds and caters to both novice and advanced users. Figure 30 shows the visual representation of the system's feedback to the user.



(a) **(b)**

Figure 29: Text-based representation of zoom-in function (a) one finger zoom-in is visually presented to 1x (b) all finger zoom-in is indicated using 5x.

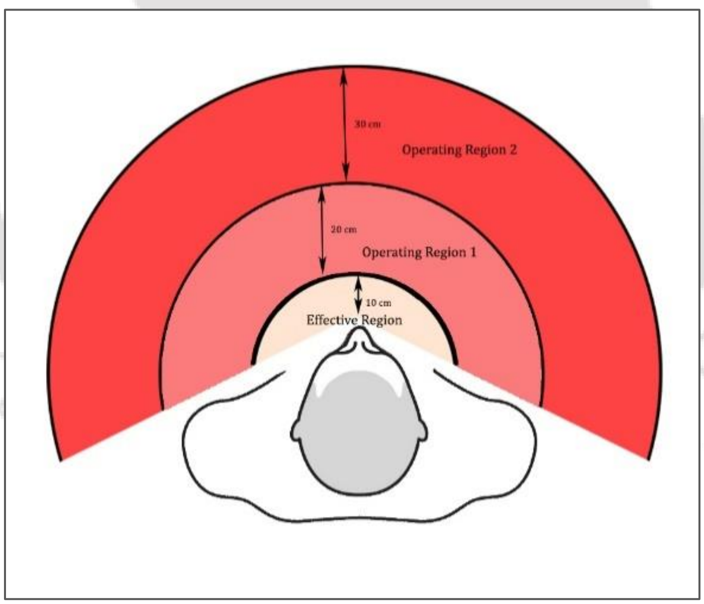


Figure 30: The effective region (10 cm in radius) is an unobstructed space created to perform the pinch-in/out gestures, and two operating regions designed to control the speed of the zoom-in. Operating region 1 (20 cm) zooms at a rate of 1 m/s and operating region 2 (30 cm) zooms in at a speed of 5m/s.

6.2 Evaluation of AMAZE

This section provides a comprehensive overview of the evaluation study conducted to assess the effectiveness of the AMAZE technique. We first introduce the baseline techniques used for the evaluation study, followed by a description of the hypothesis. Subsequently, we outline the study details, including the environment design, participants, study setup and apparatus, methodology, data collection method, and study results.

To evaluate the efficacy of AMAZE, we compared it with two baseline techniques, namely Expand and Pinch-to-select. The primary objective of the study was to investigate the accuracy, task completion time, ease of use, ease of learning, and naturalness of the AMAZE technique in comparison to the baseline techniques.

6.2.1 Research Hypothesis

For this experiment, we formed four hypotheses:

H1: It is hypothesized that the AMAZE technique will result in faster selection times than the Expand and Pinch-to-select techniques when selecting small and distant targets in a dense virtual environment.

H2: It is hypothesized that the AMAZE technique will result in a lower error rate than the Expand and Pinch-to-select techniques.

H3: It is hypothesized that the AMAZE technique will be perceived as easier to learn and use compared to the Expand and Pinch-to-select techniques.

H4: It is hypothesized that the AMAZE technique will be perceived as more natural than the Expand and Pinch-to-select techniques

6.2.2 Baseline Techniques

In order to validate the effectiveness of the AMAZE technique, we conducted a comparative study with two existing techniques from the literature: the Expand technique (Cashion et al., 2012) and the pinch-to-select technique (Oculus, 2022). The Expand technique is a refinement of the SQUAD technique, which is a popular method for selecting targets in dense VEs. It utilizes a multi-stage zoom approach to identify and select a target. The pinch-to-select technique, on the other hand, is widely used in advanced VR and mixed reality platforms such as the Meta Quest VR, Microsoft Hololens etc. HMD based interfaces. Given their widespread usage in the field, we selected these two techniques for the study. We provide a detailed explanation of these techniques below.

The Expand technique is a selection technique that builds upon the SQUAD selection technique (Kopper et al., 2011). In the Expand technique, objects are selected using a sphere-casting metaphor, where a virtual sphere is cast from the user's hand to the objects in the VE. The objects are then rearranged in a virtual grid, allowing the entire screen to be utilized. In the second stage, the user selects the target from the grid. Once a target has been selected, the objects are automatically transitioned back to their original location. This technique is commonly used for selection in dense VEs and is claimed to be useful for scenarios where objects are tightly clustered together.

The pinch-to-select technique is a widely used selection technique in advanced head-mounted display (HMD) based virtual reality (VR) and mixed reality (MR) platforms such as Meta Quest VR and Microsoft Hololens (Kress et al., 2017). This technique utilizes a pinch metaphor for object selection. When the user performs a pinch gesture, a floating cone visualization appears, and the size of the circular cursor used for selection increases or decreases based on the amount of pinch. The pinch gesture can also be used to control the pointing direction of the cursor. When the user pinches their fingers together, it squeezes the cone, and the cursor then acts like a laser pointer that can be used to make selections on a distant screen. The use of a pinch metaphor for selection is claimed to provide a natural and intuitive interaction technique for users (Kang et al., 2020; Pfeuffer et al., 2017).

6.2.3 Design of VE

The VE was developed using the Unity 3D game engine and comprised a dense molecular structure consisting of 1192 spheres. The molecular structure was designed to resemble a small-molecule compound used in molecular modeling by biologists to assess potential effects on humans. The spheres were colored red and green to serve as distractors, while the target object was colored blue. Each sphere could be selected as an individual target. The static spheres were 0.5 cm in diameter. The VE was room-scale sized, with dimensions of 20 x 15 x 11 ft, and was located at a distance of 4.9 m from the participant. Upon target selection, audio feedback was provided to inform the participant that they had successfully selected the target. Additionally, visual feedback was provided by changing the color of the target object from blue to green. Once the target was selected, it would disappear, and the zoomed-in region would return to its original location. The next target was then indicated in red for selection.

6.2.4 Participants

18 non-paid participants (10 males, 8 females) between the ages of 18-33 years (mean= 26.42, SD= 4.17) were recruited for the study. Only right-handed participants were included as prior research has shown a strong influence of handedness on spatial perception (Plaumann et al., 2013). All participants were university students enrolled in design courses and reported prior experience using HMD-VR platforms (e.g., Oculus Rift and HTC Vive) for approximately 10 hours within the last six months.

6.2.5 Study Setup and Apparatus

The study employed the Oculus Quest HMD-VR device for data collection, which comes equipped with a built-in camera capable of detecting the positioning, orientation, and finger configuration of the hands. A computer with an i7-8700 quad-core processor, Nvidia Geforce 1060 GPU, 8GB RAM, and Microsoft Windows 10 operating system was utilized to facilitate the connection. In order to capture the participants' gestures, a video camera was positioned diagonally to their location. A moderator was also present during the study, stationed diagonally to the participants in order to observe and manage the proceedings. Figure 31 shows the experimental setup used for the study.

Figure 31: The experimental setup used for the study. For the AMAZE technique, the user performs the pinch-out gesture to zoom in on the VE with a non-dominant hand.

6.2.6 Task

Participants were instructed to complete a set of five object selection tasks, with the objective of selecting the blue target located within the molecular structure. Upon successful selection, an audio cue was triggered indicating the selection status (i.e., "you have selected the target"), while a visual cue was presented by changing the color of the target object from blue to green. After each selection, the target object would disappear, the zoomed-in region would revert to its original location, and the subsequent target location would be highlighted in red, prompting participants to proceed with the next selection task. In total, participants were required to select 15 targets, with each set of five targets corresponding to one task.

6.2.7 Study Procedure

The study was conducted in a controlled laboratory setting at a university. A verbal introduction to the procedures and techniques was given by the moderator to each of the 18 participants, who were subsequently trained on all three techniques until they felt confident in their ability to use them. The training period for each participant lasted approximately 30 minutes. Following this, participants were randomly assigned one of the three techniques and instructed to complete a set of five object selection tasks. The task involved selecting a blue target located within a dense molecular structure. Audio and visual feedback were provided upon successful target selection. Post-task interviews were conducted to collect subjective feedback on the overall perception of the techniques, and were video recorded with participants' consent for further analysis. The study lasted approximately 90 minutes per participant.

6.2.8 Data Collection Method

The experiment involved 18 participants, who completed a total of 270 trials (3x5x18) using three techniques (AMAZE, Expand, Pinch-to-select) to select 5 targets. The application recorded the completion time and error rate for each trial. The task completion time was calculated after participants pressed a red button placed near the feet of the user on the VE. The study began when the target changed color from red to green, triggered by the user pressing a red button near their feet in the VE. An error was counted when the target selected by the user did not match the target specified during the task. The moderator tracked the number of pinches, including the number of fingers used, by the participants to zoom in on the VE. Additionally, the moderator administered a 7-point semantic differential scale to gather feedback on the ease of use, ease of learning, and naturalness of each technique. Participants were provided with a printed paper to collect their responses on ease of use, learning and naturalness. Participants ranked their preference for each technique at the end of the study, and the post-experiment interviews were recorded for further analysis.

6.3 Results

To collect user performance data, we employed system logs and user preference questionnaires. The system logs recorded information pertaining to task completion time and selection errors. We conducted the Shapiro-Wilk test to assess normality of the data. For normally distributed data, we performed a one-way ANOVA test with a Bonferroni correction to determine statistically significant differences. For the Likert-scale data, we used the non-parametric Friedman test with Wilcoxon Signed Ranks post-hoc tests.

6.3.1 Task Completion Time

A one-way ANOVA was conducted to examine the differences in task completion time across the three techniques (AMAZE, Expand, and Pinch-to-select) at a significance level of 0.05. The results indicated a statistically significant difference in task completion time across the techniques ($F(2,51)=46.47$, $p=.001$). Post hoc analyses using Tukey's test showed that the task completion time for Pinch ($M=66.34$, $SD=22.89$) was significantly faster than both AMAZE ($M=127.34$, $SD=16.11$) and Expand ($M=121.83$, $SD=23.28$) ($p<0.05$). There was no significant difference between Expand and AMAZE ($p>.05$). Figure 32 shows the mean and SD values for task completion time (in seconds) for the techniques AMAZE, Expand, and Pinch-to-select.

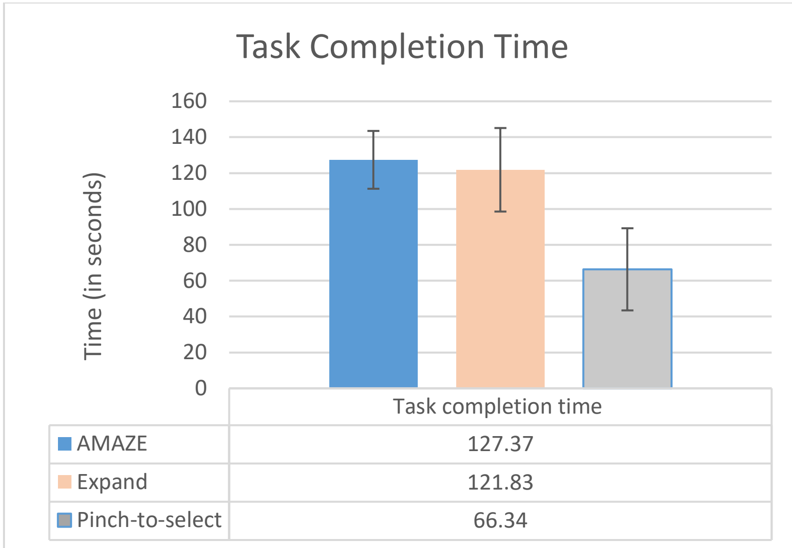


Figure 32: The mean and SD values for task completion time (in seconds) for the techniques AMAZE, Expand, and Pinch-to-select

6.3.2 Error Rate

We conducted ANOVA to examine the differences in error rate across the three techniques at a significance level of 0.05. The post hoc Tukey test indicated that the error rate for Pinch (M=1.72, SD=1.48) was significantly higher than both AMAZE (M=0, SD=0) and Expand (M=0, SD=0) ($p<0.05$). Notably, AMAZE and Expand techniques resulted in zero errors. Figure 33 shows the mean and SD values for error rate for the techniques AMAZE, Expand, and Pinch-to-select.

Figure 33: The mean and SD values for error rate for techniques AMAZE, Expand and Pinch-to-select

6.3.3 Easy to Use

A Friedman test was performed to compare the ease of use ratings among the three techniques. The test revealed a statistically significant difference in ease of use ratings among the techniques ($\chi2(2)=15.136$, $p=0.001$). Wilcoxon Signed Ranks post-hoc tests were conducted to further analyze the differences between the techniques. The results indicated that there were significant differences in the ease of use ratings between Pinch-to-select and AMAZE ($Z=-2.80$, $p<0.05$), as well as between Pinch-to-select and Expand ($Z=-2.97$, $p=0.003$). However, no significant difference was found between AMAZE and Expand ($Z=-1.83$, $p>0.05$). Figure 34 shows the mean and SD values for easy to use for the techniques AMAZE, Expand, and Pinch-to-select.

6.3.4 Ease of Learning

A non-significant result was obtained for the ease of learning variable, as determined by the Friedman test ($\chi2(2)=3.96$, $p>0.05$). Wilcoxon Signed Ranks post-hoc tests further revealed that there were no significant differences in ease of learning between AMAZE and Expand ($Z= -0.81$, $p>0.05$), Expand and Pinch-to-select ($Z=-1.94$, $p>0.05$), and AMAZE and Pinch-to-select ($Z=-1.53$, $p>0.05$). Figure 32 shows the mean and SD values for ease of learning for the techniques AMAZE, Expand, and Pinch-to-select.

6.3.5 Naturalness

We conducted a Friedman test to compare the naturalness of all the techniques, which showed a significant difference, $\chi2(2)=21.75$, $p=0.001$. Further Wilcoxon Signed Ranks post-hoc tests revealed significant differences in naturalness between AMAZE and Expand ($Z= -3.38$, $p<0.05$), as well as between AMAZE and Pinch-to-select ($Z=-3.41$, $p<0.05$). However, there was no significant difference between Expand and Pinch-to-select ($Z=-0.04$, $p>0.05$). Figure 32 shows the mean and SD values for naturalness for the techniques AMAZE, Expand, and Pinch-to-select.

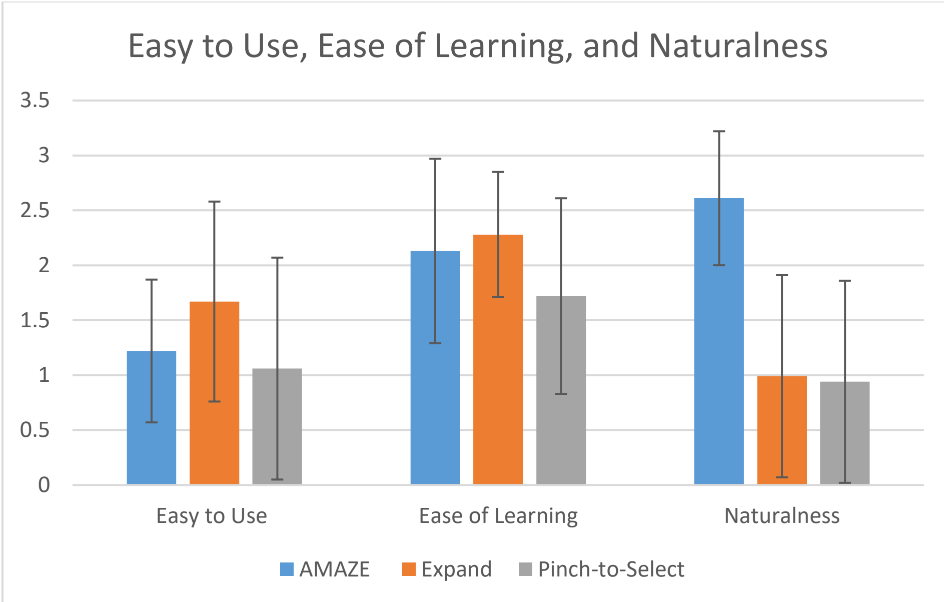


Figure 34: The mean and SD values for ease of use, ease of learning, and naturalness for techniques AMAZE, Expand and Pinch-to-Select

6.3.6 Preference

Out of the 18 participants, 16 (82%) indicated AMAZE gesture as their top preference. In contrast, 10 participants chose pinch-to-select as their second preference. 2 and 8 participants chose expand as their first and second preference.

6.4 Discussion

In this section, we present the results, hypothesis validation, and other findings of our study. Our investigation focused on comparing task completion times, error rates, ease of use, learning, naturalness, and user preference across three object selection techniques for small and distant objects within a dense VE. The three techniques we tested were AMAZE, Expand, and Pinch-to-select. We will discuss the statistical results, validate our hypotheses, and provide a comprehensive overview of the subjective findings of the study.

H.1 predicted that AMAZE would result in faster task completion times compared to Expand and Pinch-to-select techniques. However, statistical analysis revealed that AMAZE took significantly longer to select small, distant objects compared to Expand and Pinch, leading to the rejection of H.1. Three primary reasons contributed to the higher task completion times for AMAZE. Firstly, participants needed to zoom in multiple times (7-9 times) to bring the objects within hand reach, which took up to approximately 50 seconds. A participant commented, *"I might* *have zoomed in multiple times just to bring the objects within my arms reach. This process took up a lot of time and and it was also frustrating to have to constantly zoom, but it was necessary in order to bring the target near."* Secondly, the initiation gesture of zooming in and out by bringing the hand inside and out of the effective region also added to the slower performance. Effective regions were not present in other techniques, and the time taken to cross them contributed to the higher task completion times. A participant commented, *"I feel the initiation gesture by bringing my hand inside and out of the effective region actually added to a slower performance. It also took some time for me to perform zoom-in and again go in the effective region and perform zoom-out."* Lastly, participants spent more time in the ballistic phase to bring the targets near and make minor adjustments. These findings align with previous research by Liu and Liere, (2009), who reported longer times taken during the ballistic phase. In contrast, Pinch emerged as the fastest technique for object selection tasks, as it is a one-step technique that requires participants to select targets quickly with a pinch metaphor. Pinch eliminates the need to zoom in, saving time and enabling faster completion times. The Expand technique required a two-stage selection process, where participants had to select a group of targets placed in a virtual grid in the second step. Participants reported that this process caused them to lose the original context of the VE and spend more time searching for the target in the virtual grid before selecting it. This led to higher task completion time than pinch-to-select technique.

For **H.2**, our hypothesis was that AMAZE and Expand techniques would result in lower error rates compared to Pinch-to-select technique. The results showed that AMAZE had no errors in object selection, while Expand also offered error-free selection. Thus, we partially accept H.2. Participants using AMAZE mostly used 1-finger and 2-finger pinch-in to make micro-adjustments, bringing targets within precise hand reach distance for accurate selection. This resulted in no errors, as targets were accurately placed within arm's reach with better visibility and accessibility. A participant commented, *"I used 2-finger and then 1-finger to zoom in and adjust the targets near me. I was able to accurately place the target where I wanted it to be without having to zoom out. Being able to see the targets clearly and in the exact location where I want it to be helped me to select the target accurately."* Further, it reduced the need to use the fine motor skills or precise selection for object selection in a dense VE. Previous studies have highlighted the effectiveness of the ability to zoom-in and bring objects closer to the user's arm's reach in improving visual search performance and reducing the reliance on fine motor skills to achieve better performance (Mendes et al., 2017). In the Expand technique, the participants were required to select a group of targets placed in a virtual grid in the second step. This reduced the density of the objects, which resulted in error-free selection for most participants. This finding is consistent with previous studies that have shown that selecting objects from a grid layout can help users locate target more easily (Kopper et al., 2011; Cashion et al., 2012) and reduce errors (Bowman et al., 1997; Cashion et al., 2012). In contrast, Pinch-to-select technique was found to be error-prone, with half of the participants committing at least three errors. The occurrence of errors during pinch-to-select object selection technique was attributed to hand jitter and Heisenberg selection, a phenomenon where the participants used the circular cursor to indicate the target while simultaneously maintaining the pinch gesture and confirming the selection through a pinch. The literature reveals that similar issues of hand jitter and Heisenberg selection have been reported in earlier studies concerning object selection tasks in VR.

We reject **H3,** as our results indicate that AMAZE was not easier to learn and use compared to the other two selection techniques. Participants reported that the AMAZE gestures, particularly the pinch-in gesture, required them to bring their hand out of the effective region and back in to initiate the zoom-out gesture, which was challenging to remember for initial interactions. Additionally, participants found it difficult to recall the two operating regions for speed control (30cm region - operating region 1and 50cm region - operating region 2), and the sudden changes in these two regions led to confusion among the participants when migrating from one region to another. To make AMAZE easier to learn and use, participants suggested removing the separate operating regions. A participant commented, *"I found it difficult to remember gestures required to initiate the zoom-in/out, particularly during the initial interactions. Having to bring my hand out of the effective region and back in to initiate the zoom-out gesture was not intuitive and took some time to get used to."* Another participant commented, *"I found it difficult to recall the two different operating regions for speed control, which added to the overall complexity of the zooming process."*

In contrast, participants found the pinch-to-select technique difficult to use due to the requirement of holding the pinch position to select small targets using a cursor, and the need to precisely point the cursor to select small targets in a dense VE. The difficulty in holding the pinch position and difficulty in precisely pointing the cursor in pinch-to-select have been report earlier as well (Mutasim and Stuerzlinger, 2021). However, Expand was the easiest technique to use and learn as participants only needed to point to select objects. The selected group of objects were then placed in a virtual grid in the second step, making it easier for participants to select targets as they were aligned in proximity.

We found support for hypothesis 4 as the AMAZE technique was perceived as significantly more natural than the other two techniques. This could be attributed to two factors. Firstly, the AMAZE technique was derived from a user-generated gesture study which involved identifying and selecting gestures that were intuitive and familiar to users. This may have resulted in the AMAZE technique being perceived as more natural by users. A participant commented, *"Even in real life we go near or being the object near to select it. We also often use pinch to select a small object. This was possible using the AMAZE technique"* Secondly, the use of pinching as a scaling gesture is a well-established interaction metaphor in mobile devices, which participants were accustomed to. This adoption of their existing mental models have further contributed to the perceived naturalness of the AMAZE technique. Several studies have highlighted the importance of user familiarity and intuition in the design of natural user interfaces (Norma, 1988; Nielsen, 1994). In the case of the pinch-to-select technique, the pinch gesture is commonly associated with scaling and zooming in mobile devices, where it is used to resize images, maps or web pages. Hence, using the same gesture to hold the gesture to point and select objects in a VE seemed unnatural to the participants, as it deviated from their existing mental model of the pinch gesture. The Expand technique was also perceived as unnatural by the participants, mainly due to the lack of a real-world counterpart for the conversion of virtual objects into a grid after selection. The unfamiliarity of this interaction paradigm, which deviates from conventional physical interactions, led to a perception of not being natural.

6.5 AMAZE-X: Enhancing AMAZE Technique

The AMAZE technique demonstrated superior accuracy and speed compared to the Expand and Pinch-to-select techniques, and was also rated as the most natural. However, it was found to be difficult to use and learn, primarily due to the presence of two operating regions (20cm and 30cm) that led to confusion among participants and increased the learning curve. Feedback from participants suggested that eliminating these operating regions would improve the ease of use and learning for the AMAZE technique. Additionally, participants often had to zoom in multiple times during the ballistic phase to place the target within arms' reach, which increased the overall task completion time. The effective region designed for zooming in and out by bringing the hand inside and outside of the region also contributed to the slower speed of the technique. Participants also reported difficulty in perceiving depth during the corrective phase, often placing the target too close and requiring zooming out, which further increased the selection time.

Based on our analysis of the factors affecting the usability of AMAZE, we have redesigned the technique to improve its speed, ease of use, and ease of learning, particularly in scenarios where targets are small, positioned at a distance, and in a dense VE. We call it AMAZE-X. Two key design changes were implemented in AMAZE-X to achieve these goals. Firstly, modifications were made to the zooming process during the ballistic phase to reduce the number of iterative steps required to place the target within arm's reach. Specifically, the motor space was scaled by increasing the control-display ratio up to 1:5 during the ballistic phase. Secondly, a visual element in the form of dashed lines was added to the corrective phase to help users accurately infer distances in a dense configuration (Figure 35). We believe that this will aid in improving selection time during the corrective phase. We believe that the addition of this visualization will significantly reduce task completion times by accurately conveying the distance between the user and the target. Figure 10 displays the visualization of dashed line added in AMAZE-X.

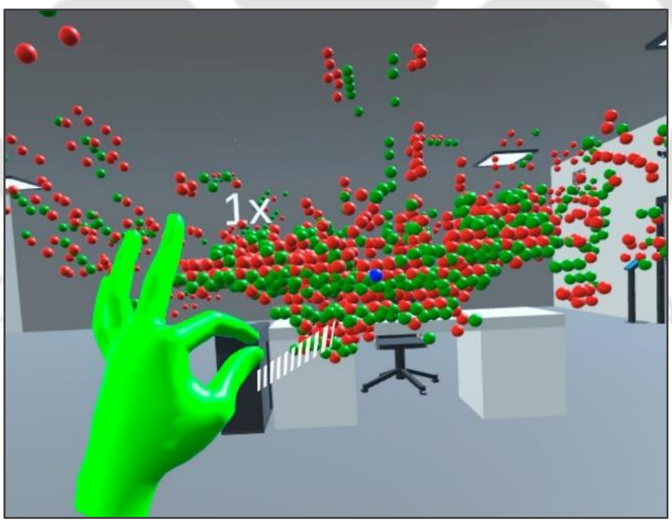


Figure 35: An additional visual element (in white) during the corrective phase indicates the distance to the user in the design of AMAZE-X.

6.6 Evaluation of AMAZE-X

The aim of this study is to assess the efficacy of AMAZE-X in comparison to other techniques, as well as to examine its ease of use and learnability relative to baseline techniques.

6.6.1 Hypotheses

H1: It is hypothesized that the AMAZE-X technique will result in faster selection times than AMAZE, Expand and Pinch-to-select techniques when selecting small and distant targets in a dense virtual environment.

H2: It is hypothesized that the AMAZE-X technique will result in a lower error rate than AMAZE, Expand and Pinch-to-select techniques.

H3: It is hypothesized that the AMAZE-X technique will be perceived as easier to learn and use compared to AMAZE, Expand and Pinch-to-select techniques.

H4: It is hypothesized that the AMAZE-X technique will be the most preferred than AMAZE, Expand and Pinch-to-select techniques

6.6.2 Baseline Techniques

For the purpose of comparison, we selected three techniques from study 1, AMAZE, Expand, and Pinch-to-select, as baseline techniques against which the effectiveness of AMAZE-X was evaluated.

6.6.3 Participants

In this study, a sample of 12 participants (7 males, 5 females) aged between 18 and 33 years (Mean= 22.54, SD= 3.87) was recruited. None of the participants had participated in Study 1. The participants were selected from among university students who were enrolled in design courses. All participants had previous experience using HMD-VR platforms such as Oculus Rift and HTC Vive.

6.6.4 Design of VE and Study Procedure

The VE utilized in this experiment was constructed with Unity 3D game engine and consisted of 1192 spheres arranged to represent a small-molecule compound used in molecular modeling. The target object, which was colored blue, was placed at a distance of 4.9 meters from the participant, while the distractor objects were colored red and green. All techniques were tested with identical target placements. Audio feedback was provided to the participants upon target selection, and a visual cue, indicating the change in color from blue to green, was also presented. The experimental design and procedures used in this study were identical to those employed in the previous study that evaluated the efficacy of the AMAZE technique.

Prior to the experiment, the moderator provided a verbal introduction of the study procedure and the four techniques to each participant. Following the introduction, training sessions on all three techniques were conducted with each participant. The duration of the training sessions for each participant was approximately 30 minutes. Participants were then asked to perform the task, with each technique being randomly assigned. Post-task interviews were conducted to obtain feedback regarding the participants' perception of the proposed techniques. These comments were recorded, and further analysis was requested by the moderator. The experiment was video recorded and lasted approximately 90 minutes per participant.

6.6.5 Data Collection Method

The experiment involved 12 participants, who completed a total of 240 trials (4x5x12) using three techniques (AMAZE-X, AMAZE, Expand, Pinch-to-select) to select 5 targets. The application recorded the completion time and error rate for each trial. The task completion time was calculated after participants pressed a red button placed near the feet of the user on the VE. The study began when the target changed color from red to green, triggered by the user pressing a red button near their feet in the VE. An error was counted when the target selected by the user did not match the target specified during the task. The moderator noted the hand size selected by the participant. Additionally, the moderator administered a 7-point semantic differential scale to gather feedback on the ease of use, ease of learning, and naturalness of each technique. Participants were provided with a printed paper to collect their responses on ease of use, learning and naturalness. Participants ranked their preference for each technique at the end of the study, and the post-experiment interviews were recorded for further analysis.

6.7 Results

For this experiment, we collected user performance data using system logs and user preference data through questionnaires. The system logs recorded task completion times and any errors made by the participants. As in the previous study, participants rated their perception of ease of use and learning on a 7-point Likert scale. In addition, we gathered user preference and effort data, which were ranked.

6.7.1 Task Completion Time

We conducted an analysis of variance (ANOVA) to examine the differences in task completion time across all four techniques, using a significance level of 0.05. The results showed a statistically significant difference in task completion time across all four techniques ($F(3,44)=111.01$, $p<0.001$). A Tukey post-hoc test was conducted to further investigate pairwise differences between the techniques. The results indicated that the task completion time for AMAZE-X ($M=127.34$, $SD=16.11$) was significantly faster from that of AMAZE ($M=127.34$, $SD=16.11$) ($Z=-3.05$, $p<0.01$), Expand ($M=66.34$, $SD=22.89$) ($Z= -3.67$, $p<0.05$), and Pinch-to-select ($M=66.34$, $SD=22.89$) ($Z=-2.21$, $p<0.05$) techniques. The task completion time for Expand ($M=121.83$, $SD=23.28$) was significantly faster from that of Pinch-to-select ($M=66.34$, $SD=22.89$) ($Z=-4.31$, $p<0.05$) technique. Notably, AMAZE-X was the fastest technique with an average completion time of 40.35 seconds, followed by Pinch-to-select (avg=59.1 s), Expand (avg=116.53 s), and AMAZE (avg=119.23 s). Figure 36 shows the mean and SD values for task completion time for AMAZE-X, AMAZE, Expand, and Pinch.

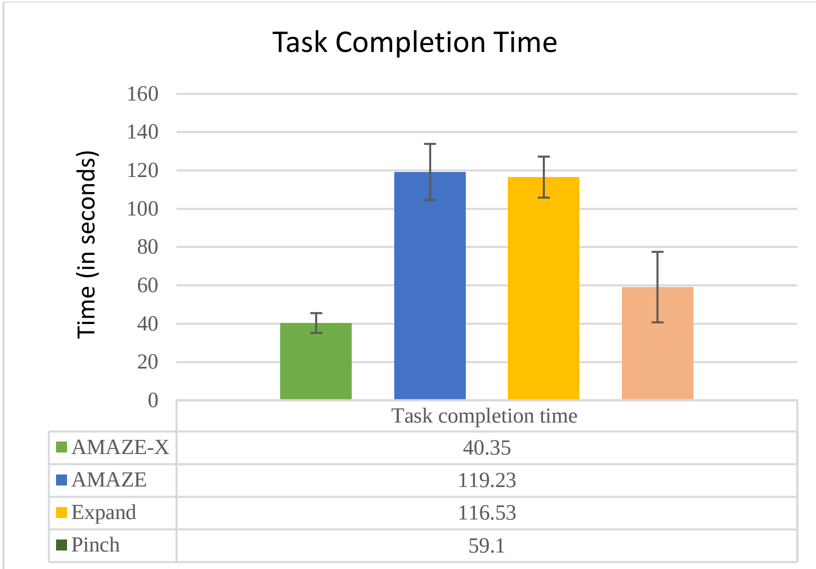


Figure 36: Mean and SD values for task completion time for AMAZE-X, AMAZE, Expand, and Pinch. AMAZE-X was the fastest, followed by Pinch, AMAZE, and Expand.

6.7.2 Error Rate

We conducted ANOVA to analyze the differences in error rates among the four techniques with a significance level of 0.05. Consistent with the findings of the previous study, participants did not make any errors while using Amaze and Expand techniques. Similarly, there were no errors made while using AMAZE-X. The Tukey post hoc test showed that the error rate for Pinch-to-select technique (M=1.72, SD=1.48) was significantly higher than that of AMAZE-X (M=0, SD=0), AMAZE (M=0, SD=0), and Expand (M=0, SD=0) techniques ($p<0.05$). Thus, AMAZE-X, AMAZE, and Expand techniques resulted in zero errors. Figure 37 shows the mean and SD values for error rate for AMAZE-X, AMAZE, Expand, and Pinch.

Figure 37: The mean and SD values for error rate for AMAZE-X, AMAZE, Expand, and Pinch-to-select techniques. The participants performed no errors for AMAZE-X, AMAZE, and Expand.

6.7.3 Easy to Use

We conducted a Friedman test to compare the ease of use across all four techniques. The results indicated a significant difference in ease of use among the techniques, $\chi 2(3)=25.11$, $p=0.001$. Wilcoxon Signed Ranks post-hoc tests were conducted to further explore these differences. We found significant differences in ease of use between AMAZE-X and AMAZE ($Z=-2.72$, $p<0.05$), AMAZE-X and Expand ($Z=-3.80$, $p<0.05$) and between AMAZE-X and Pinch-to-select ($Z=-3.73$, $p<0.05$). Significant differences were also observed between AMAZE and Expand ($Z=-2.80$, $p<0.05$) and between AMAZE and Pinch-to-select ($Z=-2.43$, $p<0.05$). However, there was no significant difference in ease of use between Expand and Pinch ($Z=-3.58$, $p>0.05$). Figure 38 shows the mean and SD values for ease of use for AMAZE-X, AMAZE, Expand, and Pinch

6.7.4 Ease of Learning

The ease of learning results were analyzed using a Friedman test, and a significant difference was observed among the techniques, $\chi 2(3)=4.89$, $p>0.05$. Wilcoxon Signed Ranks post-hoc tests were carried out, which indicated significant differences for ease of learning between AMAZE-X and AMAZE ($Z=-1.89$, $p<0.05$). Significant differences were also observed for AMAZE-X and Expand ($Z=-1.81$, $p<0.05$), and AMAZE-X and Pinch-to-select ($Z=-1.04$, $p<0.05$). However, no significant differences were observed between AMAZE and Expand ($Z=-2.04$, p>0.05) and AMAZE and pinch to select (Z=-1.34, p>0.05) and between Expand and Pinch-to-select (Z=-1.43, p>0.05). Figure 38 shows the mean and SD values for ease of learning for AMAZE-X, AMAZE, Expand, and Pinch

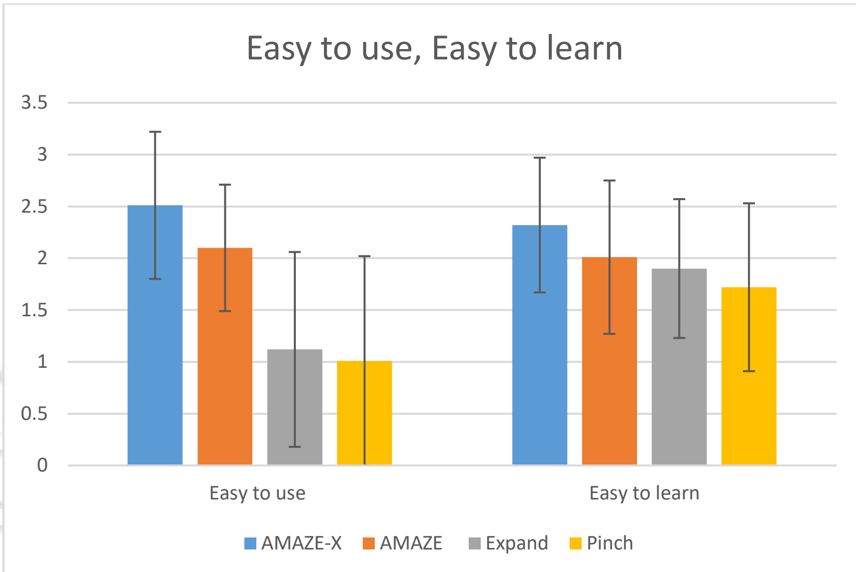


Figure 38: Mean and SD values for ease of use and ease of learning for AMAZE-X, AMAZE, Expand, and Pinch

6.7.5 User Preference

Out of 12 participants, all rated the AMAZE-X technique as their first preference. Accordingly, AMAZE was preferred by 8 participants as their second preference and 2 participants preferred Expand and pinch-select as their second preference. 2 participants preferred pinch-to-select as their third preference and Expand was preferred by 4 participants as their third preference and 6 participants ranked Expand as their third preference.

6.8 Discussion

The present study aimed to investigate the performance of four object selection techniques (AMAZE-X, AMAZE, Expand, and Pinch-to-select) in a dense virtual environment, focusing on task completion times, error rates, ease of use, and ease of learning. Our findings indicate that AMAZE-X had significantly lower task completion times and higher ease of learning than the other techniques. Additionally, AMAZE-X had no errors, and participants found it easier to use and learn than Expand and Pinch-to-select techniques.

In terms of selection time, AMAZE-X was found to be significantly faster than the other techniques (with a performance that was three times better than that of AMAZE). While the ability bring the objects within arm's reach reduced the visual search time (similar to AMAZE), there are other factors that contributed in significantly reducing the task completion time for AMAZE-X. Our findings posited that increasing the Control Display (CD) ratio for distant targets and adding user-to-target visualization would improve performance. These findings are consistent with previous research, as documented in several studies, which have demonstrated that increasing the CD ratio reduces task completion time for object selection in VR environments (Casiez et al., 2008). Post-study interviews revealed that participants preferred the user-to-target visualization, which they found intuitive. Participants reported that this feature made it easy for them to understand how close to bring the target and perform the zoom function. One participant stated, *"There were only two dash lines, so I used two fingers to zoom in and precisely place the targets near me. The visualization made it easy for me to understand how much to zoom to select the targets accurately."* Hence, we accept our first hypothesis H1. The ability to zoom-in and bring the objects within arm's reach also reduced the visual search time, and the need to use fine motor skills for object selection. It also improved the selection time according to the data analysis—these two observations supported our hypothesis.

In terms of error rate, the study found that the AMAZE-X technique had no errors, which was similar to the results of AMAZE and Expand techniques. The near and accurate placement of targets for selection by participants resulted in error-free selections for AMAZE-X and AMAZE techniques. Participants expressed confidence in their ability to select small targets accurately using these techniques. The Expand technique also had no errors, as it employs a virtual grid with targets placed in proximity, resulting in reduced density and error-free object selection. However, the Pinch-to-select technique was found to be error-prone due to hand jitter [2] and Heisenberg selection. The study partially accepted hypothesis 2, where the AMAZE-X technique was found to have significantly lower errors than the Pinch-to-select technique, and was equally efficient to AMAZE and Expand techniques in terms of error rate.

Our findings support H3, which indicate that AMAZE-X was significantly easier to use and learn compared to the other three techniques. This result is attributed to two factors. First, the removal of two operating regions (30cm and 20cm) used to increase speed eliminated confusion among participants. Second, the user-to-target visualization helped participants infer the required zoom level to position targets accurately for small object selection. Participants commented positively on the visualization's usefulness, with one noting, *"I looked at the visual element to ensure the distance and how much to zoom. It was very easy to select the targets thereafter,"* while another participant said, *"The visualization helped to distinguish minor distances accurately. I know how much to zoom to bring the target near me and where exactly to place the target. I could position the targets very near, and it was easy to select it" (P6).* In contrast, we found that participants perceived the pinch-to-select technique to be significantly more difficult to use and learn than AMAZE-X, similar to the findings of the previous study (while studying AMAZE). Pinch-to-select was also difficult to use and learn than Expand technique. This difficulty is attributed to two reasons as the earlier one: first, the requirement of holding the pinch position to select small targets using a cursor, and second, the need to precisely point the cursor to select small targets in a dense VE. Previous research has also reported that holding the pinch position and precisely pointing the cursor in pinch-to-select is challenging (Pfeuffer et al., 2017).

Participants also preferred AMAZE-X technique as compared to other techniques. Participants preferred the AMAZE-X technique as was easier and they felt more intuitive to navigate and use the technique effectively to select a small target. A participant commented, *"The AMAZE-X technique was easy to use and helped me complete the task more quickly and accurately than the other techniques."* Participants could also adjust and bring the target very close to them without the need to zoom-out. This reduced frustration and increased the overall experience of using the technique.

6.9 Design Recommendations

Based on the results of our two studies, we have identified and present four recommendations for designing selection techniques for small, distant targets in dense VEs.

Our first recommendation, R1, is to design techniques that bring objects within the user's hand's reach for small and distant object selection in dense VEs. This is because in such environments, small and distant objects are often difficult to view and access due to their small size, increased density, and low proximity between objects. As a result, users require higher cognitive effort to visually locate and confirm their selection, which negatively impacts user performance. Similar observations were made by Xu et al., (2022) where they report that the object density and distance target size significantly impacts the user performance, and the performance decreases as object density increased and target size decreases. However, bringing objects within the user's hand's reach can increase their visibility to reduce visual search time, increase hit area, and ultimately selection accuracy. This, in turn, can reduce cognitive load, increase user confidence, and improve selection accuracy. Overall, our recommendation suggests that bringing objects within the user's hand's reach can alleviate the perceived difficulty and improve the efficiency of object selection in dense VEs.

Our second recommendation, R2, is to design a shorter, faster and user-controlled method that optimize the user's physical movement to bring objects immediately within the hands reach to improve user performance and satisfaction. This strengthens the ability to reduce task completion time and perform selection tasks quickly supporting the need of novice and advanced users to interact with VR applications that require frequent object selection and manipulation. We recommend designers to devise an object selection technique that does not require repeated gestures, but optimize the user's physical movement and enable the user to perform the desired action in a single, straightforward manner. We also recommend to design a shorter and faster technique that allows the user to define a preferred zone or area for scaling and scales the defined zone or area simultaneously. For instance, AMAZE-X optimizes the user's physical movement and improves performance by defining a scaling area and simultaneously zooming-in to bring objects within arm's reach. This technique combines two steps in one, resulting in improved user performance. Additionally, AMAZE-X enables users to bring objects within arm's reach faster by performing pinch-in with multiple fingers, such as 1, 2, 3, or 4 fingers, which further shortens the time required for object selection. Designing a shorter technique has been recommended in earlier studies, and have reported improved selection time and reduced cognitive load (Schjerlund et al., 2021).

Our third recommendation R3, is to implement a visual representation in VEs that clearly displays the user-to-object distance during corrective phases of object selection. This representation should be unambiguous and un-occluding to facilitate accurate selections in dense VEs. A virtual hand that is visually displayed and illustrates its proximity to the object enhances users' judgment and control over their movements, thereby resulting in faster and more accurate selections. It is critical that designers take into account the position and orientation of the virtual hand and the object when creating the visual representation to enable users to confidently interact with virtual objects. The role of visualization based on the position and orientation in enhancing users' learning is crucial. This was demonstrated in the case of AMAZE-X, where the ease of learning was significantly higher compared to other selection techniques. Previous studies have also indicated the importance of the visualization based on the position and orientation in enhancing users' learning (Yu et al., 2019). The use of simple visual representations such as straight, dotted, or dashed lines is recommended to avoid unwanted occlusion for the user. Additionally, clear and intuitive visual representations of object location and orientation have been found to enhance users' presence and engagement in VR, as reported in previous studies (Gruenefeld et al., 2017).

Our fourth recommendation, R4 is to use a Control Display (CD) ratio 1:5 for far distances during the ballistic phase, whereas, for near object selection during the corrective phase, the CD ratio has to be 1:1. This can be particularly important in time-sensitive or high-pressure situations where speed or time is critical. By adjusting the CD ratio based on the distance from the user, users will be able to reduce the cognitive load required to complete the task, since they will not have to constantly adjust their movements and inputs to account for changes in distance. This can be particularly important in complex or multi-step tasks where cognitive load can be a significant barrier. By decreasing the CD ratio during the corrective phase for near object selection, users will be able to achieve greater precision and accuracy in their selections, since they will have finer-grained control over their movements for precise selection of small targets. This can be particularly important in situations where accuracy is paramount, such as medical or surgical procedures. It is important to note that these guidelines may need to be adjusted based on individual differences in user preferences and abilities, as well as the specific task requirements.

6.10 Chapter Summary

In this chapter, we proposed the AMAZE technique, a combination of zoom using pinching with multiple fingers to bring the target near for selection and point that allows users to select small objects in dense environments with accuracy and precision. A user evaluation was then conducted, where AMAZE was compared with two other selection techniques. Our results indicate that AMAZE was faster, performed no errors and was the most natural technique as compared to other techniques. However, AMAZE was not very easy to use and learn. This was primarily due to the two operating regions (30cm and 20cm) that created confusion among the users and increased the learning curve. Participants also felt depth perception while placing the target in the corrective phase hence participants suggested a visual element to infer distance more accurately. We considered these factors for further development and redesign of AMAZE. We extended AMAZE and designed the AMAZE-X technique to make it significantly faster, easy to use and easy to learn for users considering a scenario where targets are small, positioned at a distance and in a dense VE. We made the following design changes in AMAZE-X: (i) modifications during iterative steps while zooming during the ballistic phase. The motor space was scaled by adjusting/ increasing the control-display ratio up to 1:5 during the ballistic phase. (ii) the addition of a visual element in the corrective phase to help users infer distance more accurately in a dense configuration. To improve selection time during the corrective phase, we added dashed lines as distance indicator between the user and the target. Thus, helping users to distinguish minor distances accurately in the corrective phase. We performed a second study to evaluate AMAZE-X with techniques from study 1: AMAZE, Expand and Pinch-to-select. AMAZE-X proved to be significantly faster than other techniques and it performed zero errors. The lack of errors and uniform completion times across all participants tested make it suitable to select small targets placed at a distance in dense VEs. It was also easy to use, easy to learn, and most preferred among all techniques.

Extending AMAZE-X zoom to be better suited for dense environments, using different visualization, and additional refinement mechanisms, are some other features we intend to incorporate as future work. For instance, a better approach in the final refinement step could take into account the actual position of objects/targets and use some heuristics to determine the distance-to-target and target-to-target distance. Furthermore, a long-term study over a wider age range of participants could be considered to evaluate the generalizability of the techniques.

In all, we believe the development of AMAZE and AMAZE-X is an important step toward improving the selection of small, distance targets in dense virtual environments. In addition, from the two experiments, we have extracted recommendations that can help the design of object selection techniques for 3D virtual environments.

Chapter 7

7. Conclusion

This thesis explored the design space of body gesture-based selection in dense VE and primarily focused on small object selection for targets placed at arm's reach and at a distance. The first two were elicitation studies to propose natural and intuitive body gestures in dense VE for small object selection within arms' reach (VE1) and small object selection for distant targets (VE2). In total we identified 52 (23+29) unique gesture for VE 1 and VE 2. A single gesture for the specific VE was identified considering the ease of performance, gesture appropriateness, body-part suitability, preferences and effort. The finalized gesture for VE1 was the *point and tap* gesture and the finalized gesture for VE2 was *pinch in/out the VE and point and tap* gesture. We redesigned the gesture identified for VE1 considering the specific challenges of the environment and redesigned *Locked Dwell Time based Point and Tap (LDTPT)* technique. We performed a comparative study of the technique with techniques identified from the literature. The results suggest that the *LDTPT* technique was the fastest and produced low errors for selecting small targets at arm's reach. It was also the most natural technique, preferred among the participants and low in effort. However, the LDTPT technique was not easy to use and learn considering the feedback visualization provided and the projection of hand over small targets. We further improved upon *LDTPT* technique and explored the possibility of decreasing the hand size for the selection of small targets in dense VE where targets are positioned within arms' reach. We came up with the concept of *Tiny hands* which reduces the size of virtual hands to select small targets We compared *tiny hands* with existing techniques from the literature. Our results show that *tiny hands* was significantly faster and more accurate in selecting small targets as compared to other techniques. The *tiny hands* technique was also significantly easy to use, learn, natural and playful. It was the most preferred technique and exhibited less effort during the study. Overall, we could conclude that *tiny hands* is a versatile technique for selecting small objects in dense VE where targets are placed within arms' reach. We then present the design and evaluation of the gesture identified for VE2, dense VE, where targets are small and placed at a distance. The finalized gesture for VE 2 was *pinch in/out the VE and point and tap*. We redesigned the gesture and presented a novel technique, **AMAZE** (**A** **M**ulti-finger **A**pproach to **Z**oom in **D**ense **E**nvironments) that offers zoom using multiple fingers in VR. We performed a comparative study of AMAZE with techniques from the literature. Our results indicate that the AMAZE technique was significantly accurate and natural. However, it took higher selection time as compared to other techniques. With our analysis supported by participants' comments, we redesigned AMAZE and presented AMAZE-X. We improved upon AMAZE-X's zoom factor with the increased CD ratio of 1:5 and added subtle dash lines as visual elements during the corrective phase to improve distance to the user for faster selection. In the second comparative study, we evaluated AMAZE-X with techniques from the first study. We found that AMAZE-X was significantly faster, easier to perform, and easier to learn than other techniques. We also found AMAZE-X to be the most preferred technique. Finally, we presented a set of design recommendations that designers and developers can use to design efficient and effective gesture-based object selection techniques for small object selection in dense VE where targets are placed within arm's reach and at a distance.

Overall, this research work laid out 6 major contributions that were reported at the beginning of this thesis. Following are the contributions reiterated.

1. User-centric gestures for controller-less selection of small object placed at arm's reach in a dense virtual environment. We elicitated 196 gestures (considering the frequency of the gestures) and 23 unique gestures (after categorization) for the specific VE where targets are small and kept within arms' reach in dense VE.
2. User-centric gestures for controller-less selection of small object placed at a distance beyond arm's reach in the dense virtual environment. We elicitated 194 gestures (considering the frequency of the gestures) and 29 unique gestures (after categorization) for the specific VE where targets are small and at a distance in dense VE.
3. Categorization and a taxonomy of a set of natural and intuitive whole-body gestures for small targets placed at arms' reach and distant position in dense virtual environment. We proposed the gesture taxonomies from 52 (23+29) unique gestures extracted from the study 1. Gestures were classified based on the hand dominance in performing the task and the motion performed by hands. We also observed the posture and the sequence of gestures to

propose the taxonomies. We introduce two taxonomies: hand-dominance and multiple body-part movement gesture, each with sub-categories (dominant hand only, non-dominant hand first, equal hand dominance).

4. Design and evaluation of Locked Dwell Time based Point and Tap (LDTPT) and Tiny hands, two novel gesture based selection technique for small objects in arm's reach in dense virtual environment. We improved upon the user centric gesture identified for VE1 (Study 1 and Study 2) and designed the two techniques to accurately select small objects in dense VE where objects are placed within arms' reach. We were able to establish that these two techniques proved to be efficient in terms of task completion time and error rate. LDTPT and Tiny hands technique was also natural, easy to use and easy to learn, most preferred and low in effort in selection of small targets in dense VE where objects are within arms' reach.
5. Design and evaluation of AMAZE and AMAZE-X, a novel gesture-based object selection technique for small, distant object selection in dense virtual environment. We designed and evaluated AMAZE for accurate selection of distant targets. We established that AMAZE and AMAZE-X proved to be better in terms of task completion time and error rate. It was also easy to use, easy to learn, natural and most preferred among the techniques.
6. Recommendations and design guidelines for designers and developers to design efficient gesture-based selection techniques for small object selection in dense virtual environment for targets placed within arm's length and at a distance. Depending on the context and different application scenarios designers can consider using the different techniques. Considering a medical surgery context (objects are small placed in a dense VE within arms' reach) where accuracy is of utmost importance, the tiny hands gesture could be used as the user is able to navigate with the small size of the hand and accurately select target.

7.1 Scope of Future Research

The qualitative findings identified through direct observations and analysis of video recordings during the study established new research directions in gesture-based small object selection for small objects within arm's reach and at a distance. Although these findings did not present statistical relevance, they reported a substantial impact on the design, preference, and final results of the proposed gestures and the redesign of the selection techniques. The following sections elaborate on the key findings and new proposals for further research in investigating the overall effectiveness of small object selection in dense VE.

The research dwelled on gesture-based object selection in dense VE for selecting small targets within arm's reach and at a distance which was not sufficiently explored. The elicited gestures from our study for each of the specific VE were categorized into unique gestures for each VE. This presents scope to redesign and compare each of the (29 and 23) unique gestures considering a more realistic use case and application context such as human anatomy, scientific molecular visualization, medical surgery etc. This would allow for a more comprehensive evaluation of the effectiveness and usability of different gestural techniques in different application.

The current studies only have used one level of target sizes. This limits the generality of results and one could address this limitation with additional experiments using a broader range of target sizes, distances, dynamic objects and number of objects in the VE, yielding a more comprehensive range of task difficulties. This would allow for better prediction of indices of difficulty for small object selection. To improve the validity of the results, the proposed techniques needs to be validated with a larger sample size. One can also integrate multi-tasking scenarios i.e. when the user is walking, standing, or where manipulation tasks are integrated along with selection tasks. Selection in occluded dense VE is also an essential scope in future to establish robustness and reliability of the technique. To increase the applicability of the technique, it is viable to perform comparative studies with multiple other techniques present in the literature e.g. LenSelect, (Wegele, 2020) and Precious (Mendes et al., 2017) etc.

Many current benefits and limitations of our main findings are broadly applicable only across the class of VR devices used for the study. One could also use additional types of VR systems for such experimentation, such as Microsoft Hololens, Meta Quest Pro which are the most advanced commercially available VR headsets. This could have implications on the usability and effectiveness of different techniques over various platforms.

Lastly, the study was primarily done with tech-literate participants with a good VR experience and familiarity with newer technologies. However, given that the use of VR is still in its early stages and has limited adoption in India, there is room for improvement in the techniques employed in the study to cater to a broader range of users and applications. It is also important to consider how user factors such as age, gender, and experience level may impact selection performance. A cross-cultural study can be conducted to investigate the generalizability of the proposed taxonomy across different cultures and nationalities. This can provide insights into the cultural differences in gestural interactions in VR. Such a study could provide valuable insights into the ways that culture shapes our use and interpretation of gestural interactions in virtual environments. For example, some cultures may have different expectations or norms around physical proximity, eye contact, or touch, which could impact the effectiveness and acceptability of different gestural techniques. By examining these cultural differences, researchers could gain a more nuanced understanding of the factors that influence the usability and effectiveness of gestural interactions in VR. The study can also be extended to investigate the physiological and cognitive aspects of gestural interactions in VR, such as fatigue, cognitive load, and user experience.

Publications

Journal and conference

Bhowmick, S., Biswas, N., Kalita, P. C., & Sorathia, K. (2023). Design and evaluation of AMAZE: A Multi-Finger Approach to Select Small and Distant Objects in Dense Virtual Environments. Displays, 80, 102539.

Bhowmick, S., Kalita, P., & Sorathia, K. (2020, November). A Gesture Elicitation Study for Selection of Nail Size Objects in a Dense and Occluded Dense HMD-VR. In *IndiaHCI'20: Proceedings of the 11th Indian Conference on Human-Computer Interaction* (pp. 12-23).

Bhowmick, S., Kalita, P., & Sorathia, K. (2021, November). Understanding Gesture Performance for Object Selection in VR: Classification and Taxonomy of Gestures in HCI. *In Proceedings of ACM IndiaHCI conference (IndiaHCI'21). ACM, India*, 6 pages.

Extended Abstract and Posters

Bhowmick, S., Biswas, N., Kalita, P. C., & Sorathia, K. (2023, April). Wow! I Have Tiny Hands: Design And Evaluation Of Adaptive Virtual Hands For Small Object Selection Within Arms' Length in Dense Virtual Environment. In proceeding of CHI EA '23, April 23–28, 2023, Hamburg, Germany. https://doi.org/10.1145/3544549.3585686

Bhowmick, S., Panigrahi, A., Borah, P., Kalita, P., & Sorathia, K. (2020, October). Investigating the Effectiveness of Locked Dwell Time-based Point and Tap Gesture for Selection of Nail-sized Objects in Dense Virtual Environment. In *Symposium on Spatial User Interaction* (pp. 1-2).

Bhowmick, S. (2021, March). Exploring Body Gestures for Small Object Selection in Dense Environment in HMD VR for Data Visualization Applications. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 713-714). IEEE.

Bhowmick, S., Kalita, P. C., & Sorathia, K. (2022, November). Tiny Hands are Cute: Adaptive Virtual Hands to Accurately Select Nail-Size Arm's Reach Virtual Objects in Dense Immersive VR. In HCI

International 2022–Late Breaking Posters: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part I (pp. 520-527). Cham: Springer Nature Switzerland.

References

1. Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., & Koh, J. T. K. V. (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Microsoft Research TechReport MSR-TR-2012-111, 2, 30.
2. Argelaguet, F., & Andujar, C. (2008). Improving 3D selection in VEs through expanding targets and forced disocclusion. In *Smart Graphics: 9th International Symposium, SG 2008, Rennes, France, August 27-29, 2008. Proceedings 9* (pp. 45-57). Springer Berlin Heidelberg.
3. Argelaguet, F., Andujar, C., & Trueba, R. (2008, October). Overcoming eye-hand visibility mismatch in 3D pointing selection. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology* (pp. 43-46).
4. Argelaguet, F., & Andujar, C. (2009). Efficient 3D pointing selection in cluttered virtual environments. *IEEE Computer Graphics and Applications*, *29*(6), 34-43.
5. Argelaguet, F., & Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, *37*(3), 121-136.
6. Ariza, O., Bruder, G., Katzakis, N., & Steinicke, F. (2018, March). Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 327-334). IEEE.
7. Batmaz, A. U., & Stuerzlinger, W. (2019, March). Effects of 3D rotational jitter and selection methods on 3D pointing tasks. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 1687-1692). IEEE.
8. Bacim, F., Kopper, R., & Bowman, D. A. (2013). Design and evaluation of 3D selection techniques based on progressive refinement. *International Journal of Human-Computer Studies*, *71*(7-8), 785-802.

9. Bowman, D. A., & Hodges, L. F. (1997, April). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics* (pp. 35-ff).
10. Bowman, D. A., Johnson, D. B., & Hodges, L. F. (1999, December). Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology* (pp. 26-33).
11. Bowman, D. A., Kruijff, E., LaViola, J. J., & Poupyrev, I. (2001). An introduction to 3-D user interface design. *Presence*, *10*(1), 96-108.
12. Bowman, D. A., Wingrave, C. A., Campbell, J. M., Ly, V. Q., & Rhoton, C. J. (2002). Novel uses of Pinch Gloves™ for virtual environment interaction techniques. *Virtual Reality*, *6*, 122-129.
13. Casiez, G., Vogel, D., Balakrishnan, R., & Cockburn, A. (2008). The impact of control-display gain on user performance in pointing tasks. *Human–computer interaction*, *23*(3), 215-250.
14. Cashion, J., Wingrave, C., & LaViola Jr, J. J. (2012). Dense and dynamic 3d selection for game-based virtual environments. *IEEE transactions on visualization and computer graphics*, *18*(4), 634-642.
15. Cashion, J., & LaViola, J. J. (2014, March). Poster: Dynamic adaptation of 3D selection techniques for suitability across diverse scenarios. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 165-166). IEEE.
16. Cockburn, A., & Firth, A. (2004). Improving the acquisition of small targets. In *People and Computers XVII—Designing for Society: Proceedings of HCI 2003* (pp. 181-196). Springer London.
17. Cockburn, A., & Brewster, S. (2005). Multimodal feedback for the acquisition of small targets. *Ergonomics*, *48*(9), 1129-1150.
18. Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993, February). Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces* (pp. 193-200).
19. Delamare, W., Daniel, M., & Hasan, K. (2022, April). MultiFingerBubble: A 3D Bubble Cursor Variation for Dense Environments. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1-6).

20. El Jamiy, F., & Marsh, R. (2019). Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing*, *13*(5), 707-712.
21. Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. MIT Press
22. Fekete, J. D., Elmqvist, N., & Guiard, Y. (2009, April). Motion-pointing: target selection using elliptical motions. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 289-298).
23. Findlater, L., Jansen, A., Shinohara, K., Dixon, M., Kamb, P., Rakita, J., & Wobbrock, J. O. (2010, October). Enhanced area cursors: reducing fine pointing demands for people with motor impairments. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (pp. 153-162).
24. Forsberg, A., Herndon, K., & Zeleznik, R. (1996, November). Aperture based selection for immersive virtual environments. In *Proceedings of the 9th annual ACM symposium on User interface software and technology* (pp. 95-96).
25. Frees, S., & Kessler, G. D. (2005, March). Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* (pp. 99-106). IEEE.
26. Grossman, T., & Balakrishnan, R. (2005, April). The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 281-290).
27. Grossman, T., & Balakrishnan, R. (2006, October). The design and evaluation of selection techniques for 3D volumetric displays. In *Proceedings of the 19th annual ACM symposium on User interface software and technology* (pp. 3-12).
28. Gruenefeld, U., Ennenga, D., Ali, A. E., Heuten, W., & Boll, S. (2017, October). Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. In *Proceedings of the 5th symposium on spatial user interaction* (pp. 109-118).
29. de Haan, G., Griffith, E. J., Koutek, M., & Post, F. H. (2006, May). Hybrid Interfaces in VEs: Intent and Interaction. In *EGVE* (pp. 109-118).

30. Guiard, Y. (1987). Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. Journal of motor behavior, 19(4), 486-517.

31. Jacob, R. J. (1990, March). What you look at is what you get: eye movement-based interaction techniques. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 11-18).

32. Jiang, H., Weng, D., Dongye, X., & Liu, Y. (2022). PinchText: One-Handed Text Entry Technique Combining Pinch Gestures and Hand Positions for Head-Mounted Displays. *International Journal of Human–Computer Interaction*, 1-17.

33. Kang, H. J., Shin, J. H., & Ponto, K. (2020, March). A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 275-284). IEEE.

34. Karam, M., & Schraefel, M. C. (2005, April). A study on the use of semaphoric gestures to support secondary task interactions. In CHI'05 Extended Abstracts on Human Factors in Computing Systems (pp. 1961-1964).

35. Kopper, R., Ni, T., Bowman, D. A., & Pinho, M. (2006, March). Design and evaluation of navigation techniques for multiscale virtual environments. In *Ieee virtual reality conference (vr 2006)* (pp. 175-182). Ieee.

36. Kopper, R., Bowman, D. A., Silva, M. G., & McMahan, R. P. (2010). A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, *68*(10), 603-615.

37. Kopper, R., Bacim, F., & Bowman, D. A. (2011, March). Rapid and accurate 3D selection by progressive refinement. In *2011 IEEE symposium on 3D user interfaces (3DUI)* (pp. 67-74). IEEE.

38. Kress, B. C., & Cummings, W. J. (2017, May). 11-1: invited paper: towards the ultimate mixed reality experience: HoloLens display architecture choices. In *SID symposium digest of technical papers* (Vol. 48, No. 1, pp. 127-131).

39. LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3D user interfaces: theory and practice*. Addison-Wesley Professional.

40. Liang, J., & Green, M. (1994). JDCAD: A highly interactive 3D modeling system. *Computers & graphics*, *18*(4), 499-506.

41. Lin, J., & Schulze, J. P. (2016). Towards naturally grabbing and moving objects in VR. *Electronic Imaging*, *28*, 1-6.

42. Liu, L., & van Liere, R. (2009, November). Designing 3D Selection Techniques Using Ballistic and Corrective Movements. In *EGVE/ICAT/EuroVR* (pp. 1-8).

43. Lucas, J. F. (2005). *Design and evaluation of 3D multiple object selection techniques* (Doctoral dissertation, Virginia Tech).

44. Matulic, F., & Vogel, D. (2018, April). Multiray: multi-finger raycasting for large displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

45. Mayer, S., Wolf, K., Schneegass, S., & Henze, N. (2015, April). Modeling distant pointing for compensating systematic displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 4165-4168).

46. Mayer, S., Schwind, V., Schweigert, R., & Henze, N. (2018, April). The effect of offset correction and cursor on mid-air pointing in real and virtual environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

47. McGuffin, M., & Balakrishnan, R. (2002, April). Acquisition of expanding targets. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 57-64).

48. Mendes, D., Fonseca, F., Araujo, B., Ferreira, A., & Jorge, J. (2014, March). Mid-air interactions above stereoscopic interactive tables. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 3-10). IEEE.

49. Mendes, D., Medeiros, D., Sousa, M., Cordeiro, E., Ferreira, A., & Jorge, J. A. (2017). Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement. *Computers & Graphics*, *67*, 95-102.

50. Mine, M. R. (1995). Virtual environment interaction techniques. *UNC Chapel Hill CS Dept*.

51. Moore, A. G., Hatch, J. G., Kuehl, S., & McMahan, R. P. (2018). VOTE: A ray-casting study of vote-oriented technique enhancements. *International Journal of Human-Computer Studies*, *120*, 36-48.

52. Mutasim, A. K., Batmaz, A. U., & Stuerzlinger, W. (2021, May). 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* (pp. 1-7).
53. Nacenta, M. A., Kamber, Y., Qiang, Y., & Kristensson, P. O. (2013, April). Memorability of pre-designed and user-defined gesture sets. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1099-1108).
54. Nanjappan, V., Liang, H. N., Lu, F., Papangelis, K., Yue, Y., & Man, K. L. (2018). User-elicited dual-hand interactions for manipulating 3D objects in virtual reality environments. *Human-centric Computing and Information Sciences*, *8*(1), 1-16.
55. Norman, D. A. (1988). *The psychology of everyday things*. Basic books.
56. Norman, D. A. (2010). Natural user interfaces are not natural.Interactions,17(3), 6–10.
57. Norman, D. A., & Nielsen, J. (2010). Gestural interfaces: A stepbackward in usability.Interactions,17(5), 46–49.
58. Ni, T., Bowman, D. A., North, C., & McMahan, R. P. (2011). Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *International Journal of Human-Computer Studies*, *69*(9), 551-562.
59. Nickel, K., & Stiefelhagen, R. (2003, November). Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. In *Proceedings of the 5th international conference on Multimodal interfaces* (pp. 140-146).
60. Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.
61. Nielsen, M., Störring, M., Moeslund, T. B., & Granum, E. (2004). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers 5 (pp. 409-420). Springer Berlin Heidelberg.
62. Set Up Hand Tracking | Oculus Developers. (n.d.). Set up Hand Tracking | Oculus Developers. Retrieved from https://developer.oculus.com/documentation/unity/unity-handtracking/
63. Pereira, A., Wachs, J. P., Park, K., & Rempel, D. (2015). A user-developed 3-D hand gesture set for human–computer interaction.Human Factors: theJournal of the Human Factors and Ergonomics Society,57(4), 607–621

64. Periverzov, F., & Ilieş, H. (2015, March). IDS: The intent driven selection method for natural user interfaces. In *2015 IEEE symposium on 3D user interfaces (3DUI)* (pp. 121-128). IEEE.
65. Pierce, J. S., Forsberg, A. S., Conway, M. J., Hong, S., Zeleznik, R. C., & Mine, M. R. (1997, April). Image plane interaction techniques in 3D immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics* (pp. 39-ff).
66. Piumsomboon, T., Clark, A., Billinghurst, M., & Cockburn, A. (2013). User-defined gestures for augmented reality. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 955-960).
67. Pfeuffer, K., Mayer, B., Mardanbegi, D., & Gellersen, H. (2017, October). Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th symposium on spatial user interaction* (pp. 99-108).
68. Poggi, I. (2002, May). From a typology of gestures to a procedure for gesture production. In Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, GW 2001 London, UK, April 18–20, 2001 Revised Papers (pp. 158-168). Berlin, Heidelberg: Springer Berlin Heidelberg.
69. Poupyrev, I., Billinghurst, M., Weghorst, S., & Ichikawa, T. (1996, November). The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th annual ACM symposium on User interface software and technology* (pp. 79-80).
70. Poupyrev, I., Ichikawa, T., Weghorst, S., & Billinghurst, M. (1998, August). Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum* (Vol. 17, No. 3, pp. 41-52). Oxford, UK and Boston, USA: Blackwell Publishers Ltd.
71. Plaumann, K., Weing, M., Winkler, C., Müller, M., & Rukzio, E. (2018). Towards accurate cursorless pointing: the effects of ocular dominance and handedness. Personal and Ubiquitous Computing, 22, 633-646.
72. Ortega, M. (2013, March). Hook: Heuristics for selecting 3D moving objects in dense target environments. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 119-122). IEEE.

73. Ramcharitar, A., & Teather, R. J. (2018, June). EZCursorVR: 2D selection with virtual reality head-mounted displays. In *Proceedings of the 44th Graphics Interface Conference* (pp. 123-130).

74. Ren, G., & O'Neill, E. (2013). 3D selection with freehand gesture. *Computers & Graphics*, *37*(3), 101-120.

75. Rosa, D. A. W., & Nagel, H. H. (2010, November). Selection techniques for dense and occluded virtual 3d environments, supported by depth feedback: Double, bound and depth bubble cursors. In *2010 XXIX International Conference of the Chilean Computer Science Society* (pp. 218-225). IEEE.

76. Ruiz, J., Li, Y., & Lank, E. (2011, May). User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 197-206).

77. Schjerlund, J., Hornbæk, K., & Bergström, J. (2021, May). Ninja hands: Using many hands to improve target selection in vr. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

78. Seixas, M., Cardoso, J., & Dias, M. T. G. (2015). One hand or two hands? 2D selection tasks with the leap motion device.

79. Spittle, B., Frutos-Pascual, M., Creed, C., & Williams, I. (2022). A review of interaction techniques for immersive environments. *IEEE Transactions on Visualization and Computer Graphics*.

80. Steed, A., & Parker, C. (2005). Evaluating effectiveness of interaction techniques across immersive virtual environmental systems. *Presence*, *14*(5), 511-527.

81. Steinicke, F., Ropinski, T., & Hinrichs, K. (2006). Object selection in virtual environments using an improved virtual pointer metaphor. In *Computer Vision and Graphics: International Conference, ICCVG 2004, Warsaw, Poland, September 2004, Proceedings* (pp. 320-326). Springer Netherlands.

82. Stoakley, R., Conway, M. J., & Pausch, R. (1995, May). Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 265-272).

83. Sutherland, I. E. (1965, May). The ultimate display. In Proceedings of the IFIP Congress (Vol. 2, No. 506-508, pp. 506-508).

84. Teather, R. J., and Stuerzlinger, W. (2011, March). Pointing at 3D targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 87-94). IEEE.
85. Tian, F., Lyu, F., Zhang, X., Ren, X., & Wang, H. (2017). An empirical study on the interaction capability of arm stretching. *International Journal of Human–Computer Interaction*, *33*(7), 565-575.
86. Tu, H., Huang, S., Yuan, J., Ren, X., & Tian, F. (2019, May). Crossing-based selection with virtual reality head-mounted displays. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-14).
87. Vafaei, F. (2013). Taxonomy of gestures in human computer interaction.
88. Vanacken, L., Grossman, T., & Coninx, K. (2007, March). Exploring the effects of environment density and target visibility on object selection in 3D virtual environments. In *2007 IEEE symposium on 3D user interfaces*. IEEE.
89. Vanacken, L., Grossman, T., & Coninx, K. (2009). Multimodal selection techniques for dense and occluded 3D virtual environments. *International Journal of Human-Computer Studies*, *67*(3), 237-255.
90. Vatavu, R. D. (2012, July). User-defined gestures for free-hand TV control. In *Proceedings of the 10th European conference on Interactive tv and video* (pp. 45-48).
91. Vogel, D., & Balakrishnan, R. (2005, October). Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th annual ACM symposium on User interface software and technology* (pp. 33-42).
92. Vosinakis, S., & Koutsabasis, P. (2018). Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. *Virtual Reality*, *22*(1), 47-62.
93. Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge.
94. Wilkes, C., & Bowman, D. A. (2008, October). Advantages of velocity-based scaling for distant 3D manipulation. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology* (pp. 23-29).
95. Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(3), 225-240.

96. Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009, April). User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1083-1092).

97. Wonner, J., Grosjean, J., Capobianco, A., & Bechmann, D. (2012, December). Starfish: a selection technique for dense virtual environments. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology* (pp. 101-104).

98. Worden, A., Walker, N., Bharat, K., & Hudson, S. (1997, March). Making computers easier for older adults to use: area cursors and sticky icons. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (pp. 266-271).

99. Wu, H., Wang, J., & Zhang, X. (2016). User-centered gesture development in TV viewing environment. *Multimedia Tools and Applications*, *75*, 733-760.

100. Wyss, H. P., Blach, R., & Bues, M. (2006, March). iSith-Intersection-based spatial interaction for two hands. In *3D User Interfaces (3DUI'06)* (pp. 59-61). IEEE.

101. Yu, D., Liang, H. N., Fan, K., Zhang, H., Fleming, C., & Papangelis, K. (2019). Design and evaluation of visualization techniques of off-screen and occluded targets in virtual reality environments. *IEEE transactions on visualization and computer graphics*, *26*(9), 2762-2774.

102. Zaiţi, I. A., Pentiuc, Ş. G., & Vatavu, R. D. (2015). On free-hand TV control: experimental results on user-elicited gestures with Leap Motion. *Personal and Ubiquitous Computing*, *19*, 821-838.

103. Zhai, S., Buxton, W., & Milgram, P. (1994, April). The “silk cursor” investigating transparency for 3d target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 459-464).

104. Zhao, Y., Du, N., Xu, X., Gu, Q., Wang, L., Gao, Z., & Wang, C. (2014).The influence of user-tracking feedback format on gestural interaction’s user experience: A Kinect-based usability study.Chinese Journal of Applied Psychology,20(4), 367–374

105. Wegele, W. (2020). LenSelect-Dynamic Object Scaling in Virtual Environments for Object Selection.