Ph.D. Thesis

Enhancing Mid-Air Selection and Manipulation for Complex Virtual Reality Interaction

February 21, 2023

Difeng Yu

ORCID: 0000-0002-7753-4591

School of Computing and Information Systems
Faculty of Engineering and Information Technology
The University of Melbourne, Australia

Submitted in total fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

Object selection and manipulation are fundamental to interacting with objects in Virtual Reality (VR) systems. Existing object selection and manipulation techniques in VR are primarily based on mid-air interaction with virtual hands or ray pointers. They are simple and intuitive but are often criticized in the literature for being imprecise, inefficient, and cumbersome. Specifically, these techniques are insufficient for complex VR interaction scenarios that contain small, distant, and occluded targets and require efficient, precise, versatile, and prolonged operations.

This thesis presents occlusion visualization techniques, integration strategies of complementary modalities of gaze and body surfaces, and predictive systems based on target prediction models to enhance virtual hands and ray pointers for complex VR interactions. Findings from a series of user studies demonstrated that the proposed solutions could select and manipulate small, distant, and occluded targets in an effective, efficient, comfortable, and satisfying manner. Overall, our technical solutions and findings can inform the future design of more usable and useful 3D user interfaces for VR systems.

DECLARATION

This is to certify that

(1) the thesis comprises only my original work towards the Ph.D.,

(2) due acknowledgment has been made in the text to all other material used,

(3) appropriate ethics procedure and guidelines have been followed to conduct this research,

(4) the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies, and appendices.

Difeng Yu

February 21, 2023

PREFACE

This thesis is submitted to fulfill the requirements for the degree of Doctor of Philosophy at The University of Melbourne. The research was conducted during my study at The University of Melbourne under the supervision of A/Prof. Jorge Goncalves, A/Prof. Eduardo Velloso, and Dr. Tilman Dingler.

The thesis includes four peer-reviewed articles (Article I-IV as listed below) following The University of Melbourne guidelines for a thesis with publication[1]. Among the four articles, Articles I and III were completed fully within The University of Melbourne, Article II was completed in collaboration with X-CHI Lab at Xi'an Jiaotong-Liverpool University, and Article IV was completed during my remote internship at Reality Labs Research, Meta Inc.

While several collaborators have contributed to the articles, I declare that I am the primary author and have more than 50% contributions to each of the publications. More specifically, I proposed the research questions, designed the solutions, planned the study design, developed the VR software, (co-)conducted the user studies, and performed the data analysis. Furthermore, I drafted the full research articles and subsequently revised them until they got published.

I am grateful for the contributions of the listed co-authors, who provided valuable feedback on the works, helped conduct the user studies, and contributed to preparing the research articles. Hence, I use the term "we" throughout this thesis to recognize my co-authors' contributions.

- **Article I**: Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. "Fully-Occluded Target Selection in Virtual Reality." *IEEE transactions on visualization and computer graphics* 26, no. 12 (2020): 3402-3413. https://doi.org/10.1109/TVCG.2020.3023606
- **Article II**: Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. "Gaze-Supported 3D Object Manipulation in Virtual Reality." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-13. 2021. https://doi.org/10.1145/3411764.3445343
- **Article III**: Difeng Yu, Qiushi Zhou, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. "Blending On-Body and Mid-Air Interaction in Virtual Reality." In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 637-646. IEEE, 2022. https://doi.org/10.1109/ISMAR55827.2022.00081

[1]The University of Melbourne. 2009. *Graduate Research Training Policy (MPF1321)*. Retrieved from https://policy.unimelb.edu.au/MPF1321

- **Article IV**: Difeng Yu, Ruta Desai, Ting Zhang, Hrvoje Benko, Tanya R. Jonker, and Aakar Gupta. "Optimizing the Timing of Intelligent Suggestion in Virtual Reality." In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1-20. 2022. https://doi.org/10.1145/3526113.3545632

ACKNOWLEDGEMENTS

To begin with, I am deeply grateful to my incredible advisors, Jorge Goncalves, Eduardo Velloso, and Tilman Dingler. Jorge, with his enlightened wisdom, always led me in the correct direction, research-wise and life-wise. I will never forget the mindset you taught me to deal with stressful conditions and your encouragement during my downtimes. You are always responsive to my messages, and I had a terrific time during our meetings (e.g., discussing ART and human reaction time). Eduardo can often point out any unclearness in my papers with constructive criticism. I appreciate your insights and honesty that empowered new perspectives. I will be less vague (next time). Tilman, with his extensive reading experience, provided a deep understanding and triggered engaging discussions on various topics. Let's find a time to eat more xiaolongbao. I am also thankful for the online game events, beer sessions, and retreats the three of you have organized (and for including me in these enjoyable occasions). I am honored to have had all of you as my advisors. I also thank Prof. Rui Zhang and Prof. Lars Kulik for being my advisory committee chair and for providing perspicacious feedback.

I offer my heartfelt gratitude to Weiwei and Andrew for their company during the lockdowns and the 2022 ISMAR conference in Singapore. We faced many challenges in the in-lab work during the pandemic, but we managed to navigate the difficulties and produced several thought-provoking side-projects. We should have more Hainanese chicken in the Hawkers center to reward ourselves. I also want to convey my appreciation to Qiushi Zhou, Brandon Syiem, and Benjamin Tag for collaborating on my Ph.D. projects. Your expertise and feedback have been invaluable in bringing these works to fruition. In addition, I must remember the unwavering support from the big sister of the lab—Zhanna Sarsenbayeva. We've had great Xingjiang food and microwaved dumplings.

I sincerely appreciate the esteemed faculty members, including Prof. Frank Vetere, Jarrod Knibbe, Jenny Waycott, Melissa Rogerson, Prof. Vassilis Kostakos, Wafa Johal, and Prof. Wally Smith, for their guidance, support, and insightful discussions. We had good times in the many group events, and I learned a lot from you in our impromptu talks.

Furthermore, I was fortunate to have the support and camaraderie of many friends and colleagues I met in Melbourne. Especially, I want to thank Chaofan Wang, Chunxue Wei, Danula Hettiachchi, Ebrahim Babaei, Elsy Garcia, Gabriele Marini, Hao-Ping (Hank) Lee, Henrietta Lyons, Jing Wei, Joe Brailsford, Joshua Newn, Martin Reinoso, Maximilian Tränkler, Kangning Yang, Mo Zhang, Nattapat Boonprakong, Romy Gruber, Samangi Wadinambi Arachchi, Sara Khorasani, Saumya Pareek, Senuri Wijenayake, Songyan Teng, Sophie Freeman, Thomas van Gemert, Ulan Kelesbekov, and Ying Ma. I owe a debt of gratitude for the countless laughs and thoughtful conversations during coffee breaks, lunchtimes, and retreats. I hope you keep the

Lunch Train going. Chew choo! I also thank the friends I met in CIS-GReS, the tutoring team, and the badminton groups for the joyful moments.

I express my gratitude to the wonderful people I met at Meta Reality Labs, especially Aakar Gupta, Ruta Desai, Ting Zhang, and Tanya Jonker. While the internship was online during the pandemic, our discussions on the nitty-gritty of the work reminded me of the excitement of conducting research. Your suggestions can always seize the bigger picture while picking up on nuances and subtleties that are essential to ensure the validity of the work. Thanks for letting me know that doing research in industry is as fun as in academia.

Special thanks go to my remarkable undergraduate advisor—Hai-Ning Liang. You led me to the field of human-computer interaction and shaped my initial research perspectives and skills. I remember your patience in teaching me research methodologies and correcting my writing. I am also indebted that we continued our collaboration during my Ph.D., which allowed me to meet a lot of fantastic collaborators at Xi'an Jiaotong-Liverpool University, such as Xueshi Lu, Yushi Wei, Rongkai Shi, Lei Chen, Kaixuan Fan, Heng Zhang, Feiyu Lu, and Wenge Xu.

I also extend my appreciation to the faculty and staff at the University of Melbourne for their support and guidance, including but not limited to Nicole Barbee, Allen Mari Pilares, Rhonda Smithies, and Emma Russo. I also acknowledge that the University of Melbourne and Meta fund my Ph.D. research—I am grateful for their investment in my academic pursuits. Furthermore, I thank many participants involved in my studies who have contributed to the understanding and knowledge presented in this thesis.

Additionally, I thank the creator teams of games like StarCraft 2, Diablo II: Resurrected, and Beat Saber, as well as many live streamers and online content makers such as SCBoys. These games and videos helped me through the days and nights of the pandemic. I was thrilled that TIME (well, Oliveira) won the IEM Katowice 2023 SC2 championship just before my thesis submission. I've been watching your games since the beginning of my Ph.D.

Finally, I want to take a moment to express my sincerest thanks to my friends and families. I thank my parents and my partner, Mingxuan, for everything you did and sacrificed to support me during my Ph.D. Your love, encouragement, and inspiration are indispensable parts of my life. I also thank my best friends Huiyi and Yudong for sharing all the joys and tears. As I said, although my research endeavors can be easily replaced or continued by other researchers, I am irreplaceable in your life, and you are the reason for my existence. I hope to spend more time with you for the rest of my life.

Melbourne, February 21, 2023
Difeng Yu

CONTENTS

Chapter 1

INTRODUCTION

Object selection and manipulation are indispensable for interacting with virtual objects in Virtual Reality (VR) headsets [13, 21]. Users perform selections to identify the target of interest and execute manipulations, including translation, rotation, and scaling, to further transform the target into a desired configuration. Compared to desktop- or tablet-based systems, interacting with VR headsets fundamentally differs because users are fully immersed in a 3D digital space with co-located virtual objects. They can observe a target from different angles, touch, grab, point, pull, push, and even squeeze the object. Because of this significant difference in experiencing the 3D world, VR technology requires unprecedented, new ways of interaction.

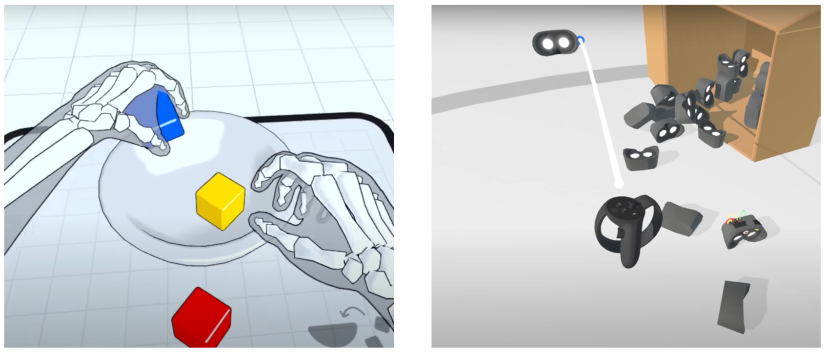


Fig. 1. Left: A user is grabbing a blue cube through Virtual Hand in *Hand Physics Lab*. Right: A user is pointing at a virtual goggle through Raycasting in *Virtual Virtual Reality*.

Historically, there are two seminal selection and manipulation techniques that both leverage mid-air gestures and movements for input: Virtual Hand and Raycasting [5, 83] (see Figure 1). Virtual Hand creates a virtual replica of users' physical hands in the VR space, and the user can use the virtual hands to grab and manipulate virtual objects. Raycasting emanates a virtual ray into the environment from (typically) the physical hand position, and the user can control the ray to point and interact with objects. These techniques are simple, straightforward, and intuitive for 3D interaction and have been employed in many off-the-shelf applications.

However, the literature has also pointed out known usability issues with these techniques. Performing actions in 3D space is inherently difficult [62, 64]. Simple mid-air interaction techniques such as Virtual Hand and Raycasting can be imprecise and inefficient in completing

a 3D interaction task [5, 83], especially when users cannot feel the physical properties (e.g., shapes, textures, weights) of virtual objects. Furthermore, the Heisenberg effect—where inputs such as a button click could disturb the position of the input device and result in a different selection point [189]—also plague these techniques. Additionally, it can be cumbersome to use Virtual Hand and Raycasting for a prolonged period because of the gorilla arm effect—a feeling of heaviness in the arm [16, 63].

Meanwhile, VR interaction scenarios can be complex because of the added depth dimension. For example, VR application scenarios such as immersive data analytics [101], medical training [145], and interior design [75] may involve complicated visualizations (see Figure 2 left). Therefore, targeted objects of interest can be small, distant, off-screen, and even fully occluded. It is difficult to acquire and manipulate such targets with Virtual Hand and Raycasting. In other application scenarios like 3D modelling [77], the task may require interaction techniques that are efficient, precise, versatile, and comfortable for prolonged usage (see Figure 2 right). While Virtual Hand and Raycasting may work fine for generic interactions with unoccluded, properly-sized buttons, menus, and virtual objects, they may not be sufficient for these more complex applications because of the aforementioned usability issues.

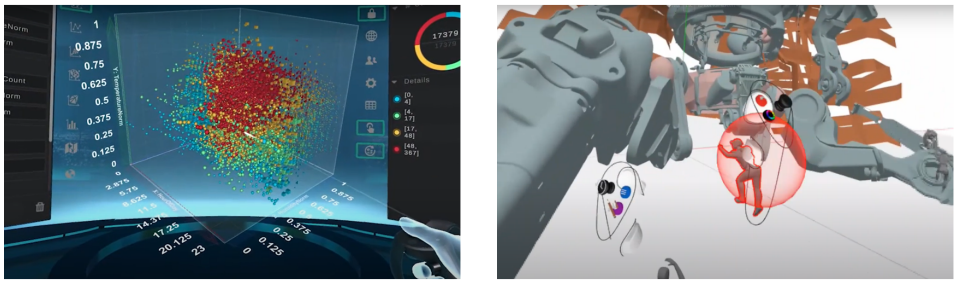


Fig. 2. Left: An immersive data analytics scenario in *Virtualitics* that involves selecting and manipulating small, distant, and occluded targets. Right: A 3D modeling scenario in *Gravity Sketch* that requires precise, versatile, and prolonged operations in VR.

1.1 Research Question

To summarize what we have discussed so far: Virtual Hand and Raycasting, which are the most prevalent mid-air techniques for object selection and manipulation in VR, have limited capability in dealing with more complex application scenarios that contain small, distant, and occluded targets and require efficient, precise, versatile, and prolonged operations. This challenge motivated the following overarching research question (**RQ**) in this thesis.

RQ. How to enhance Virtual Hand and Raycasting for target selection and manipulation in complex VR interaction scenarios?

To further specify the **RQ**, it is essential to clarify the meaning of "complex VR interaction scenarios" and "enhance" within the scope of this thesis.

1.1.1 Complex VR Interaction Scenarios

As briefly mentioned in the previous section, the "complexity" of VR interaction scenarios comes from two viewpoints: *environment* and *task*.

- *Environment*: a complex environment may contain (1) small targets, typically smaller than 1° in angular size, to be deemed as challenging for selection [194], (2) distant targets, which are outside of the arm-reach distance, and (3) occluded targets, which are partially- or fully-obscured by other distractors in an environment.
- *Task*: a complex task may require (1) precise and efficient input in completing the assignment, (2) low-fatigue, comfortable, and satisfying user experience in a prolonged interaction scenario, (3) versatile input that can support a multitude of functional requirements. These task requirements are closely connected to the measurements that will be introduced next.

1.1.2 Measurements

To "enhance" Virtual Hand and Raycasting in fulfilling the task requirements in a complex VR interaction scenario, we aim to optimize the proposed solutions in the following five measurements (namely the *5Es*). We modify and extend ISO-9241 [71], the international standard of usability measures, and the established usability metrics in human-computer interaction (HCI) [66] to a more granular, domain-specific version.

- *Effectiveness.* The accuracy with which users achieve specific goals. Common metrics include error rate, the percentage of incorrect completions in the tested set of trials, and error distance, the distance offset between target and user-completed configurations.
- *Efficiency.* The time used in relation to the results achieved. Example metrics include selection time, the time taken to complete a successful target selection, and manipulation time, the time taken to manipulate a target into a desired configuration.
- *Ergonomics.* The physical and mental workload associated with results achieved. The physical workload can be assessed through, for example, hand/arm movements and questionnaires like Borg-CR10 [15]. The mental workload is normally quantified through NASA-TLX [58].
- *Experience.* Users' feelings and satisfaction when performing tasks with the evaluated solutions. These data are normally collected from questionnaires such as UEQ-S [150].

Table 1. The solutions provided in this thesis are promising in small, distant, and occluded target selection and manipulation. They have been demonstrated to improve the existing solutions in the measurements of the *5Es*.

		Article I Occluded Visualization	**Article II** Gaze Support	**Article III** On-Body Support	**Article IV** Intelligent Suggestion
Env.	Small	✓	✓	✓	✓
	Distant	✓	✓	✓	✓
	Occluded	✓		✓	
Task	Effectiveness	✓		✓	✓
	Efficiency	✓	✓	✓	✓
	Ergonomics		✓		✓
	Experience	✓	✓	✓	✓
	Expressivity	✓	✓	✓	✓

- *Expressivity.* The solution's ability to be applied for a wide range of interactive applications or new use cases. Expressivity is typically demonstrated through sample applications.

1.2 Contribution

In this thesis, we contribute solutions to address **RQ**. More specifically, we enhance Virtual Hand and Raycasting for target selection and manipulation in complex VR interaction scenarios by incorporating occlusion visualizations, additional input/output modalities, and computational models (see Table 1). All solutions consist of interaction techniques or frameworks that have been proven to help handle complex VR interaction scenarios, with additional findings from user studies to guide interface designs.

Our solutions are distributed in the four research articles (Article I-IV). These solutions aim to expand the human-computer communication channel for more complicated VR interaction scenarios and optimize the communication process to make it more usable and useful (see Figure 3). More specifically, the solutions consider how users may benefit from receiving helpful task-related information with additional virtual contents (e.g., occlusion visualizations), extending their inputs and outputs to other modalities (e.g., eye gaze and on-body surfaces), and automating their input commands to a VR system (e.g., intelligent suggestions).

Article I discusses how occlusion visualizations such as multiple viewports, virtual X-rays, and object displacements can improve object selection, especially for fully-occluded targets. With the help of additional visualizations displayed in the virtual world, users are provided with extended capabilities to manually adjust their views and selections for completing a task.

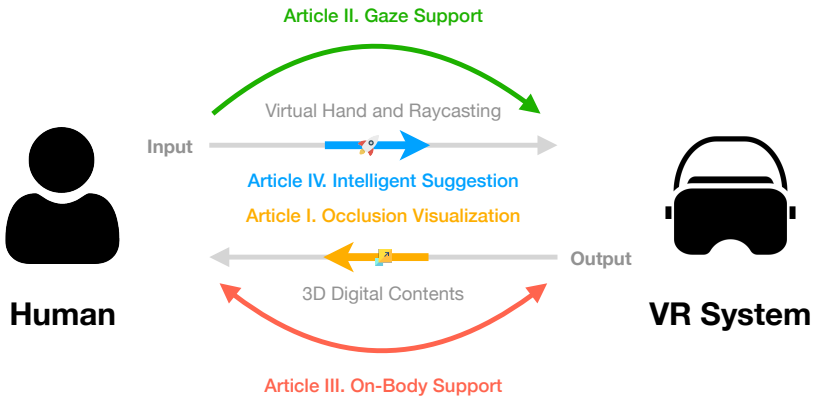


Fig. 3. The proposed solutions in this thesis aim to expand the human-computer communication channel and optimize the communication process for complex VR interaction scenarios.

Articles II and III illustrate how complementary modalities, including eye gaze and body surfaces, can augment the mid-air selection and manipulation process. By designing how mid-air interfaces may collaborate with other modalities, users can expand the human-computer communication process to other effective channels, which delievers a richer set of interaction vocabulary in VR.

Article IV demonstrates the optimal use of context-aware computational models to offer prompt, intelligent suggestions to boost user performance and experience. Rather than entirely relying on users' manual input, the predictive models infer users' intentions implicitly and provide helpful task automation.

Our solutions and findings advance the understanding of more usable and useful 3D user interfaces and will benefit future research and applications in handling a variety of novel VR interaction scenarios.

1.3 Thesis Outline

The rest of the thesis is organized as follows. Chapter 2 provides a systematic literature review that aims to identify ongoing research challenges in VR object selection and manipulation, summarize the corresponding solutions, and categorize measurements that are essential in determining the success of a solution. Next, Chapter 3 describes the methodologies that we employed to complete the research work, such as design prototyping, user studies, and data analysis. We also discuss the ethical considerations within our studies.

Chapters 4, 5, 6, and 7 present four articles (Article I-IV) that introduce new solutions for enhancing Virtual Hand and Raycasting for target selection and manipulation in complex VR

interaction scenarios. Specifically, Chapter 4 presents interactive visualizations that can help with fully-occluded target selection. Chapter 5 and Chapter 6 illustrate designs incorporating gaze and on-body surfaces into the selection and manipulation process. Chapter 7 demonstrates an optimization framework to provide timely intelligent suggestions based on target prediction models to assist object acquisition in VR.

After presenting the research publications, Chapter 8 reflects on the findings of this thesis and summarizes the solutions that can help handle complex VR selection and manipulations. In addition, we envision the future selection and manipulation techniques and point out promising research directions. Finally, Chapter 9 concludes with a summary of this thesis.

Chapter 2

LITERATURE REVIEW

Through more than 50 years of development of 3D interactions in VR, originating from Sutherland's work on interactively determining the viewing angle through head orientations in 1968 [83, 165], a multitude of solutions have been proposed for virtual object selection and manipulation. These solutions range from artifact inventions to empirical studies [9, 127, 172], span across interaction techniques and devices to computational models [9, 61, 148, 194], and extend over user input and feedback mechanisms [7, 43, 90, 168].

With the rapid development of selection and manipulation solutions, our literature review aimed to answer the following questions: (1) What core challenges in VR selection and manipulation have researchers been trying to address? Are there new challenges emerging with the development of technology? (2) What are the state-of-the-art solutions for these challenges? Why are these solutions considered successful in solving the challenge? Answering these questions is critical to determining the backbone topics and emerging trends from the scattered endeavors and ensuring the robustness and validity of our research practices.

We conducted a systematic literature review of 106 publications on object selection and manipulation in VR headsets to answer the questions. We categorized eight research challenges that the literature aimed to tackle, including those more relevant to this thesis regarding complexity in 3D interaction scenarios (e.g., small, faraway, occluded, out-of-view targets) and emerging trends such as context integration and collaborative manipulation. We also present existing solutions to these challenges. Furthermore, we classified nine success measurements used by previous research when resolving the challenges. This thesis has applied these crucial measurements extensively, especially the *5Es* (effectiveness, efficiency, ergonomics, experience, and expressivity). Finally, we summarize our recommendations regarding research practices and directions for future VR selection and manipulation studies.

2.1 Scope, Related Surveys, and Contributions

2.1.1 Scope and Definitions

The topic covered by this review is "*object selection and manipulation in VR headsets*". This section describes our scope and clarifies the inclusion and exclusion criteria.

§1 Object Selection and Manipulation. Object selection refers to acquiring or identifying one or multiple objects from an entire set of objects available. Object manipulation concerns

the further act(s) of handling the selected object, which can be broken down into sub-tasks, including positioning (changing object position), rotation (adjusting object orientation), and scaling (modifying object size) [83]. In this work, we focus on manipulations that preserve the shape of objects (i.e., spatial rigid object manipulation [83]). Furthermore, we focus on the selection and manipulation of general virtual objects rather than solutions developed for selecting a specific object type (e.g., key selection in text entry, location selection for teleportation).

§2 Fully-immersed VR headsets. This work focuses on VR technology that completely immerses a user in a computer-synthesized virtual environment [109] (i.e., does not involve the direct presence of real-world objects). The challenges and solutions of selection and manipulation can be different in other immersive technologies that afford 3D user interfaces, such as AR and MR, compared to VR because of the involvement of real-world objects [159]. In other words, this review focuses on 3D user interfaces through the perspective of VR interaction, which may or may not be applicable to other settings. Furthermore, we focus on VR head-mounted/worn displays (HMD/HWD, or more colloquially, VR headsets), which means that the visual display devices should be coupled to a user's head. Therefore, stationary VR displays (i.e., displays that do not move with the user), such as tabletop VR displays and CAVE, which afford different interaction capabilities from VR headsets, are considered out of the scope of this research.

2.1.2 Related Surveys

Several related surveys aim to create a new classification or taxonomy of different 3D selection and manipulation techniques in the literature. Dang's 2007 review [31] provides a chronological view of 3D pointing techniques. It classifies them based on 3D pointer- or selection ray-based control and how pointing is enhanced (e.g., reducing cursor movement distance, increasing target size, or both). Argelaguet and Andujar's 2013 survey [5] not only categorizes the techniques based on their intrinsic characteristics (e.g., selection tool types and how a user controls the tool) but also covers human pointing models and factors that may influence user performance in selection tasks (e.g., target geometry and object density). LaViola et al.'s 2017 book [83] (which updates Bowman et al.'s 2005 book [21]) discusses techniques for 3D selection and manipulation based on a classification of their metaphors: grasping, pointing, surface, indirect, bimanual, and hybrid. Weise et al.'s 2019 paper [178] also classifies 3D selection and manipulation techniques according to their different characteristics (e.g., metaphor, degree-of-freedom, reference frame). Mendes et al.'s 2019 survey [105] reviews 3D virtual object manipulation techniques, from desktops to immersive environments. It proposes a taxonomy based on environment properties and types of transformations. Overall, these taxonomies provide structured ways of viewing the 3D interaction techniques in the literature and offer helpful insights into designing new 3D user interfaces.

Other than creating new classifications of the techniques, more relevant to our work are surveys that identify significant design challenges with 3D interfaces and research trends for future work. Hinckley et al.'s 1994 survey [64] synthesizes design issues and potential solutions for developing effective free-space 3D user interfaces. For example, they identify that users may have difficulty understanding 3D space and offer solutions such as multi-sensory feedback to resolve this issue. They are also concerned about issues related to, for instance, dynamic target acquisition and ergonomics. Hand's 1997 survey [57] overviewed state-of-the-art 3D interaction techniques at that time and highlighted the research opportunity of usability testing for future work. Similar to these surveys, our work aims to determine research challenges and solutions and identify future research directions. We achieved this through a systematic literature review to provide an updated view of the early surveys, given the recent advancement of VR technology.

Bergström et al.'s 2021 review [13] derives guidelines on how to conduct and report object selection and manipulation studies in VR. Task types, experimental settings, target parameters, and dependent variables of such studies were analyzed in detail. The goal is to inform the design of future research studies. Other surveys overlap with our topic and inform the analysis in this paper [1]. These include, but are not limited to, a review of mid-air interaction [79], a survey of interaction with large displays [4], and a review on distant object selection methods [88].

2.1.3 Contributions

This review focuses on determining (1) the primary challenges research papers aimed to solve in VR object selection and manipulation research and (2) the existing solutions to these challenges. While numerous research papers are published annually, it is essential to summarize the scattered research endeavors and analyze critical research challenges and the corresponding state-of-the-art. This helps us reflect on the existing practices and identify the backbone topics and emerging trends in the research field. Furthermore, our work surveys and evaluates (3) how researchers measure their success under each research challenge. These essential measurements guide the development of our solutions.

2.2 Methodology

We followed the PRISMA guidelines [113] to select relevant publications for analysis. Our initial information sources of publications came from online databases and the most relevant literature review papers. We then applied the four-step process (identification, screening, eligibility, inclusion) to derive our final corpus. Figure 4 gives an overview of this filtering procedure.

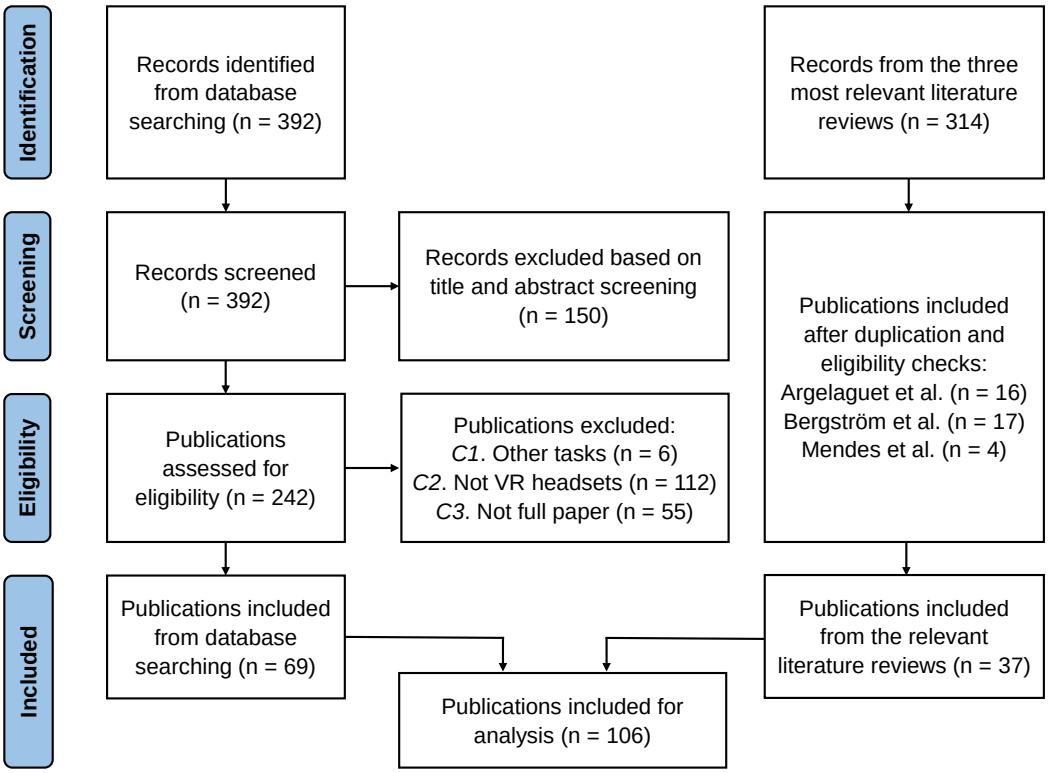


Fig. 4. PRISMA flow diagram of our systematic review.

2.2.1 Systematic Query Searches within Online Databases

To identify relevant, high-impact papers on object selection and manipulation in VR headsets, we first performed systematic query searches within online databases, including ACM Digital Library, IEEE Xplore, Wiley Online Library, Scopus, Taylor & Francis Online, and Springer Link. The publication venues included in the search were CHI, UIST, VRST, SUI, CSCW, Ubicomp, DIS, IUI, TOG, IMWUT, PACM HCI, TOCHI, IEEE VR (including 3DUI), ISMAR, TVCG, Computer Graphics Forum, IJHCS, Computer & Graphics, IJHCI, and Springer VR. These venues were selected based on the authors' expertise in HCI and VR, as well as their impact, according to Google Scholar Metrics.

To identify publications that are primarily relevant to object selection and manipulation in VR headsets, we used "*selection*", "*manipulation*", and "*virtual reality*" as our initial search terms in publication titles and iteratively derived their synonyms based on the literature present in the publication venues mentioned above. The new terms identified were "*pointing*", "*acquisition*", "*VR*", "*3D*", and "*immersive*". We did not include the term "*interact*" (as for object interaction)

or search the publication abstracts for keywords as they returned a large number of irrelevant records from the online databases. We documented our detailed search process and results in our supplementary material. A simplified example query in ACM Digital Library, without including the publication venues, is:

```
Title:((acqui* OR point* OR select* OR manipulat*) AND (virtual OR VR OR 3D OR Immers*))
```

Here, * denotes any number of unknown characters (wild cards). We thus were able to include other word forms such as "*manipulate*", "*manipulating*", and "*manipulation*". The word "*virtual*" was used to capture similar wordings of virtual reality environments such as "*virtual environment*" and "*virtual object manipulation*". In total, we obtained 392 records from searching the databases.

After obtaining these initial records, we first screened their titles and abstracts to exclude papers that were irrelevant to our exploration (e.g., constructing a 3D point cloud). This process left us with 242 publications. Next, we assessed the full text of these publications for eligibility according to three criteria: (1)not about object selection and manipulation; (2) not in VR headsets; (3) not a full paper. The first two criteria were based on the scope of this research. We also excluded posters and extended abstracts as they do not usually have the same level of maturity as full papers. At the end of this filtering procedure, we were left with 69 publications.

2.2.2 Records from Relevant Literature Reviews

In addition to performing query searches, we also examined all references in the three most relevant literature review papers to extract further papers relevant to our topic. This was to ensure that we included impactful papers that were not published in the selected publication venues or did not use our keywords in the title (e.g., object *interaction* instead of *selection* or *manipulation*). The literature review papers we used were: Argelaguet and Andujar's survey on 3D object selection techniques for virtual environments in 2013 [5], Bergström et al.'s papers on guidelines for evaluating VR object selection and manipulation in 2021 [13], and Mendes et al.'s survey on 3D virtual object manipulation in 2019 [105]. We assessed the papers' titles and full texts to exclude less relevant papers using the same criteria and made sure to remove duplication in the collected papers. At the end of this process, we were left with 37 publications.

2.2.3 Dataset and Coding Process

In total, we collected 106 publications (69 from online database query searches and 37 from the three most relevant literature reviews) as the corpus for further analysis. With this corpus, we first coded the challenges, research goals, proposals/methods, and measurements of success in text fields by collecting quotations from the papers. We then iteratively defined and categorized challenge types across the papers and further distilled 8 core challenges. Both preliminary and

final challenge types were coded categorically. We also coded relevant information such as contribution types, solution types, study types, and success measure types categorically by referencing the categories in previous research [85] and iteratively defining them. Some papers had made multiple contributions and proposed various solutions, and we thus distinguished their primary and secondary contributions and solutions in our coding. Readers can find more details in our coding manual.

2.3 Overview of Contribution Types

We investigated the contribution types of the 106 publications in our corpus according to Wobbrock and Kientz taxonomy [188]. Figure 5 summarizes the results. A significant portion of the papers contributed new *artifacts* (42 papers, 39.6%), including, for example, new interaction techniques for occluded target selection [152, 174, 198], systems for grasping rendering [34, 119], and novel haptic devices [7, 43, 87]. Another mainstream of the papers focused on *empirical* contributions (49 papers, 46.2%), where user studies were carried out to evaluate or compare technological solutions [80, 127], fine-tune design parameters [144, 182], investigate the effects of a factor [10, 78], or explore design possibilities [91, 191]. There were four *methodological* papers (3.8%) on standardizing the research practices in VR object selection and manipulation [13, 19, 20, 137]. Eight were *survey* papers (7.5%) that have provided a new taxonomy of the techniques [31, 105] or intended to answer specific research questions [36]. Three papers (2.8%) have a *theoretical* emphasis on initiating new design spaces or frameworks that could motivate new interaction techniques [111, 129, 160]. Note we classified qualitative models, such as models that predict selection endpoints [61, 194], as either empirical or artifact contributions. While these models may have predictive power, they do not aim to provide a systematic set of statements that explains the fact (e.g., why the endpoints distribute in a certain way), which is an essential component of a theoretical contribution [141]. None of the papers has the primary contribution of datasets or opinions.

2.4 Research Challenges and Existing Solutions

We identified eight research challenges and their corresponding solutions for VR object selection and manipulation research. We iteratively defined these eight core challenges by surveying the problems research publications in our corpus aimed to solve and their research goals. Throughout our categorization process, we wanted to capture popular research challenges that have already attracted many researchers to try to address them. We were also interested in identifying emerging topics that now have a limited number of publications but may still be promising for future research to consider. Table 2 summarizes these research challenges and solutions.

Fig. 5. Number of publications under the contribution types proposed by Wobbrock and Kientz [188].

We note that our categorization of the research challenges is not mutually exclusive, and many papers presented within each section may tackle one or more challenges. Our goal was to capture and classify the primary obstacle that a research paper aimed to resolve and the main solution offered by the paper. Through this process, we can draw a clear picture of the representative themes in the VR selection and manipulation literature. We also note that we excluded general surveys that do not tackle specific challenges (but summarize them or their solutions) [5, 31, 57, 64, 79, 105, 163] and an early programming implementation of basic interaction techniques [142] in this analysis.

2.4.1 Complexity in 3D Interaction Scenarios

Though VR technology may create unprecedented opportunities for new types of interaction, developing appropriate VR interfaces for selection and manipulation is not trivial. As illustrated in the introduction of this thesis, Virtual Hand and Raycasting may not be sufficient for more complex scenarios that contain small, faraway, and occluded targets or require precise, versatile, and prolonged operations. Therefore, the papers under this theme aim to develop optimal selection and manipulation interfaces for simple and more complex 3D VR interaction scenarios.

In the following, we present thirty-four papers that address complexity in 3D interaction scenarios. Among the selected papers, most of them (27 papers, 79.4%) primarily contributed new artifacts, including (1) *interaction techniques*, the fusion of input and output for users to complete tasks in human-computer dialogues [45, 169], (2) *devices*, the hardware pieces employed by users to communicate with a computer [24, 65, 99], and (3) *models*, the computational assistance that improves the usability of an interface through user behavior prediction [61]. The remaining papers (7 papers, 20.6%) primarily presented empirical contributions. Empirical user studies were carried out to derive (1) *design knowledge*, the body of knowledge that can be used in similar application scenarios, e.g., the advantages and disadvantages of a technique [172],

Table 2. A summary of research challenges and solutions on VR object selection and manipulation.

1	**Complexity in 3D Interaction Scenarios** Challenge - Selecting and manipulating 3D virtual objects in VR headsets can be challenging because the interaction scenarios may contain small, faraway, occluded, out-of-view, and multiple targets and the tasks may require precise, versatile, and prolonged control. Solution - Developing optimal selection and manipulation designs for simple and complex (e.g., distant, occluded, out-of-view targets) VR scenarios.
2	**Underexplored Interaction Spaces and Factors** Challenge - Understanding new opportunities (i.e., design spaces or ways of interaction) and considerations (i.e., factors that influence user behavior or responses) of 3D user interfaces. Solution - Conducting usability studies on (1) possible ways to offer new interaction (e.g., 3D eyes-free selection) and (2) scrutinizing how specific factors (e.g., the presence of multimodal feedback and visual avatar) influence user performance, experience, and behavior.
3	**Unknown Comparative Usability** Challenge - The lack of understanding or guidelines of the relative usability between different solutions to inform "which method(s) to choose under a given situation". Solution - Conducting usability studies on comparing alternative choices of devices (e.g., game controller vs. 3D pen-like device), modalities (e.g., gaze vs. hand vs. head), and techniques (e.g., Raycasting vs. Virtual Hand).
4	**Ergonomic Issues: Workload and Fatigue** Challenge - Limitations regarding users' physical interaction space (e.g., space constraints). Solution - Developing techniques that aim to fulfill users' comfortable requirements.
5	**Imprecise Rendering of Visual and Haptic Realism** Challenge - Enabling realistic visual and haptic rendering during object selection and manipulation under hardware limitations and form factor constraints. Solution - (1) Proposing algorithms for realistic hand rendering, (2) building devices for simulating different haptic features (e.g., textures, shapes, and stiffness), and (3) conducting usability studies to explore methods that can improve perceived visual and haptic realism.
6	**Underdeveloped Evaluation Methodology** Challenge - Standardizing the practices of evaluating selection and manipulation solutions to allow the generalization of results across studies. Solution - Building relevant testing framework, testbeds, and guidelines.
7	**Limited Support for Collaborative Object Manipulation** Challenge - Simultaneous manipulation of a virtual object with multiple users. Solution - Building framework and techniques to enable simultaneous object manipulation.
8	**Context Integration and Workflow Optimization** Challenge - Integrating selection and manipulation into the "broader" context and workflow. Solution - Developing techniques that consider the context and simplify the workflow.

(2) *design recommendations and guidelines*, the explicit set of "rules" that inform future designs [107, 168, 195], (3) *desired design parameters*, the setting of design parameters where the proposed solution can be the most useful [90], and (4) *models* [194], verbal or mathematical representations that describe and predict the characteristics of human-computer interactive dialogues [99]. We elaborate on existing solutions for VR selection and manipulation as follows.

§1 Selection Approaches. Many proposed techniques improved the selection efficiency and accuracy by adjusting the criteria of how the selection of a target is determined. Rather than requiring a tiny virtual pointer to be exactly "on" the targeted object, an enhanced technique may select the closest object to the pointer [9, 161], scale up the cursor size [46, 96], leverage computational models to predict the intended target [61, 194], or introduce crossing-based [168] or multi-step selection techniques [106]. Techniques also added an extra dimension of movement (moving along the depth dimension) to the Raycasting pointer [9] or distributed multiple 3D cursors across the space [148]. Moreover, they incorporated multi-modality support with pen-based input (that leveraged dexterous movements of fingers) [90] and synergetic gaze and head-based input [153].

Other selection techniques were developed to handle more complex 3D VR environments that contain distant, occluded, out-of-view, or multiple targets. While a user can only select objects within the arm-reach distance with Virtual Hand, assistant techniques may extend the movement of the virtual hand [18, 134] or create a reachable replica of the virtual environment or its elements [128, 163]. For partially or fully-occluded targets, existing techniques, including our research presented in Chapter 4, leveraged dis-occlusion visualizations (e.g., making distractors transparent or translating candidate objects into new locations) to identify the target [174, 198]. Techniques also modified selection mechanisms (e.g., gaze-based outline pursuits [152], Bézier curve-modified selection ray [44]) to acquire such occluded targets more robustly. For an out-of-view target, proposed techniques may guide the user towards its location through, for example, vibrotactile cues [82]. If there were multiple targets in the scene, techniques could create a selection volume via, for example, a volumetric cube, a lasso, or a virtual tablet and further progressively refine the selection [72, 114, 162].

§2 Manipulation Approaches. The literature presented two main methods to improve the usability of VR object manipulation: degree-of-freedom (DoF) separation and control-display ratio (CD ratio) adjustment. DoF separation-based techniques reduced the number of DoFs being controlled simultaneously compared to Virtual Hand (which has 3 axes for translation, rotation, and scaling). For example, researchers adapted 3D virtual widgets similar to those used in desktop CAD software (e.g., Unity, Blender) for VR headsets [23, 86, 107, 108]. They further enabled user-defined 3D anchor points or transformation axes [51, 108]. CD ratio adjustment-based techniques dynamically increased or decreased the movements of the virtual hand compared to the corresponding physical hand [108, 134]. For example, scaling up the

movement may allow coarse, rapid manipulation while scaling it down may enable more fine-grained transformation [47, 48]. Additionally, previous research has also combined Virtual Hand and Raycasting [157, 172], designed finger gestures for rotation control [157], allowed users to impersonate an under-manipulated object [173], and incorporated gaze input into the manipulation process [195] (our research in Chapter 5).

2.4.2 Underexplored Interaction Spaces and Factors

One primary goal of HCI research is to understand users' needs towards computing interfaces and map out new spaces of designs. Shifting from traditional 2D interfaces like PC screens and tablets, many research questions exist on how to best leverage the 3D virtual space for interactions [83]. Specifically, there is a need to understand the new opportunities (i.e., design spaces or ways of interaction [56, 60]) and considerations (i.e., factors that may influence user behavior or responses) that 3D interfaces may bring. Therefore, determining underexplored interaction spaces and factors is another major challenge that many papers aimed to resolve in the literature.

Twenty-eight papers in our corpus aimed to explore new interaction spaces and factors that may enhance VR selection and manipulation. The majority (24 papers, 85.7%) focused on empirical contributions through discovering design knowledge, design recommendations and guidelines, desired design parameters, and models. There was one survey paper on conducting a meta-analysis to derive guidelines [36], two theoretical papers on a framework [111] and a conceptual model [160] of underexplored spaces, and one artifact paper on a novel device to offer new ways of interaction [55].

A collection of papers examined new design spaces to offer interaction. They considered, for example, the feasibility of eyes-free target acquisition [111, 190, 191, 202], the practicability of freehand pointing without a selection ray [30, 104], and the usefulness of modifying control-to-display mappings (input scale [49, 81, 182], direct vs. indirect input [81], and cursor offset [89]). They also tried to understand how users prefer to select and manipulate objects in VR [121, 184]. Moreover, they investigated the locations of providing 3D virtual interfaces (e.g., arm-anchored [91], smartphones [81], fovea and periphery regions [76], user's own body [111], and a display attached to the face [55]) and the spatial and temporal aspects of selection [160].

Another set of works scrutinized how specific factors presented in user interfaces may influence performance, kinematic features, and user perception during VR selection and manipulation tasks. These factors include multimodal feedback [6], interaction fidelity (e.g., widgets vs. physically grabbing items) [143], the presence of a virtual avatar [33, 36], the aptitude and experience of individuals [183, 185], perception of redirection [32], the absence of haptic feedback during VR manipulation [97], and object features like size and distance (e.g., [183, 194]).

Other works explored the impact of device-related factors on VR selection, including vergence-accommodation conflict [10], stereo deficiency [11], and jitter of input device [12].

2.4.3 Unknown Comparative Usability

While numerous new solutions have been developed for interactions in VR every year, their comparative usability is not always clear, such as effectiveness, efficiency, and satisfaction [66]. The lack of understanding or guidelines of the relative usability makes it hard to choose a more suitable approach for different applications. To solve this challenge, some research is dedicated to comparing the usability among different VR selection and manipulation solutions. These studies aim to inform the design decision of "*which method(s) to choose under a given situation*".

Notably, comparing usability among different solutions is common in the relevant literature. The unique point of the research studies summarized in this category is that they typically do not propose new interactions or explore new interaction spaces. In contrast, they leverage existing solutions and compare them under new conditions.

We identified fifteen papers in our corpus where the primary goal was not to develop new methods but to perform rigorous empirical evaluation studies that compare choices of devices [3, 17, 80, 112, 127], modalities [29, 39, 68, 112, 124, 138], and techniques [78, 103, 135, 136, 177]. They all were empirical contributions, focusing on developing design knowledge, guidelines, and recommendations. For example, existing studies compared displays (e.g., VR, AR, and PC screens) and input devices (e.g., handheld controller, bare hand, 3D pen-like device, and mouse) for object selection and transformation tasks [3, 80, 127]. A few studies measured the performance of different input modalities (e.g., eye, hand, head, and muscle contraction) for VR object selection [29, 68, 124, 138]. They also analysed feedback modalities like auditory and force and derived design guidelines based on the study results [39]. Moreover, researchers also conducted empirical studies to compare the ability of DoF control during object manipulation [78], visualization techniques for precise object alignment [103] and fixed vs. handheld menus for selection [177].

2.4.4 Ergonomic Issues: Workload and Fatigue

Ergonomic assessments on workload and fatigue have been applied extensively to evaluate and compare different selection and manipulation approaches [13]. Measurements through self-reports (like NASA-TLX [58] and Borg CR10 [15]) are often included in studies as accompanying metrics. However, existing VR interactions may still require large, cumbersome body movements, overlooking the limits of a user's physical interaction space, comfortable requirements, and mobility issues [115, 179]. Therefore, recent research investigates the challenge of improving user comfort within constrained, physical, and operational spaces during VR interactions.

Our corpus presented two papers that address ergonomic issues related to workload and fatigue during VR interaction. Both papers primarily contributed new artifacts that leveraged new interaction techniques, while one also proposed design recommendations [179]. Montano et al. [115] proposed an optimization-based retargeting strategy to relocate visual targets to more convenient reaching positions. Wentzel et al. [179] investigated non-linear virtual hand amplification functions to improve arm ergonomics while maintaining body ownership. Both methods made the VR interaction experience more comfortable and accessible.

2.4.5 Imprecise Rendering of Visual and Haptic Realism

With advances in optics and audio technologies, current VR headsets can provide people with an improved sense of presence in simulated realities, creating a fully immersive experience [8]. However, realism often breaks when users attempt to grab and manipulate virtual objects: their virtual hands/fingers can pass through the object [119], and they cannot “feel" the physicality of the object in the real world [144, 149]. For Virtual Hand-based selection and manipulation methods that mimic real-world experience, the challenge is to discover realistic visual and haptic rendering techniques under hardware limitations and form factor constraints.

We identified four papers on achieving visually realistic grasping of objects during VR manipulation. Three were primarily artifact contributions on new rendering systems, and one was an empirical contribution that evaluated alternative visual representations. Oprea et al. [119] proposed a system that automatically fits a hand to the shape of virtual objects during grasping. Delrieu et al. [34], and Sorli et al. [158], realizing there might be inherent mismatches in the tracked hand and the virtual hand during hand-object manipulation without a real physical object, introduced strategies that balance between the tracked and the simulated hand to enable fine manipulation. Dewez et al. [35] considered the visual realism of users’ avatars when using techniques that adjust the CD ratio during selection and manipulation (e.g., Go-Go [134]) and examined dual representations of a user’s virtual body.

Our corpus also included five papers on providing *active* or *passive haptics* to enable haptic renderings like textures, shapes, stiffness, and weight of objects during VR manipulation. Four were artifact contributions on new haptic devices, and one was empirical contributions on determining design parameters for more believable haptics. A few papers focused on active haptic techniques that exert forces onto virtual contact areas through haptic devices to simulate a compelling interaction experience [7]. Schorr and Okamura [149] and Lee et al. [87] built wearable devices to trigger haptic feedback on users’ fingertips. In contrast, others examined passive haptic approaches that leverage a pre-defined set of physical props as proxies of virtual objects. For instance, Arora et al. [7] used custom-designed LEGO bricks to simulate various object shapes. Feick et al. [43] further used composable shape primitives and connectors to simulate the haptic sensations of a complex virtual model. While providing a matching physical prop for every virtual object is not scalable, Samad et al. [144] created illusions of the changed

weight of virtual objects with limited physical props by adjusting the CD ratio within an appropriate range.

2.4.6 Underdeveloped Evaluation Methodology

A valid, reliable, and reproducible evaluation methodology is the cornerstone for assessing the usefulness and effectiveness of a new method for selection and manipulation [188]. Results yielded under rigorous evaluation methodologies can accumulate replicable findings, provide design guidelines, and potentially enable the comparison of techniques across studies [13].

The initial obstacle of this space was to build a representative set of VR interaction tasks, task parameters, and evaluation metrics so that the research findings could be generalized beyond a particular experimental setting [19, 20, 83, 137]. However, with the evolution of VR technology, the challenge shifted towards designing evaluation studies that may consider a variety of new, important factors that are not covered in a canonical task setting while preserving generalizability [13, 199]. Ultimately, these methodological works aim to standardize the practices in technique evaluation [13].

Our corpus contained five papers on standardizing evaluation methodologies of object selection and manipulation in VR. Four methodological contributions involved testbeds, frameworks, and design guidelines that inform how to conduct empirical studies. One empirical contribution investigated whether specific factors could influence the validity of user evaluations.

Poupyrev et al. [137] and Bowman et al. [19, 20] formalized the early testbeds for technique evaluation. Poupyrev et al. [137] presented *VRMAT*, a testbed containing three basic interaction tasks (select, position, and orient) with their corresponding independent variables and evaluation metrics. Bowman et al. [19, 20] further suggested that an interaction task (e.g., colouring an object) can be broken down into several sub-tasks (e.g., selecting an object, selecting a colour, and applying a colour). Each sub-task can be achieved by various interaction techniques, which can be evaluated by manipulating important outside factors (like task characters and environments). More recently, Yu et al. [199] investigated the potential issue of disengagement with long, repetitive selection experiments and evaluated motivational strategies to incentivise participants during such experiments. Bergström et al. [13] analyzed research works in evaluating object selection and manipulation from 2000 to 2019 and proposed recommendations and checklists on task design and result reporting for guiding future studies.

2.4.7 Limited Support for Collaborative Object Manipulation

When multiple users collaborate in VR, a common need is to move and modify objects within the virtual space cooperatively [41, 140]. For example, users may need to assemble a complex object together [175], modify a 3D data visualization concurrently for exploration [14, 38], and place digital furniture at different locations for configuration testing [140]. Existing research identifies the challenge of simultaneous manipulation of a virtual object with multiple

users [84, 129, 130, 156, 175]. When two or more users want to manipulate the same virtual object, it is essential to determine who should control the object for better efficiency and user experience.

Our corpus captured two papers on providing simultaneous object manipulation in VR headsets. There was one theoretical contribution and one artifact contribution. Pinho et al. [129] introduced a conceptual framework (*Collaborative Metaphor*) that considers which input technique to use, how to combine them, and how to display a user's action to others in a collaborative task. They also presented interaction techniques that, for example, allowed users to control different transformations (like managing either translation or rotation) or employ different input techniques (using either Raycasting or Virtual Hand). Wang et al. [175] proposed an interaction technique that determines the dominant manipulator based on a viewport quality function that examines quantities like object visibility and distance of the target.

2.4.8 Context Integration and Workflow Optimization

Though selecting or manipulating an object is typically treated as individual tasks in research studies, they are associated with scene and interaction contexts in realistic applications. For example, a selected object may belong to a group of closely-related objects, which are often manipulated together [176]. A manipulation gesture may result in multiple consequences because the same gesture is used for several tasks [27, 100]. Integrating selection and manipulation techniques into the "broader" context and workflow is another challenge based on the literature.

We identified three recent papers that proposed new artifacts (specifically, interaction techniques) on this topic. Mardanbegi et al. proposed *EyeSeeThrough* that simplifies the process of tool selection and application: users can visually align a target object with the tool at the line-of-sight to apply the tool to the object, rather than performing a tedious two-step operation of first selecting the tool and then selecting the target [100]. Chen et al. proposed a technique that resolves ambiguous hand manipulation gestures (e.g., hand movements can either displace or stretch an object) with a pop-up menu that can be interacted with head gaze [27]. Wang et al. developed a method that considers scene context information, such as object semantics and interrelations, when selecting or moving an object. For example, the technique can automatically adjust the yaw of a chair during translation to make it face a nearby table [176].

2.4.9 Summary Statistics

We analyzed how the number of publications under each challenge changed over the years. The results are summarized in Figure 6. The total number of publications has significantly increased in recent years (since 2017) because of the advancements made in the off-the-shelf

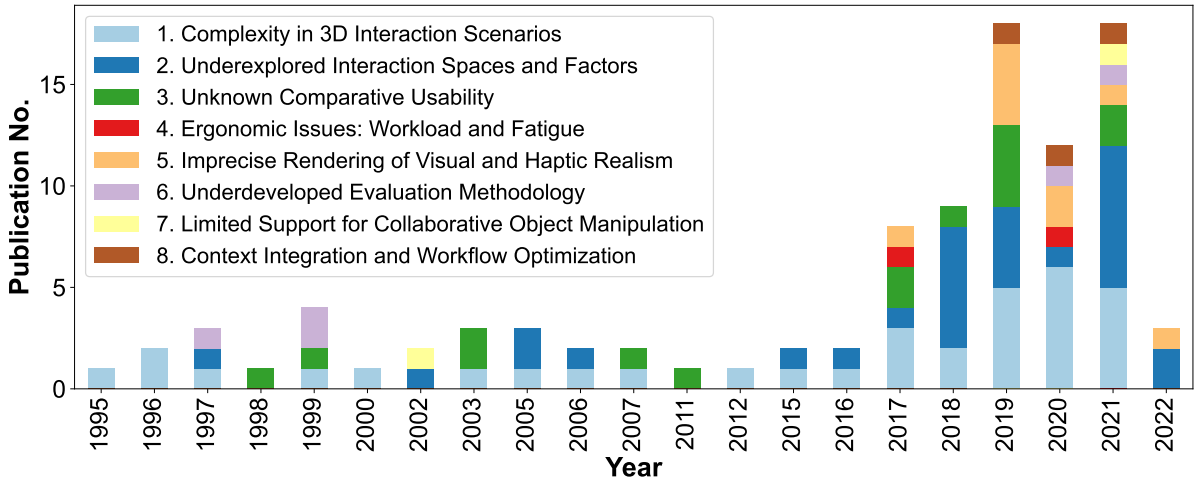


Fig. 6. Number of publications under each challenge by year.

headsets and development kits. Note that our literature review was initiated in early 2022, so limited publications were captured for this year.

The topics of complexity in 3D interaction scenarios, underexplored interaction spaces and factors, and unknown comparative usability have remained attractive and mainstream since the 1990s. There are also emerging themes where all relevant publications appear in more recent years. Researchers gained interest in resolving ergonomic issues related to workload and fatigue, rendering more precise and believable visuals and haptics, and integrating selection and manipulation techniques into a broader interaction context. One interesting observation is that the publications on evaluation methodology were present early in 1997 and 1999, remained silent between 2000 and 2019, and were picked up again until more recently (2020 and 2021). Also, the topic of collaborative object manipulation appeared in 2002 and was continued in 2021.

2.5 Measuring Success

According to the presented research challenges and solutions, we further investigated and reflected on how authors of the selected papers measure the success of their solutions. Based on the literature, we first categorized nine success measures (effectiveness, efficiency, ergonomics, experience, robustness, expressivity, realism, behavior, and consistency). We then analyzed how these measurements were applied in each research challenge.

2.5.1 Measures

We first summarised and categorized the success measures used in the papers. During the iterative development process, we borrowed concepts from Hornbæk's work [66] on usability measures while extending the original classifications (effectiveness, efficiency, and satisfaction) to a more detailed, domain-specific version with nine measures. For example, we distinguished efficiency in terms of completion time and workload into two different usability measures (efficiency and ergonomics) to improve the granularity. We also introduced new dimensions more relevant to VR selection and manipulation research, such as robustness, expressivity, and realism.

- *Effectiveness.* Effectiveness represents "the accuracy and completeness with which users achieve specific goals" as according to ISO 9241 [71]. Specific measures used in our corpus include error rates (e.g., percentage of incorrect selections [11, 96]), error distances or rotations (e.g., offsets between the target and the actual input [87, 104, 107, 175]), false positives/negatives (e.g., in a group selection scenario [114, 162]), and task completion (e.g., completion rates [48, 51]). They also involve prediction accuracy of a model [30, 61, 194] and may get incorporated into throughput measures [6, 68].

- *Efficiency.* The ISO 9241 (2018 version) defines efficiency as "resources used in relation to the results achieved" such as time, human effort, and materials [71]. To make it more specific to our tasks, we considered efficiency as the time cost associated with the results achieved. The typical measure in our corpus is task completion time (e.g., cursor movement time [127], selection time [199], manipulation time [34]). They also get involved in throughput measures [91].

- *Ergonomics.* While ergonomics is a broad term in certain contexts, we here restrict it to the physical or mental workload associated with the results achieved. Objective quantification (approximation) of ergonomics employed in our corpus include hand/arm movement distance [96, 195] and RULA (rapid upper limb assessment) score [115]. Subjective measures related to ergonomics contain questionnaire results from NASA-TLX [55], Borg CR10 [179], customized scales of fatigue and comfort [96, 127], and qualitative feedback [90].

- *Experience.* This encapsulates users' feelings and satisfaction when performing tasks with the evaluated solutions [65]. These data are normally collected from questionnaires. The measures include, but are not limited to, overall impression [43], general user experience [176, 198], satisfaction [68], preference [49, 111, 127], sense of control [51, 68], body ownership [33, 35, 87, 144, 179], ease-of-use [153, 172], fun [51, 107], perceived performance [81, 96, 173], perceived ease of learning [27, 47], perceived usability [103], intuitiveness [27, 96], sense of presence [86, 175], immersion [143], engagement [143], obtrusiveness [112, 198], and sickness [182, 191].

- *Robustness.* A robust solution remains useful under different testing conditions, especially if the solution has been evaluated to achieve good performance under more challenging scenarios. It can also mean that a derived conclusion performs consistently across multiple studies. For example, researchers have tried to evaluate their solutions under difficult scenarios (e.g., wider or untested conditions for a predictive model [194] and high occlusion scene [96, 152]) to test their robustness. They have also performed meta-analyses to determine robust conclusions [13, 36].

- *Expressivity.* This means that a solution can be applied for a wide range of interaction scenarios or even enable new use cases. To demonstrate expressivity, researchers often present a section of application scenarios in the paper (e.g., [43, 176]). For instance, when introducing the haptic device *VirtualBricks* [7], the authors also detailed example applications such as its use in first-person-shooter games, fishing, disco, etc. Additionally, a framework or testbed may illustrate its expressivity through sample techniques and use cases [20, 129, 137].

- *Realism.* Realism (sometimes dubbed as naturalness [34, 158]) is defined as how well the way of interacting with virtual objects corresponds to the way of interaction in the physical world. We consider it separately from experience measures as it emphasizes the cognitive judgment of physical-virtual mismatches more than the interaction experience itself. Realism is also different from body ownership, the psychological mapping of one's real body to a virtual body [155], and sense of embodiment, the illusion that the co-located virtual body has effectively replaced their physical body [52]. Realism is typically assessed through customized scales [34, 112, 158], discrimination tasks (e.g., weight discrimination [144, 149]), or a qualitative interview [144].

- *Behavior.* User behaviors are likely to change if a new solution is adopted. Several papers in our corpus have demonstrated that different approaches could influence interaction strategies [172, 195], movement profiles [6, 182, 194], and Fitts's law parameters [68, 168]. A few showed that their solutions could encourage positive behaviors in an interaction context. For example, the solutions can increase user participation [175], cursor speed [127], and maximum reach distance [35]. They can also decrease the number of iterations to complete a task [135], the number of target re-entries [10, 11], and the number of times that users press the trigger button [49].

- *Consistency.* Consistency, in our case, means that a solution could maintain its performance over an extended period. A few papers in our corpus have checked the performance of their solutions in a prolonged interaction scenario. They found the performance (e.g., completion time and error rates) could be influenced by learning/practicing [35, 168, 183], fatiguing, and disengagement [199].

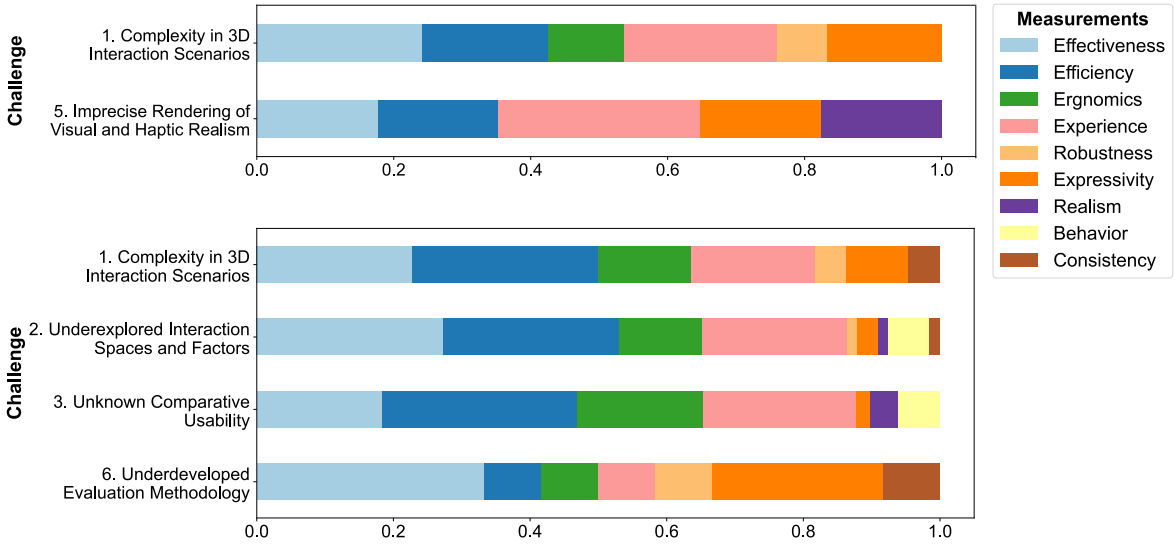


Fig. 7. Top: The likelihood of success measurements being used by the authors to argue their interaction techniques, models, devices, and systems to be "better than" or "comparable to" previous or other alternative approaches in artifact papers. Bottom: The likelihood of success measurements being used in empirical, methodological, theoretical, or survey papers to evaluate the potential solutions.

2.5.2 How Solutions Address Each Research Challenge

We assessed how success measures were used in each paper, which aimed to address the aforementioned research challenge. In artifact papers, a success measure refers to the evidence the authors provide to claim their proposed interaction techniques, models, devices, and systems, to be "better than" or "comparable to" previous or other approaches. In empirical, methodological, theoretical, or survey papers, all the evaluation metrics were considered as the success measures—we assumed that the authors considered the evaluation metrics essential for a successful solution to use them in the study. Because of this inherent difference between contribution types, we analyzed them separately.

We first investigated the likelihood (percentage) of success measures being used to evaluate the solutions to each research challenge (Figure 7). We removed a challenge category for measurement percentage analysis if it consisted of fewer than 5 papers (i.e., a small sample size). We then performed additional analysis to offer insights tailored to the contribution types.

§1 Artifacts. For the challenge of managing the complexity in 3D interaction scenarios, the corresponding artifact papers emphasized more on effectiveness (13/27, 13 out of 27 papers),

efficiency (10/27), experience (12/27), and expressivity (9/27) to demonstrate the success of their solutions. That is, the proposed solution was often argued to achieve better performance, such as faster completion and fewer errors, provide more satisfactory user experiences, and be suitable for various application scenarios. Ergonomics (6/27) and robustness (4/27) measures were less often used in the arguments, and there was little attention to realism, behavior, and consistency measurements.

Regarding the challenge of the imprecise rendering of visual and haptic realism, the dominant measurement was experience (5/7), followed by effectiveness (3/7), efficiency (3/7), expressivity (3/7), and realism (3/7). This indicated that while the solutions might be proposed to improve realism, they could also enhance user experience (e.g., body ownership, sense of embodiment) and performance. The solutions were often demonstrated to remain useful in many application scenarios.

Further analysis of all artifact papers suggested that when a solution achieved better performance (effectiveness or efficiency), the probability that it outperformed other solutions in the experience measure was 76.9% and in the ergonomics measure was 34.6%. If performance was improved, the likelihood that the artifact performed superior in both effectiveness and efficiency measures was 34.6%. There were 24.4% of the solutions performed better in more or equal to four measurements.

§2 Empirical, methodological, theoretical, and survey. For the challenge of complexity in 3D interaction scenarios, the empirical papers employed similar measurements as in the artifact papers. More papers evaluated effectiveness (5/7), efficiency (6/7), ergonomics (3/7), and experience (4/7), with limited analysis on robustness (1/7). One paper (1/7) measured consistency in learning the techniques over time.

When exploring new interaction spaces and factors, a large number of papers focused on standard measurements such as effectiveness (18/27), efficiency (17/27), and experience (14/27). Ergonomic measures, including fatigue and workload, were also used in some cases (8/27). Only a few papers assessed robustness (1/27), expressivity (2/27), realism (1/27), and consistency (1/27) measures. As behavior measures(4/27), there were explorations on whether the potential solutions could encourage positive user behaviors, such as decreasing the re-entry rate.

Similarly, more frequent measurements when comparing alternative solutions were effectiveness (9/15), efficiency (14/15), ergonomics (9/15), and experience (11/15). There were also limited analyses on expressivity (1/15), realism (2/15), and behavior (3/15).

For developing evaluation methodologies, the papers mainly demonstrated that their framework or testbeds could achieve the desired purposes (effectiveness: 4/5) and be adapted to new application scenarios (expressivity: 3/5). One empirical paper also investigated the effect

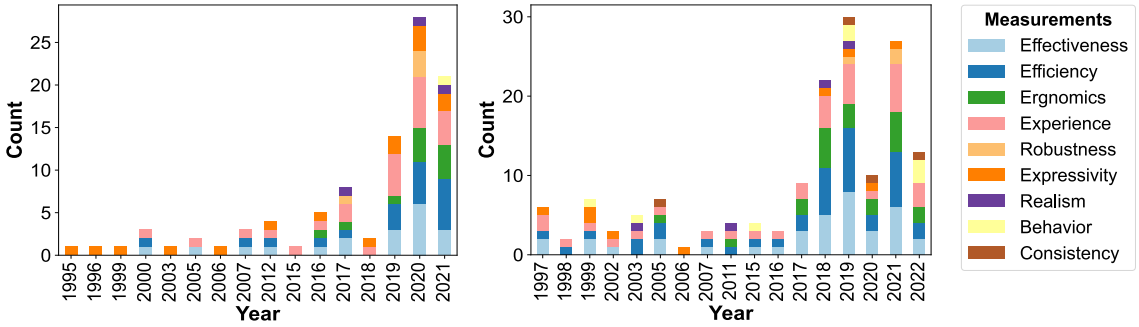


Fig. 8. Left: How new artifacts outperform alternative baselines in different success measurements across years. Right: How successful measurements were employed in empirical, methodological, theoretical, and survey papers across the years.

related to efficiency, ergonomics, experience, and consistency when adjusting the evaluation methodology.

§3 The usage of different measurements over time. In Figure 8, we summarized how success measurements were applied in different types of papers over the years. For artifact papers, there seemed to be a trend that more recent solutions were evaluated to outperform baselines in more diverse sets of successful measurements (i.e., more colors in the stacked bars). For empirical, methodological, theoretical, and survey papers, different measurements were commonly employed to evaluate different perspectives of the solutions across years.

2.6 Discussion

Based on our literature review, we first discuss findings on classic challenges and emerging trends in the field of VR selection and manipulation, together with an overview of the solutions. We then discuss how success measurements have been leveraged to address research challenges.

2.6.1 Classic Challenges and Emerging Trends

By categorizing the research challenges the publications aimed to solve, we see both classic hurdles that have been actively investigated throughout the years and emerging trends that only contain a small number of papers for now but will potentially have significant influences.

Classic research challenges were raised in the early days when there was little understanding of designing appropriate VR user interfaces for selection and manipulation (Challenges 1-3). Researchers prototyped solutions to interact with objects in the surrounding 3D virtual space, explored new design spaces and features unique to 3D interaction, and compared

alternative solutions for the best performance and interaction experience. With a more advanced understanding of the space, the broad research challenges have been expanded to a multitude of specific sub-areas such as target occlusion [152, 174, 198], eyes-free acquisition [111, 190, 191, 202], multi-modality integration [90, 102, 153, 197], and interaction with virtual avatars [33, 36]. We have detailed descriptions of these challenges and their solutions in Section 2.4.

There are also small but emerging topics in the field that are worth attention: coping with the limitations in a user's physical space, rendering precise visual and haptic realism to reduce noticeable sensory mismatches, and integrating the selection and manipulation tasks into broad contexts and workflows (Challenges 4, 5, and 8). We expect further evolution of these research challenges and their solutions to a more mature state in the future. For example, with the rise of the challenge of ergonomic issues, it seems that the researchers have been putting more emphasis on designing for users themselves rather than their performance improvements. The topic could be further extended to consider accessibility issues, where users might have physical constraints with their bodies or environment-imposed situational disabilities (e.g., a person holding groceries might not be able to use their arms for other tasks) when interacting with VR systems [94, 116, 201]. We also envision the application of AI technologies and novel research concepts to help systems better understand the environmental context and the user's needs and provide timely assistance [73, 126, 193] and more believable experiences [42, 166] during VR object selection and manipulation.

Two challenges were investigated in the early days and were revisited more recently. One is the challenge of underdeveloped evaluation methodology (Challenge 6). Though the existing evaluation framework is still helpful in ensuring internal validity (i.e., study rigor), experimental factors that could influence the study results (e.g., target size, distance, arrangement, density, occlusion, the presence of virtual avatar, background setting, etc.) are becoming too overwhelming to be fully-crossed in a user study. It is thus difficult to determine to what extent the study results could be generalized to the intended applications and whether it is suitable to compare the results across studies [13]. The other returning challenge is the limited support for collaborative object manipulation (Challenge 7). With the growing commercialization of VR systems, it would be advantageous if users could complete tasks that require simultaneous manipulation of virtual contents with collocated or remote peers [53, 54, 130, 140, 180]. We expect to see more explorations that consider the unique affordances of immersive VR headsets and the cooperation of multiple devices (e.g., VR headsets with AR glasses, tablets, and desktops) in object selection and manipulation.

2.6.2 Usage of Success Measurements

According to our results, current measurements of artifacts and empirical studies center mainly around the *5Es* (effectiveness, efficiency, ergonomics, experience, and expressivity). They cover

both objective performance measures and subjective feelings, and the solutions should also be demonstrated to be well-suited for various use cases. These appropriate and essential measurements pave the way for us to resolve the research challenges.

Meanwhile, considering success measurements related to robustness, consistency, and behavior can give a more comprehensive picture of the solution. While these measures might not apply to every scenario, we should reflect upon them when evaluating new VR object selection and manipulation solutions. Some example questions include (1) Robustness: does the solution remain helpful in more extreme scenarios? (2) Consistency: does the usability of the solution increase or decay over time, either in the short term or in the long run? (3) Behavior: how does the solution reshape user behavior? Does the solution encourage positive user behavior?

We should also note that the successful measurements may correlate or conflict with each other. Our analysis has shown that performance measures (effectiveness and efficiency) correlated with experience measures quite well; 76.9% of the proposed artifacts outperformed the baselines in experience measures if they achieved better performance. In other cases, researchers and designers might need to decide the tradeoff between the measures like speed vs. accuracy tradeoff [133] and flexibility (expressivity) vs. efficiency tradeoff [92]. For example, while improving performance regarding effectiveness and efficiency can be essential, it might not be desirable if achieved at the expense of increased cognitive load [5]. Moreover, users may prefer an interface that does not necessarily improve performance [125, 139]. These previous findings served as helpful guides for the research in this thesis.

2.6.3 Limitations of the Literature Review

§1 Completeness. While we employed both systematic query searches and key literature identifications to build our corpus, we acknowledge that we could inadvertently miss some relevant papers or extended abstracts (e.g., [132]). In other words, this corpus cannot be treated as an exhaustive and complete list of VR object selection and manipulation research. We highlight that our goal was to identify key challenges, solutions, and success measures in the domain, and the current corpus serves as a representative subset of the most relevant papers. We aim to address this inherent limitation of a systematic review by making our dataset and search queries transparent and open-source for future research to iterate and expand upon.

§2 Reality-Virtuality Continuum. In this research, we only included relevant research built with VR headsets but not other types of VR displays. Our rationale was based on Bowman et al.'s research [17, 98] on comparing multiple interaction techniques under different displays (i.e., CAVEs and HMDs). They found migration of techniques to other displays could sometimes work but could also cause serious usability problems due to their different display properties. Thus, it was difficult to justify whether solutions that worked on other displays could also be transferable to VR headsets. Therefore, we restricted our scope to VR headsets for simplicity.

Furthermore, we only investigated fully-immersive VR rather than AR and MR. Many VR headsets offer video see-through or pass-through mode, enabling both AR and MR (e.g., Sutherland's early work on The Sword of Damocles [83, 165]). We excluded them because the presence of real-world objects may significantly influence the interactions [159]. However, we want to highlight that solutions proposed in other VR displays and AR and MR scenarios (e.g., CAVE, AR glasses, volumetric displays) may also be adaptable for fully-immersed VR headsets [83, 110, 131].

2.7 Summary

We have teased out eight research challenges in VR selection and manipulation through the literature review. In this thesis, we provide novel solutions for the research challenge of coping with complex 3D interaction scenarios. Specifically, we develop techniques for fully-occluded target selection (Chapter 4), incorporate gaze and on-body input to offer precise, versatile, and more comfortable interaction (Chapter 5 and 6), and leverage computational models to lower the friction in acquiring small and distant objects (Chapter 7). Meanwhile, our solutions should inspire future researchers to resolve other challenges. For example, our computational models in Chapter 7 consider optimizing the use of contextual information to provide the most appropriate suggestions to users. We also explore a new design space for integrating gaze into the existing workflow based on mid-air interaction in Chapter 5, and the techniques should offer helpful support in a collaborative VR environment.

This thesis focuses on the success measurements of effectiveness, efficiency, ergonomics, experience, and expressivity (the *5Es*). For example, we evaluate the techniques through performance measures such as task completion time and error rate. We assess workload and fatigue measures through hand movement distances and questionnaires, including NASA-TLX [58] and Borg CR10 [15]. We also deploy questionnaires, such as the single easement questionnaire [147] and the short version of the user experience questionnaire (UEQ-S) [150], and perform interviews to measure interaction experiences. We also apply the techniques to various application scenarios to demonstrate their expressivity. These measures cover both the objective and subjective perspectives of the interaction experience. At the same time, we consider robustness, behavior, and consistency measures when feasible and appropriate.

Chapter 3

METHODOLOGY

This chapter introduces the general methodology that we applied in the research of this thesis. Specifically, we designed and implemented our interaction techniques through design space formulation and VR software prototyping. To evaluate the techniques, we took a human-centered approach that considers the end user's needs, experiences, and perspectives by conducting multiple in-lab user studies. We describe the experiment designs and procedures we employed to ensure the internal and external validity of the study results. We also performed rigorous quantitative and qualitative analyses with the collected user data. Finally, we discuss the ethical considerations when conducting user studies.

3.1 Artifact Design and Prototyping

§1 Design Space Formulation. Before building a new interaction technique, we typically consult the design space, which refers to the structured set of possibilities in which people can interact with the technology in a given application scenario or under specific design constraints. For example, in fully-occluded target selection in Chapter 4, we considered different occlusion visualizations (e.g., multiple viewports and X-rays), the size and space of the visualizations, and different selection techniques. In gaze-supported object manipulation in Chapter 5, we investigated a collaborative design space of the two input modalities (eye and hand), such as how one modality can transit to the other. These design spaces (i.e., the categorization of the different design options) helped us to consider the design trade-offs and make informed decisions about which technique to choose in a given situation.

§2 VR Software Prototyping and Demonstration. After conceptualizing the designs, experiments, and demonstrations, we implemented them in Unity 3D with the C# programming language. We then deployed the software to VR platforms, such as the Oculus Quest.

§3 Simulations and Reinforcement Learning. In Article IV, we employed reinforcement learning to identify optimized solutions in computer simulations. Specifically, we built user behavior models that simulate how users behave under a given situation. We then employed reinforcement learning agents to test different strategies by trial and error to maximize the reward (e.g., time saved for users) based on the simulated user behaviors.

3.2 User Studies

§1 Design. In Articles I, II, and IV, we conducted within-subject experiments to compare candidate solutions. In this case, each participant was tested in every experimental condition of techniques and environmental factors (e.g., target distance and densities). The environmental factors were tailored to the goal of the experiment and were carefully selected based on the literature and the pilot studies. Within-subject design helped control the effect of individual differences and increased the statistical power. We counterbalanced or randomized the testing conditions to mitigate the potential ordering effect caused by such designs. We also repeated each condition for multiple trials to improve the reliability (i.e., the resulting data accurately represented the "true" performance).

§2 Procedure. Our in-lab user studies were typically structured as follows. We first welcomed the participants to the experiments and requested them to fill in a pre-experiment questionnaire to collect their demographic information, including, for example, whether they had normal or correct-to-normal vision and their familiarity with VR equipment. We then introduced and helped them to wear the VR device and let them get acquainted with the system in a sample virtual space. We then required them to practice the interaction techniques and proceed to the formal experiment. After the experiment, we asked them to complete another questionnaire on their experience and, in some cases, also conducted an interview. We ensured participants rested enough during the study to prevent user fatigue or disengagement [199].

§3 Interview. We performed semi-structured interviews in Articles I-III to gather user feedback on the proposed solutions. We prepared a list of questions to ask before the experiment and probed into more details based on participants' answers. These interviews helped to understand the users' impressions and concerns, which were employed to iterate on our designs.

3.3 Data Analysis

We conducted quantitative analyses to determine whether our proposed solutions have significantly improved the intended objectives (e.g., completion time). We also administered qualitative analyses regarding users' feedback to advance our designs.

3.3.1 Quantitative Analysis

We performed quantitative analyses on the experimental data in Articles I, II, and IV.

§1 Metrics. The quantitative evaluation metrics centered around the *5E* measurements of effectiveness, efficiency, ergonomics, and experience. We included selection time, manipulation time, error rate, and more detailed metrics of coarse translation time and re-position time for measuring effectiveness and efficiency. We also used hand movement distance, hand rotation angles, Borg CR10 [15], and NASA-TLX scales [58] for estimating user workload and fatigue.

Moreover, We quantified user experience through preference rankings, function usage percentage, and UEQ measures of pragmatic, hedonic, and overall quality [150]. These measurements encompass both the objective and subjective aspects of the interaction experience.

§2 Outlier Removal. If time performance data were collected in the experiment, We typically considered the time above or below three standard deviations from the mean ($mean \pm 3std.$) in each condition as outliers and discarded them in the analysis. These outliers could be induced by the confusion or mind-wandering of the participants.

§3 Tests of Significance. We conducted statistical significance tests with repeated-measures ANOVA (RM-ANOVA) in Articles I and II and linear mixed models (LMM) in Article IV. Both RM-ANOVA and LMM were used to identify a significant effect of a factor on a dependent variable, such as selection time (i.e., whether the chance that a factor has an impact on the dependent measures is below a threshold). RM-ANOVA was applied when the within-subject factors were fixed (i.e., the same levels were applied to each subject).

Before conducting the significance tests, we validated the normality hypothesis through Kolmogorov-Smirnov tests, Shapiro-Wilk tests, and visual inspections. If the data appeared non-normally distributed, we performed transformations such as aligned rank transformation (ART) [187] to normalize the data. We also adjusted the degrees of freedom with Greenhouse-Geisser correction, if appropriate. Additionally, we performed Bonferroni-adjusted pairwise comparisons to identify whether the factor levels were significantly different from each other. These analyses were common practices in the field.

§4 Complementary Tests. In addition to the significance tests, we employed effect size measures such as the non-parametric estimator of CL [171] to complement the pairwise comparison results. Effect size helped us understand the magnitude of differences between the conditions, which could not be achieved by statistical significance tests alone [164].

3.3.2 Qualitative Analysis

In Articles I-III, we conducted qualitative analyses on the interview data gathered from the participants. We transcribed the interview data and coded the issues raised by the users in a thematic manner that concerns the frequency and importance of the issue being mentioned. In other words, we emphasized those issues that appeared more frequently or were recognized as fundamentally important [2, 37]. We then inferred possible casual conditions of the occurrence of the issue and elaborated on our findings on papers. These qualitative analyses help us focus on key user experience issues of the invented techniques, which can improve understanding of interaction behaviors and guide future research.

3.4 Ethical Considerations

We implemented several measures to guarantee the ethical conduct of the research studies in this thesis. The experiment protocol was approved by the University of Melbourne Human Ethics Advisory Group (Article I and III, ethics ID: 1955876), the University Ethics Committee in Xi'an Jiaotong-Liverpool University (Article II), or the Western Institutional Review Board (Article IV).

§1 Informed Consent. We informed the participants of study-related information via a written plain language statement. Before they agreed to participate in the study, they were given sufficient time to read a consent form. Their consent was obtained by having them sign and return the consent form. The experiment and data collection only started once the participant signed the consent form. Additionally, we ensured that the participants were aware that their participation would not impact their grades if they were students at the university.

§2 Risks, Compensations, and Monitoring. We were aware of the potential risks in a VR experiment, such as causing motion sickness and eyestrain. Participants were informed of the possible adverse effects. They were allowed to take off the VR headsets if feeling uncomfortable. Further, they could withdraw from the study at any point of the experiment without explanation or prejudice. The researcher also carefully monitored the whole experiment to spot any potential risks. After completing the study, the participants were compensated with rewards such as gift vouchers and snacks, as written and agreed upon in the consent form.

§3 Data Management. We collected only the necessary data for our research, such as basic demographic information, task completion time, and oral feedback. To maintain privacy, we anonymized the collected user data so that it is impossible to link them with identifiable personal information (e.g., names). In any publication resulting from the research, the participants were identified by pseudonyms such as "P1". For safeguarding the data, we utilize secure servers protected by firewalls. Access to the server was controlled through protected password authentication mechanisms.

Chapter 4

FULLY-OCCLUDED TARGET SELECTION

4.1 Summary

In this work, we propose interaction techniques for fully-occluded target selection in VR. The presence of fully-occluded targets is common within virtual environments, ranging from a virtual object behind a wall to a data point of interest hidden in a complex visualization. However, current mid-air interactions based on Virtual Hand and Raycasting have limited functionalities in selecting such targets without repetitively moving from one place to another (locomotion) to discover the occluded target. Therefore, we developed ten techniques and conducted two user studies to explore appropriate visualizations and interactions to offer selections for fully-occluded targets.

The proposed fully-occluded target selection techniques can deal with small, distant, and partially- or fully-occluded targets. The selected techniques were optimized for effectiveness, efficiency, and user experience. They were also demonstrated to remain helpful in simple and more complex environmental settings (e.g., different occlusion layers, target depths, and object densities) and various application scenarios (e.g., 3D modeling and data exploration).

Env.			**Task**				
Small	*Distant*	*Occluded*	*Effectiveness*	*Efficiency*	*Ergonomics*	*Experience*	*Expressivity*
✓	✓	✓	✓	✓		✓	✓

4.2 Article I

Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. "Fully-Occluded Target Selection in Virtual Reality." *IEEE transactions on visualization and computer graphics* 26, no. 12 (2020): 3402-3413. https://doi.org/10.1109/TVCG.2020.3023606

Fully-Occluded Target Selection in Virtual Reality

Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, Jorge Goncalves

Abstract—The presence of fully-occluded targets is common within virtual environments, ranging from a virtual object located behind a wall to a datapoint of interest hidden in a complex visualization. However, efficient input techniques for locating and selecting these targets are mostly underexplored in virtual reality (VR) systems. In this paper, we developed an initial set of seven techniques techniques for fully-occluded target selection in VR. We then evaluated their performance in a user study and derived a set of design implications for simple and more complex tasks from our results. Based on these insights, we refined the most promising techniques and conducted a second, more comprehensive user study. Our results show how factors, such as occlusion layers, target depths, object densities, and the estimation of target locations, can affect technique performance. Our findings from both studies and distilled recommendations can inform the design of future VR systems that offer selections for fully-occluded targets.

Index Terms—Pointing selection, object selection, visualization, occlusion, virtual reality, hidden target, head-mounted displays

1 Introduction

Virtual reality (VR) enables users to achieve what may not be possible in the physical world. Though many user interfaces have been developed for simulating or adapting real-world features (such as providing realistic tactile feedback [5]), it has long been argued that the real power of VR lies in creating a "better" reality by utilizing "magical" techniques that while being unrealistic, provide a better user experience [1, 46, 63, 65, 67]. One primary advantage of such interaction techniques is to overcome human limitations in terms of cognitive, perceptual, physical, and motor capabilities [46]. For example, existing techniques enable the user to interact with distant objects [78] and teleport around virtual environments [37], which are impossible in the physical world. This research focuses on one such interaction—selecting fully-occluded targets in VR.

The challenge of interacting with fully-occluded targets is prevalent within virtual environments. Structural elements, like walls, can easily hide and prevent users from accessing the objects behind them [23, 47, 79] (see Figure 1). In another example, high-dimensional data visualizations are also likely to obscure a datapoint of interest from being acquired by analysts [6, 20, 54, 76]. Further, when building 3D models in virtual environments, it might be cumbersome to select and thus manipulate hidden components, such as an engine hidden inside a virtual model of a motor vehicle [4].

However, existing selection techniques in VR are limited in their effectiveness for selecting fully-occluded targets. Based on the available literature on the topic, we argue that the main challenges are **(1)** the deficiency of the formulation of the problem in VR and general strategies to solve it; **(2)** the lack of effort in combining 3D occlusion management techniques to facilitate the discovery phase of the selection process [3]; and **(3)** the absence of a thorough evaluation and comparison of techniques that manipulate the key factors related to fully-occluded target selection. We aim to fill these gaps in this paper.

We first formulate the fully-occluded target selection problem and frame an approach to address it. We then derive a design space, which inspired seven potential techniques for selecting fully-occluded targets in VR. We present a user study that compares these techniques based on both simple and complex tasks. Based on the study results, we refined the more promising techniques and introduced a second, more in-depth study aimed at assessing technique performance under different environmental factors including occlusion layers, target depths, object densities, and the estimation of the target locations. Following, we discuss the findings from both studies and suggest recommendations to inform the design of future VR systems that offer selections for fully-occluded targets.

• *The authors are with School of Computing and Information Systems, The University of Melbourne. E-mails: {difengy, qiushiz2}@student.unimelb.edu.au; {joshua.newn, tilman.dingler, eduardo.velloso, jorge.goncalves}@unimelb.edu.au.*

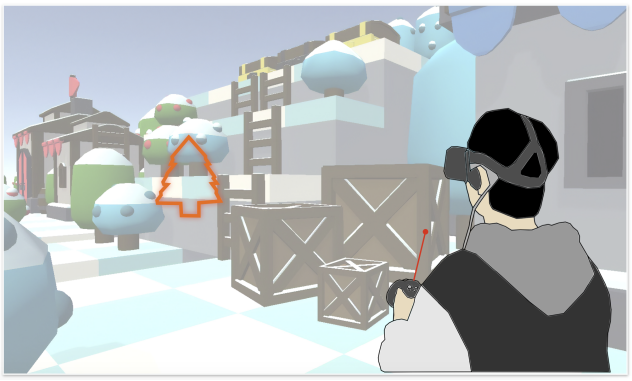


Fig. 1. Example Scenario: A user is constructing an environment in VR and intends to select and manipulate a hidden tree (outlined in orange) that is fully-occluded from the view of the user.

2 Problem Formulation and General Strategy

Considering previous work regarding the occlusion problem in 3D environments [31], we formulate the problem space and propose a general problem-solving strategy for the selection of fully-occluded targets in head-mounted display (HMD)-based VR systems.

Within a 3D virtual environment, there are *selectable*, and *unselectable* objects. Users can pick up or interact with the selectable objects, but not the objects that are unselectable since they serve other purposes within the virtual environment, such as decoration to enhance the realism of the scene. Among the selectable objects, there is commonly one primary *target* that the user intends to interact with, while all the selectable objects act as *distractors*. The target can switch when the user's intention changes. A target is defined to be fully-occluded from a viewpoint if it can not be seen from any viewing direction of the user. Different objects within a virtual environment can become fully-occluded at some point during the interaction.

To select a fully-occluded target, the user needs first to form an intention. With that intention, although the user cannot directly see the target at this stage, they typically have an awareness of the areas where it might occur—we call them *occurrence areas*. The estimated size of the occurrence area depends on the user's confidence. If the user has no idea of where the target might locate, the occurrence area

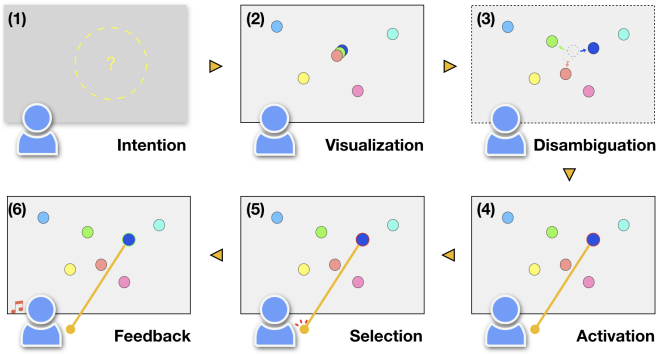


Fig. 2. To select fully-occluded targets, a user 1) forms an intention. The user is then provided with 2) visualization of the objects, 3) an optional disambiguation technique to spread the clustered objects, 4) an indicator of the current pointing object (e.g. highlighting), 5) a selection trigger to confirm the selection, and 6) feedback (e.g., visual, sound, haptic) after selection.

is the whole visualization. After that, the user can use a supporting technique to "locate" the fully-occluded target and then pinpoint the target to perform the selection.

In line with the problem space discussed above, we propose a general strategy to tackle the fully-occluded target selection problem. Once a user has a selection intention, a visualization of the target needs to be displayed to make it "visible" to the user from their viewport. However, as the technique will not know which object the user is aiming at, a group of potential objects, possibly within the user's estimated occurrence areas, will be presented. Next, the technique helps the user to disambiguate the list of selectable objects and provides activation feedback when an object is being pointed at. Finally, the technique should allow the user to select an object by pressing a trigger and receive confirmation feedback. The general strategy described above for selecting fully-occluded targets is summarized in Figure 2.

3 Related Work

In this section, we first introduce previous work regarding occluded target selection in VR. We then present the techniques related to visualization and selection, which are the two main steps in our general problem-solving strategy.

3.1 Occluded Target Selection in VR

Previous research in VR has explored object selection under cluttered and occlusion conditions, mostly for partially-occluded targets. One successful technique is Depth Ray [38], which attaches a movable marker onto the selection ray. The object intersected by a ray and closest to the marker can be selected. In another study, Depth Ray was shown to be effective for selecting a target that was completely occluded by making the objects adjacent to the ray semi-transparent [72]. However, the target was first visible to the user for 2 seconds before it got fully-occluded by distractors, which is unlikely to be the case in real applications. Furthermore, only one layer of occlusion was introduced, and the target shape was different from the distractors, which made the task much simpler. In real scenarios, multiple layers of occlusion can be presented, and the target can also be similar to the distractors [54]. Future work has refined [11], and applied similar techniques for collaborative work [2] and object manipulation [57].

Some techniques try to separate objects in a cluster by translating them into new positions. For example, Flower Ray uses a two-step approach: the user first point at an object cluster through a virtual ray, and then press to trigger to separate them into a marking menu [38, 44]. An updated version of Flower Ray uses a fixed-size cone to replace the ray to avoid missing small targets [26]. Both techniques have not been tested under dense conditions, where objects could still be partially- or fully-occluded even if they are translated to new positions.

Progressive refinement techniques that take the advantages of rearranging the objects in a more organized way, commonly into a new view, can also be suitable for selecting fully-occluded targets. SQUAD [7,42] allows users to cast a sphere onto multiple objects and interactively narrow down their selections using a quad-menu. Another technique called Expand [9, 18, 19] (not in VR HMD) enables users to zoom into the target area and reorganize the objects onto a grid for a second phase selection. Expand was shown to perform faster than SQUAD in dense environments. Later works extend such techniques by using a mobile touchscreen as input [25] and arranging objects in different layouts (circular layout rather than a grid) [49, 56]. However, none of them have been formulated under the context of fully-occluded target selection, nor have they been thoroughly compared to other techniques presented in this section. Nevertheless, we drew inspiration from these techniques when developing our techniques for fully-occluded target selection in VR.

There are some other techniques that are promising for selecting fully-occluded targets: flexible pointer [51] uses a curved ray which could bypass the distractors, iSith [77] determines the target by using the interaction point of two rays, VirtualGrasp [78] retrieves an object by simulating the gesture as if grasping the target object, X-Ray Vision [40] reveals hidden content by looking at a "scaffolding pattern", and Outline Pursuits [64] selects an occluded target by matching its outline with smooth pursuit eye movement [73]. While these techniques provide interesting concepts, substantial tweaks would be needed for them to be suitable for general fully-occluded target selection scenarios. For example, VirtualGrasp [78] can not deal with objects with an identical shape.

We summarise the following three gaps in the literature:

- The fully-occluded target selection problem has not been established in VR. Previous work normally assumed that the target location was known or only partially hid the target. In addition, important factors, such as layers of occlusions, were not identified.
- Limited work has tried to combine occlusion visualizations to support the discovery phase of the targets, as they mainly focus on the selection phase. However, as fully-occluded targets can cause some uncertainties with their locations, visualizations that help with the search phase are essential.
- A thorough evaluation and comparison of different types of techniques that could be potentially used for fully-occluded target selection are missing.

3.2 3D Occlusion Visualization

Elmqvist and Tsigas reviewed fifty 3D occlusion management techniques for visualizations [31] and extracted five design patterns from these techniques. Next, we highlight important work in the three patterns that are more relevant to our research.

Multiple Viewports. The multiple viewports pattern is characterized by embedding alternate (often separate) viewports/windows to the main view. Examples include World In Miniature (WIM) [67, 70], which generates a small, handheld copy of the entire world, and Worldlets [33], which inserts multi-perspective viewpoints of an environment into the main view [13, 58, 76]. Recent work presents 3DMini-map [79], which helps to convey distance and direction information of off-screen and occluded targets. However, selecting objects directly on these visualizations is still underexplored.

Virtual X-Ray. The virtual X-ray pattern makes objects visible by turning occlusion surfaces invisible or semi-transparent. Making front objects transparent can benefit the discovery of objects that hide behind [27,30,41,68,82]. However, it is known to suffer from the "Superman's X-ray vision" problem [48]—when there are too many occlusion layers, users are not able to make sense of the depth relationships of objects. Others have explored a cutaway view [16, 23, 28, 34], which eliminates or cuts holes over unwanted distractors.

Volumetric Probes. Volumetric probes normally use a probe to transform objects by removing or separating them. The above-mentioned disambiguation techniques, which reorganized potential targets on a new

view [74], could be counted as one substream. Other techniques have attempted to scale [22], translate [8, 15, 29, 32, 55], and distort [17, 24] objects in the scene in order to reveal the hidden objects. The transformation of the object needs to be carefully controlled so that the object is not occluded by new distractors, especially in a dense environment [32].

3.3 Selection Techniques in VR

There are two main categories of selection techniques in VR: virtual hand and virtual pointing [3]. Since a plethora of techniques have been proposed under those two categories, we direct interested readers to surveys on the topic [3, 46], and more recent works [11, 12, 71, 80]. RayCasting [11, 50] it is one of the most common techniques for 3D object selection in virtual environments. In RayCasting, a visible ray emanates from the tracked hand position to the direction of where the hand is pointing at, and the first object that is intersected by the ray can be selected [46]. Despite its usefulness, the performance of RayCasting deteriorates when selecting distant or small objects. Researchers have been actively seeking solutions to enhance its performance, especially in dense environments [36, 72]. Recent work has compared different visual feedforwards for RayCasting and suggests that highlighting the nearest target was the most efficient way in terms of selection performance [11]. Another approach is to try to minimize input noise with the use of algorithms and computational models [11, 80]. We utilized some of the techniques mentioned above to strengthen our fully-occluded target selection techniques. We illustrate this aspect in more detail in the description of the developed techniques.

4 Design Space

Our general strategy for selecting fully-occluded target suggests that the problem can be solved in five steps (visualization, disambiguation, activation, selection, and feedback). Here, we focus on the three main steps, which are visualization, disambiguation, and selection. We maintain the other two the same across all the techniques. The activation indication was provided by outlining the target, and the confirmation feedback was given by sound. Regarding the three focused steps, we have identified the following six primary considerations for designing fully-occluded target selection techniques.

Visualization Patterns: which type(s) of the visualization pattern, among the ones that are identified by Elmqvist et al. [31] (typically multiple viewports, virtual X-ray, and volumetric probe) is/are utilized to visualize the target?

Visualization Size: what is the size of the visualization area? Are we applying the visualization to only limited areas, or more extensive areas (even the whole environment)? For instance, to visualize the objects, we can make a small area transparent, however, we can also tweak the whole scene to do so.

Visualization Versatility: will users be able to specify which area(s) they want to apply the visualization? How precise can it be (in an arbitrary shape or a constrained region)? In real use cases, the users will have different estimations of where the target might occur, thus it is important to define their belief/guess accurately.

Disambiguation Invariances: when applying the disambiguation technique, which property (or properties) of the original objects will be maintained? These properties may include object position, size, relation, and appearance. For example, if we are asked to select a datapoint among other datapoints that have the same appearance, re-arranging all of them into new positions might not be ideal.

Selection Techniques: what type of selection techniques will be applied? Are we embedding selection enhancement techniques or filter out the noisy input? These decisions are likely to be highly related to selection performance. In the initial exploration, we mainly focus on the selection techniques that are based on pointing (Raycasting) and virtual hands without adding selection enhancements.

Input Modality: which input modality (modalities) are used for selecting the fully-occluded target? While many types of input modality exist (voice, gaze, gesture, etc.), we focus on controller input. A survey of currently available controllers on the market showed that most controllers were equipped with at least a touchpad or a joystick (2 degrees-of-freedom input, 2DOF), a trigger (1DOF input), and buttons (only on/off). The controller itself can be 3DOF (only rotation can be detected) or 6DOF (both rotation and translation can be recognized). Different techniques might need to employ different inputs. In our research, we used a joystick, a trigger, and a button of an Oculus Touch controller throughout the studies. The design space can be expanded in the future when investigating other input modalities to achieve the functionalities of each technique (e.g. hand-tracking).

5 Potential Techniques

Based on the design space, we developed the following nine potential techniques with several iterations and pilot tests. These techniques are summarized and visualized in Figure 3. The following technique descriptions adhere to the design space. For an explicit mapping between the design space and the techniques, please refer to our supplementary materials.

Alpha Cursor: this technique is inspired by previous work that attaches a movable cursor onto the selection ray [11, 38]. With *AlphaCursor*, users control the cursor to come closer or go deeper into the environment at a constant speed by pushing the joystick forward or backward (see Figure 3b). In contrast to previous work, if the distance between the cursor and the user is larger than the distance between an object to the user, the object becomes fully transparent. The transparency manipulation is applied to the whole environment, and all objects maintain their original position and size during the disambiguation phase. RayCasting, which uses the trigger for selection confirmation, then selects the desired object.

Flower Cone: in *FlowerCone* (see Figure 3c), users select objects in two phases. First, the user controls a cone to match the estimated area of where the target might occur. The size of the cone can be adjusted by tilting forward/backward the joystick. When pressing the trigger, the user enters the second selection phase, in which all objects within that cone are presented on a grid. The user can select the target directly on the grid with RayCasting, or, if the target is not on the grid, the user can press the button to go back and resize the cone again. This technique combines visualization and disambiguation by using the grid layout. The visualization size can be controlled through the size of the flat circular base of the cone. However, the grid layout changes the original location and size of the object.

Gravity Zone: as shown in Figure 3d, *GravityZone* translates all objects in the scene to come closer or further away in a constant speed by tilting the joystick forward or backward. If the distance between an object and the user is smaller than a threshold, the object will be fully transparent. It is similar to *AlphaCursor* in that both of them make the objects transparent based on their relative depth. However, in contrast to *AlphaCursor*, *GravityZone* manipulates all the objects in the scene rather than the cursor. The location and size of the objects are changed during the translation, but their relative position is not altered. Raycasting is used to make the selection.

Grid Wall: inspired by Expand [19], in this technique, when the user presses the controller button, all objects are arranged on a grid (see Figure 3e) with a constant scale factor. We did not use the zoom-in feature from Expand as it can make participants dizzy in VR. *GridWall* completely reorganizes all objects in the scene to a new location with a different size. The user can select the target on the grid with RayCasting. The original location information of the object is temporarily lost with the grid layout.

Lasso Grid: with *LassoGrid*, users draw a trace in any shape by long-pressing the trigger (see Figure 3f). All objects within the trace, are presented on a grid layout when releasing the trigger for the second stage of selection. If the trace is not closed, the program completes it automatically. RayCasting is used to select the target on the grid. Pressing the button allows the user to go back and draw the trace again.

Magic Ball: inspired by previous work [79] (which only explored visualization rather then selection), *MagicBall* removes unselectable distractors and creates a 3D mini-map of all the selectable objects inside a transparent sphere (see Figure 3g). The objects' size and the distance between each other are both scaled-down, but the relative size and location information are both maintained. The user can select directly on the semi-transparent object proxies by moving the tip of the virtual

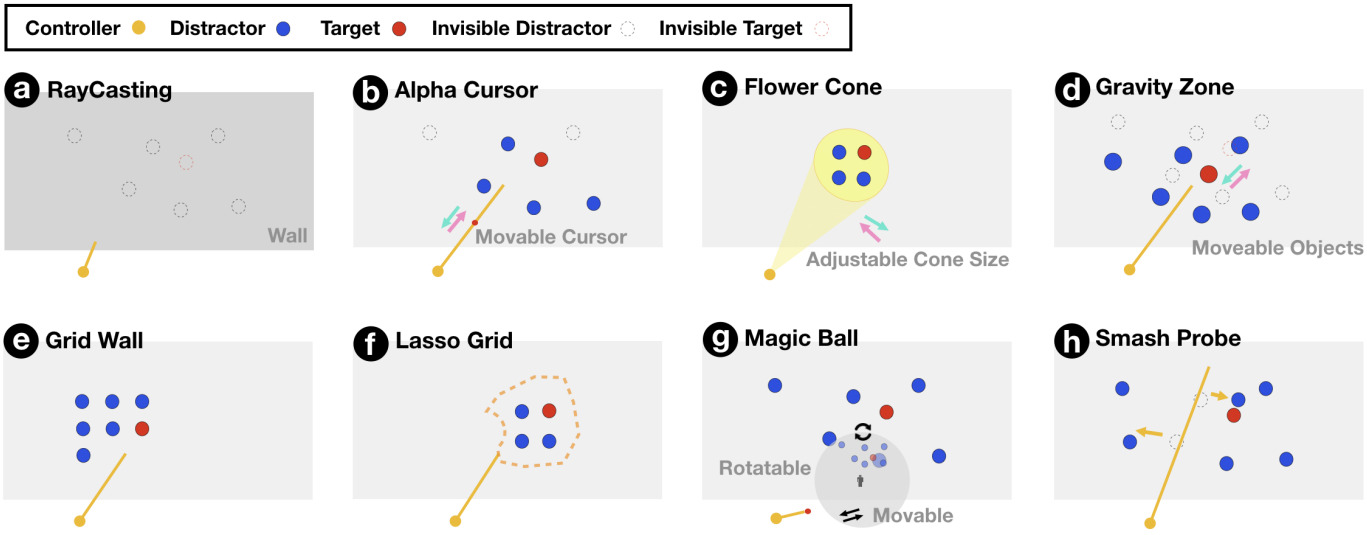


Fig. 3. The RayCasting technique (a) and techniques devised for fully-occluded target selection, including Alpha Cursor (b), Flower Cone (c), Gravity Zone (d), Grid Wall (e), Lasso Grid (f), Magic Ball (g), and Smash Probe (h).

stick onto the proxy and pressing the trigger. The user can also rotate and translate the mini-map by tilting the joystick.

Smash Probe: if more than one object intersects with the ray, *Smash-Probe* spreads these objects to a random direction within a fixed range (see Figure 3h). It only alters a small area per spread; however, multiple spreads can disarray the whole environment. The objects are translated back to their original position after a pre-defined time. RayCasting is used for selecting objects, and the user can disable or re-enable the spread function by pressing the button.

Depth Ray (discarded): previous work [38, 72] has described Depth Ray, which attaches a depth marker onto the selection ray. The objects that are close to the ray are rendered as semi-transparent so that the occluded targets could become visible. The one that is closest to the marker can be selected. However, during the pilot testing, we found that users were not able to distinguish the target and the distractors when multiple occlusion layers showed up using this technique, even with semi-transparent and border highlighting. Thus, we discarded this technique from the study.

Fly-Through (discarded): the technique allows users to fly through any objects and navigate freely across the virtual environment. However, following our pilot testing, it became clear that this technique was not efficient for this purpose and could cause motion sickness, and therefore, we also discarded this technique from the study.

All our techniques introduce a superimposed selection mode, which removes the unselectable objects in the scene for the simplicity of selectable target acquisition. While conducting user studies to optimize each technique was not feasible, and outside the scope of this research, we tuned all of their variables to the best of our ability during informal testing. The values of the variables are made available in our supplementary materials for replication purposes.

6 Evaluation Framework

6.1 Variables of Interest

We identified a set of factors that we hypothesized could impact the techniques' performance for fully-occluded target selection. As suggested by Fitts's law [35, 66], *target size* and *movement amplitude* are likely to have a significant effect on selection performance. However, rather than replicating the findings from the extensive previous work on the topic, we consider variables that are related to occlusion properties. These variables are:

- *Occurrence Area*. As discussed before, a user normally has an awareness of where a fully-occluded target might be located. The more uncertain the user is, the larger the occurrence area might be. Different sizes of the occurrence area is likely to influence the target searching time.
- *Occlusion Layer*. It specifies the number of selectable objects that can fully overlap the target from a user's point of view. It is more challenging for the user to find the target when more occlusion layers are present.
- *Environmental Density*. It is the number of selectable objects in the whole virtual space. Although some of them might not hinder the selection performance directly, it can cause distraction and are quite likely to appear in real application scenarios (unwanted objects are spread across the whole environment).
- *Target Depth*. It is the distance between the target and the user. A higher target depth value can make the target appear smaller to the user and raise more challenges for selection.
- *Density Space*. Density space [21, 38] offers more precise control of the object density within the target area. Similar to previous research, we place six distractors around the target (front, behind, up, down, left, and right). Density space is the distance between the six distractors surrounding the target.

6.2 Experimental Setup

To frame the experimental task for this research, we first consulted the past literature regarding target selection in 3D space. Existing tasks with perceivable patterns (e.g., [11, 66, 69, 81]), which users were required to select a set of fixed targets in a constant sequence, are not applicable in our case. This is because we wanted to vary the occurrence areas, which requires some randomness in the allocation of the target. Meanwhile, tasks based on interaction scenarios with the presence of some degree of unexpectedness, such as a game [19], might pose challenges to the control of variables. We decide to use more controlled tasks, which would still allow the randomization of target locations (such as [7, 38, 45, 53, 72]). However, as there is little work regarding fully-occluded target selection, we had to develop a new and reusable experimental task. Based on previous research, we designed the task as follows.

In the task, the user aimed to select a fully-occluded target among a set of distractors in a virtual environment. The target and the distractors had different colors, and the colors were generated from a pre-prepared list (we used seven colors in our case which were chosen to be easily distinguishable, see Figure 4). The task was divided into two phases:

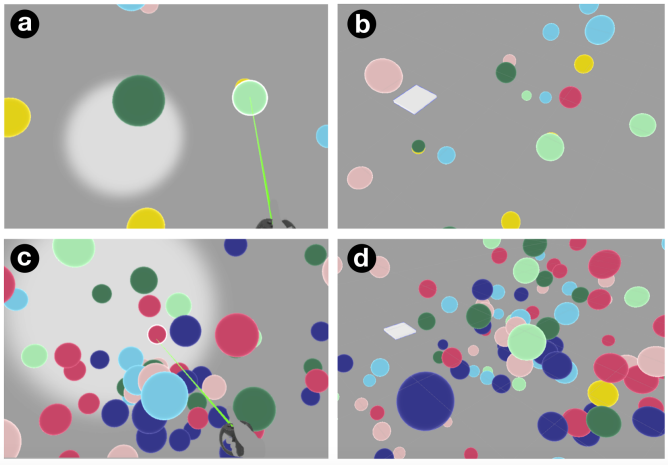


Fig. 4. Demonstrations of the experimental testbed including sparse environment first-person view (a) and third-person view (the square indicates where the user should stand) (b) and dense environment first-person view (c) and third-person view (d).

preparation and formal trial. During the preparation stage, the user started by pointing at a fixed *home object*, which had an identical color as the goal target. The home object ensured that the ray started from the same direction in the formal trial, and the user could press the selection trigger to proceed to the formal trial. During the formal trial, the objects, including both a target and distractors, and an indicator of where the target might be located (the occurrence area, marked as white in Figure 4a, c) were generated. The occurrence area could show up in any direction related to the home object, but the distance between them was always the same in our setting. The user then needed to use the corresponding input technique to select the target. We envision that, by modifying the variables of interest mentioned above, this task can capture a broad range of interaction scenarios in real cases.

7 Study 1 – Initial Exploration

We conducted an initial exploration and evaluation of the seven potential techniques for selecting fully-occluded targets in virtual environments. We aimed to extract design features that perform well in different interaction scenarios and determine potential aspects of our techniques that might need refinement.

7.1 Participants, Apparatus, and Materials

We recruited 21 participants (13F/8M), aged between 19-39 (M = 24.5 $\pm$ 4.3) with a diverse set of educational backgrounds (economics, arts, law, engineering, etc.) from a local university campus. All participants had normal or corrected-to-normal vision and rated their familiarity with VR as moderate (average 3.0 $\pm$ 1.6 out of a 7-point scale). Participants wore an Oculus Rift CV headset and interacted with our application through an Oculus Touch wireless controller.

7.2 Design and Procedure

The study employed a within-subjects design where we compared the performance of the seven developed techniques: (*AlphaCursor*, *FlowerCone*, *GravityZone*, *GridWall*, *LassoGrid*, *MagicBall*, and *SmashProbe*). The techniques were tested on two levels of task complexity (low and high). The higher complexity task had much larger occurrence areas, more occlusion layers before the target, higher environmental density, higher target depths, and larger density space than the lower complexity one. We ensured that there were considerable differences between the two levels of complexities—the detailed parameters are provided in the supplementary material. The order of the techniques was counterbalanced using a Latin Square approach, and the order of complexities was randomized. Following recommendations from previous work on target selection performance [53], we used two subsequent tasks (***search*** and ***repeat***). The search task required users to search for one target in a new scene and then select it, while the repeat task asked users to select the same target in the exact same scene.

We collected both performance data and subjective feedback from participants. Both selection time (the elapsed time between when the objects appear and when the selection is made) and error rate (the percentage of error trials for each condition) were recorded. We also measured the easiness of the techniques with the Single Easement Questionnaire [60] and the intrusiveness caused by them [52] on a 7-point scale. In addition, we asked participants to provide their preference ranking after finishing each technique and optionally also provide free-form feedback. We monitored the experiment from a computer, which showed the user's current view in VR, to observe the use of the techniques.

The whole procedure lasted around 40 minutes for each participant. At the beginning of the study, participants were briefed about the purpose of the research and signed a consent form. They also completed a pre-experiment demographic questionnaire. After that, they were introduced to the VR device and the experimental task, where we required them to finish as fast and as accurately as possible. They then wore the VR headset and familiarised themselves with the virtual environment. Next, they proceeded to the formal experiment within a fixed physical area. The experiment was divided into seven parts (corresponding to the evaluation of seven techniques). In each part, there were three phases: practice, perform formal trials, and answer questions. In the practice phase, participants were taught about how to use the technique, and they could practice it as long as they wanted until they got familiar with it. They then completed a series of formal trials. Finally, they were asked to complete the questionnaires mentioned above. Participants were allowed to rest between each condition. They were compensated with a $10 voucher at the end of the study.

7.3 Performance Results

In total, we collected 4704 data points (21 participants $\times$ 7 techniques $\times$ 2 complexities $\times$ 2 tasks $\times$ 8 repetitions) from the experiment. To analyze selection time, we discarded trials in which participants made a wrong selection (374 error trials, 8.0%), and removed outliers, in which the selection time was above three standard deviations from the mean ($mean + 3std.$) in each condition (92 trials, 2.0%). Such outliers are typically removed as they are likely to not represent the typical selection performance (e.g., small distraction during the experiment), and can skew results in a particular condition [72, 80]. The data regarding selection time were shown to be normally distributed (evidence from Kolmogorov-Smirnov tests and visual inspections), while the error rate data were not normally distributed and underwent pre-processing through Aligned Rank Transform (ART) [11, 75]. Next, we performed a repeated-measures ANOVA (RM-ANOVA) and Bonferroni-adjusted pairwise comparisons in each experiment scenario to analyze the selection time and error rate in each experimental condition[1]. The degrees of freedom produced by RM-ANOVA regarding selection time was adjusted using Greenhouse-Geisser correction. Both results are summarized in Figure 5.

7.3.1 Search Task - Low Complexity

Technique was shown to exhibit a significant main effect on selection time, with a large effect size ($F_{2.893,57.869} = 22.516, p < .001, \eta_p^2 = 0.530$) in low-complexity search task. *GravityZone* was the fastest, being significantly faster than most techniques ($p = 0.036$ for *AlphaCursor* and $p < .001$ for others) except *GridWall* ($p = .219$).

There was a statistically significant difference between Techniques regarding error rates ($F_{6,120} = 3.710, p = .002$). Post-hoc analysis indicated that *FlowerCone* had a significantly higher error rate than *AlphaCursor* ($p = .003$) and *GravityZone* ($p = .002$).

7.3.2 Search Task - High Complexity

Technique had a significant main effect on selection time, with a large effect size ($F_{3.221,64.425} = 13.276, p < .001, \eta_p^2 = 0.399$). *GridWall*

[1]For readability, we here report statistics in APA style (6th Edition). For the exact p-value when $p < .001$, please refer to the supplementary material.

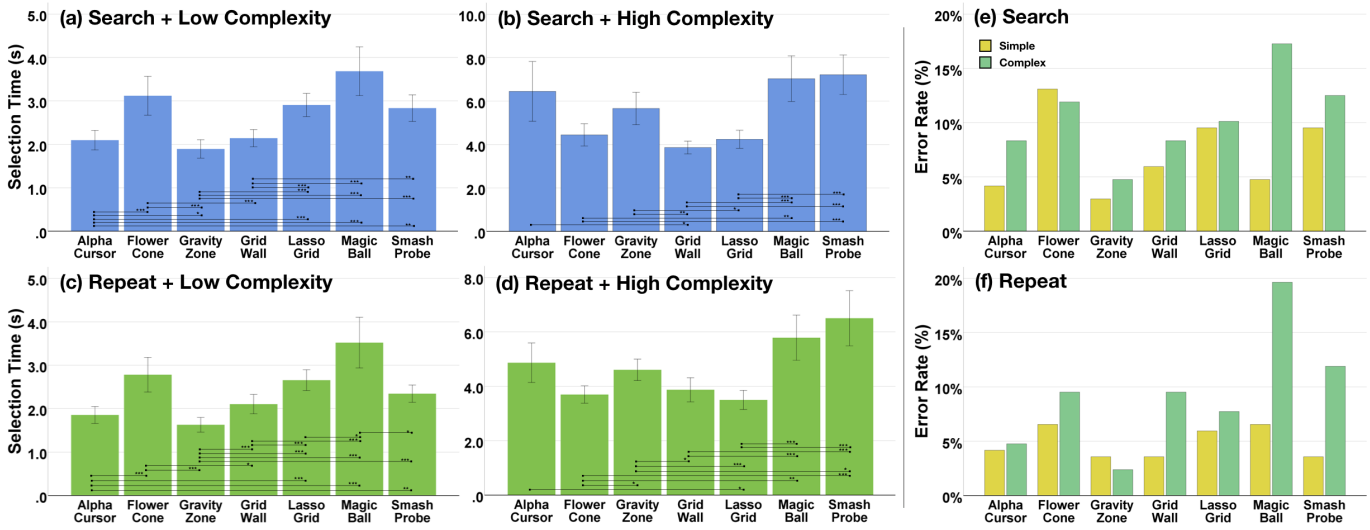


Fig. 5. Plots of selection time for the seven potential techniques regarding the search task with low complexity (a) and high complexity (b) and the repeat task with low complexity (c) and high complexity (d). Error bars indicate the 95% confidence interval. Statistical significant effects are marked (* = p $<$ .05, ** = p $<$ .01, and *** = p $<$ .001). Plots of error rate for the techniques regarding the search task (e) and the repeat task (f).

was the fastest technique for the complex task. It was significantly faster than *AlphaCursor* ($p = .017$), *GravityZone* ($p = .002$), *MagicBall* ($p < .001$), and *SmashProbe* ($p < .001$). There was no statistically significant difference when compared to *FlowerCone* ($p = .362$) and *LassoGrid* ($p = 1.000$).

There was a statistically significant difference between TECHNIQUES regarding error rates ($F_{6,120} = 3.204, p = .006$). *GravityWall* was shown to have significant lower error rate than *MagicBall* ($p = .001$).

7.3.3 Repeat Task - Low Complexity

TECHNIQUE was found to have a statistically significant effect on selection time, with a large effect size ($F_{2.379,47.581} = 26.061, p < .001, \eta_p^2 = 0.566$). *GravityZone* took the least time for selection when compared to most other techniques ($p < .001$), except *AlphaCursor* ($p = .078$).

There was no statistically significant difference between TECHNIQUES regarding error rates ($F_{6,120} = 0.945, p = .466$).

7.3.4 Repeat Task - High Complexity

TECHNIQUE was found to have a statistically significant effect on selection time, with a large effect size ($F_{3.282,65.646} = 18.412, p < .001, \eta_p^2 = 0.479$). *LassoGrid* was the fastest, but was similar to *FlowerCone* and *GridWall* ($p = 1.000$). *LassoGrid* was significantly faster than *AlphaCursor* ($p = .031$) and the remaining techniques ($p < .001$).

There was a statistically significant difference between TECHNIQUES regarding error rates ($F_{6,120} = 6.491, p < .001$). *AlphaCursor* ($p = .018$) and *GravityZone* ($p < .001$) had much lower error rates than *MagicBall*. *GravityZone* led to less error than *SmashProbe* ($p = .029$).

7.3.5 Search Task vs. Repeat Task

In terms of selection time, most techniques (all $p < .003$) had significant improvements in the repeat phase of the low complexity condition except *GridWall* ($p = .405$) and *MagicBall* ($p = .057$). For high complexity condition, there was no statistically significant effect of task on selection time for *GridWall* ($p = .970$) and *SmashProbe* ($p = .143$), but there was for all the others (*MagicBall*: $p = .031$ and others: $p < .001$).

Regarding error rates, only *MagicBall* ($p = .035$) and *SmashProbe* ($p = .043$) improved in the low complexity condition. No significant difference was revealed in the high complexity condition (all $p > .050$).

7.4 User Feedback Results

The overall easiness and intrusiveness of the techniques were calculated by averaging the 7-point Likert scale results. We also computed the mean ranking and counted the number of first/second place for each technique. The results from both questionnaires are summarized in Table 1.

In terms of the free-form feedback, the comments were mostly focused on *GridWall*, *MagicBall*, and *SmashProbe*. Several participants (N=7) felt *GridWall* was somewhat "boring" because it simply arranged all the objects in a 2D grid. In contrast, *SmashProbe* was seen as "fun" to use (N=4). Some participants thought *MagicBall* provided a good overview of the objects (N=4) but was quite difficult for selecting the target when the object number was high (N=3).

Table 1. The mean value (standard error) of easiness rating, intrusiveness rating, and preference ranking for all the techniques in Study 1. The last column shows the number of times a technique is ranked as the first/second. For Easy, higher is better; for Intrusiveness and Rank, lower is better.

Technique	Easy	Intrusiveness	Rank	#1/2
AlphaCursor	5.38 (0.33)	1.90 (0.34)	4.05 (0.41)	2/3
FlowerCone	5.81 (0.27)	1.95 (0.36)	3.38 (0.43)	4/3
GravityZone	5.86 (0.27)	1.76 (0.26)	3.05 (0.35)	5/3
GridWall	6.33 (0.16)	1.57 (0.36)	3.48 (0.42)	6/1
LassoGrid	5.76 (0.22)	1.86 (0.26)	3.86 (0.40)	1/7
MagicBall	4.33 (0.37)	3.19 (0.39)	5.19 (0.41)	0/3
SmashProbe	4.76 (0.34)	3.10 (0.28)	5.00 (0.47)	3/1

7.5 Summary and Discussion

The results show that performance improved for most techniques when participants moved from the search task to the repeat task. This is particularly true for the complex tasks, where selection time was significantly shortened in the repeat task. However, *GridWall* did not gain an advantage from the repeated selection, as the object order was randomized on the grid. *SmashProbe* did not improve significantly in the high complexity condition during the repeat phase. The selection phase of these techniques took a significantly longer time to complete when compared to the searching phase. As the repeat task was a replay of the previous task, the learning effect can also reduce the selection time and help users correct errors. Interestingly, the ranking of the techniques based on selection time almost did not change from the

search task to the repeat task. This is likely caused by the fact that the first selection only narrowed down the participants' estimation of the "occurrence area" of the target, while some searching was still needed in the subsequent selection.

For low complexity tasks, *GravityZone* and *AlphaCursor* performed better (both with shorter selection time and lower error rates). *GridWall* also yielded good performance, whereas other techniques were shown to take more time or have higher error rates. One possible cause for this is that for simpler tasks, *GravityZone*, *AlphaCursor*, and *GridWall* can reveal the target quickly with straightforward manipulations, while techniques like *FlowerCone* and *LassoGrid* required an extra layer of area specification. The performance of *SmashProbe* was comparable to *FlowerCone* and *LassoGrid*. In contrast, *MagicBall* was the slowest, mostly because it required some precision to select the small proxies of objects.

For high complexity tasks, techniques that arranged the objects on a grid were the most successful in terms of selection performance, with *GridWall*, *LassoGrid*, and *FlowerCone* clearly outperforming other techniques. For instance, performance when using *AlphaCursor* and *GravityZone* suffered when the task got complex. Searching the target became difficult for participants, as once missing the target, which was surrounded by the sea of distractors, the participant had to move back and forth (the cursor of *AlphaCursor* or the object clusters of *GravityZone*) to search for them. Navigating to the correct depth where the target located was cumbersome. Similarly, *SmashProbe* performed poorly, as the target can sometimes "jump" to places where it was still fully-occluded by others.

Furthermore, participants found that if they kept spreading all objects in such dense environments, it would lead to significant distraction. The complex scenario also further exacerbated the problems with *MagicBall*. This is because participants needed to have very high precision for selecting the duplicates of the objects, while pressing the trigger on the controller could easily cause hand tremors [14], which can lead to the wrong selection.

Regarding easiness and intrusiveness, all techniques were rated better than the middle point of the 7-point Likert scale. Participants rated *GridWall* the easiest technique, which also caused the least distraction. However, participants felt bored when using this technique as it no longer felt like 3D interaction. *GravityZone*, *LassoGrid*, *FlowerCone*, and *AlphaCursor* all got positive feedback in term of these two scales. On the other hand, *MagicBall* and *SmashProbe* were rated lower, given the difficulty of selecting targets with these techniques. *MagicBall* can cause a wrong selection due to the handshaking, and *SmashProbe* might lead to an unexpected spread of the objects when performing the selection. However, they were both seen as interesting by the participants. *MagicBall* built a nice overview of the objects, while *SmashProbe* created a level of unexpectedness, which could be fun for gaming purposes [59]. With regard to preference ranking, *GravityZone* was ranked highest, followed by techniques that employed the grid feature and *AlphaCursor*.

Based on the study results and our observations, we extracted a set of design lessons for different kinds of scenarios and application purposes regarding fully-occluded target selection.

L1. Use techniques with the grid feature (*GridWall*, *LassoGrid*, and *FlowerCone*) for dense environments. Our results showed that these techniques had much better performance in complex tasks. However, according to the design space, be aware that these techniques would not preserve the original scene (the original locations of objects).

L2. Depth-based techniques (*AlphaCursor* and *GravityZone*) provide simple solutions to lower complexity tasks. They can also preserve the location information of objects. However, when many distractors are clustered with the target, it might be difficult for these techniques to navigate to the exact depth where the target is located.

L3. A smaller-scaled duplicate of the whole environment (like *MagicBall*) can help provide location awareness in virtual environments [79]. However, requiring users to perform direct selection on the small object proxies can pose challenges, such as hand tremors [14].

L4. It can be beneficial to use techniques that have some sort of unpredictability for recreational purposes (like *SmashProbe*). However, in dense and complex environments, such unexpectedness can obstruct the primary selection task. In addition, applying 3D features in a virtual environment rather than only using 2D surfaces (e.g., *GridWall*) could lead to a more enjoyable experience.

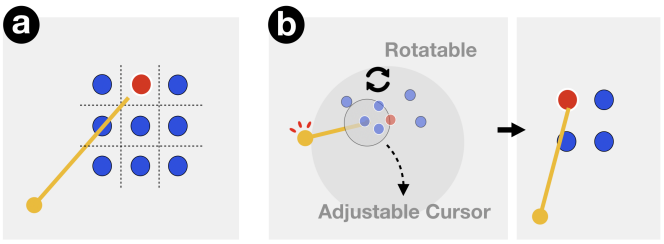


Fig. 6. (a) We implemented a selection enhancement technique on the grid layout, which would select the closest object to the ray; (b) *MagicBall+* embedded an adjustable cursor which could transform all object proxies into a grid layout for accurate selection.

After summarizing the findings from this first study, we were interested in refining the most promising techniques further. We also wanted to explore how specific environmental factors (like the size of the occurrence areas, occlusion layers, target depths, and object densities) would affect the performance of the techniques.

8 Technique Refinement

The seven techniques can be categorized into three sets: grid-based (*GridWall*, *LassoGrid*, and *FlowerCone*), depth-based (*GravityZone* and *AlphaCursor*), and others (*MagicBall* and *SmashProbe*). As the techniques in the different sets are better suited for different application purposes, and the ones within a set have similar strengths and weaknesses, we decided to improve them according to their general features. Based on the results and our observations from Study 1, we refined the techniques as follows.

We first improved the techniques that used grids (*GridWall*, *LassoGrid*, and *FlowerCone*). In the experiment, we found that the RayCasting technique for selecting objects on the grid sometimes led to errors during the fast-paced movements, as a correct selection was confirmed only when the ray was "crossing through" the target. Therefore, we decided to add selection enhancement techniques for RayCasting in the grid selection phase. We highlighted the nearest object to the ray and confirmed the selection on it when the user pressed the trigger, as this was shown to be the most efficient visual feedforward by recent work [11] (see Figure 6a). Additionally, to preserve depth information of the objects, we scaled the distances between the objects on the grid and the user according to the object's real distances to the user. This also adds some 3D features on the 2D grid surface. Although the visual size of the objects changed, our selection enhancement ensured that the effective size for selection was the same. Furthermore, we identified that randomizing the positions of the objects on the grid every time (for search and repeat tasks) was not efficient when there were a large number of objects. It could be more effective if objects were arranged by their distance to the ray (or center of the cone), from the closest to the farthest.

We also aimed to enhance the depth-based techniques (*GravityZone* and *AlphaCursor*).For some users, we observed that when the distractors in the front of the target were fairly close to it, navigating to the exact depth where the target located was laborious. Meanwhile, using a constant cursor speed might not be ideal for every user, as some might need it to be faster, while others want it to be slower. As a result, we made the cursor speed adjustable through the joystick input. The harder it was pushed/tilted, the faster the cursor became.

In addition, we found that most users had difficulties when selecting object proxies inside the mini-map, especially in the case where there was a large number of objects. To improve the selection, we combined the grid feature, which was shown to be effective for selection into *MagicBall*. Instead of employing the forward and backward movements of the joystick to translate the mini-map (which was not that useful

according to our inspection), it was used to scale up and down the transparent point cursor, which was used for selecting objects. Once the selection trigger was pressed, and the scaled-up cursor enclosed more than one object proxy, the objects that were inside the cursor would be arranged onto a grid for the second phase of selection (see Figure 6b). Users could still select objects from the mini-map directly if only one object proxy collided the cursor.

We first picked two techniques, *LassoGrid* from the grid-based techniques and *GravityZone* from the depth-based techniques, according to the empirical performance and user feedback, and applied the refinements as mentioned above. We also improved *MagicBall*, as many users preferred the small overview of objects, and the main problem with the technique was the difficulty caused by selection. Consequently, we evaluated the three refined techniques (*LassoGrid+*, *GravityZone+*, and *MagicBall+*) in the second study.

9 Study 2 - In-Depth Evaluation

To have a more thorough understanding of how different environmental factors might affect the performance of the techniques, we conducted a second study based on three refined techniques (*GravityZone+*, *LassoGrid+*, and *MagicBall+*).

9.1 Environmental Factors

Initially, we were interested in five essential environmental factors (occurrence area, occlusion layer, environmental density, target depth, and density space) which can have a substantial impact on target selection with the different techniques. However, evaluating all of them might pose a high workload for participants.

As a result, we combined environmental density and density space to one single factor called area density, as both these factors are related to the number of distractors inside a space unit. Area density specified the density of the objects within the occurrence area, intending to maintain the same level of difficulty for techniques within the targeting area. We assumed that objects within the targeting area might raise more challenges than the ones that were spread around the whole space. In this case, the density of the objects within the whole environment (outside of the occurrence area) would be set as constant. We ended up with four environmental factors, which are OccurrenceArea, AreaDensity, OcclusionLayer, and TargetDepth.

9.2 Method

We recruited another set of 16 participants (9F/7M) between the ages of 20-32 (M = 24.6 $\pm$ 3.3) with different educational backgrounds from a local university campus. All participants had normal or corrected-to-normal vision. They rated their familiarity with VR as moderate (3.6 $\pm$ 1.7 on a 7-point scale). We used the same apparatus and devices as in the first study.

The study employed a within-subjects design with five factors: Technique (*GravityZone+*, *LassoGrid+*, and *MagicBall+*), OccurrenceArea (small and large), AreaDensity (low and high), OcclusionLayer (less and more), and TargetDepth (low and high). The details of the variables are summarized in the supplementary material. Three techniques appeared in a random sequence, while for each technique, we varied the four counterbalanced environmental factors. The techniques were well-distributed in terms of their order according to our post-hoc evaluation. In this study, only the search task was used, rather than including both search and repeat task, because 1) it decreased the workload of the participants, 2) we found techniques had similar rankings based on the selection time for both tasks, and 3) the search phase is likely to be more relevant to real application scenarios. In total we collected 3072 trials of data (16 participants $\times$ 3 techniques $\times$ 2 occurrence areas $\times$ 2 area densities $\times$ 2 occlusion layers $\times$ 2 target depths $\times$ 4 repetitions).

As with Study 1, we gathered selection time and error rate data, and observed each experiment. Additionally, we used two standardized questionnaires to assess the task workload and user experience. The workload was measured by RAW NASA-TLX [39], and the user experience was quantified by the short version of the User Experience Questionnaire (UEQ-S) [61]. These questionnaires are more comprehensive than the ones used in the first study, which were used for the simplicity of the experiment given the higher number of tested techniques. Both questionnaires were presented inside the virtual environment as previous work has shown that it can reduce study duration and user disorientation [62].

Table 2. The results from the short version of User Experience Questionnaires (UEQ-S) which outline the pragmatic quality, hedonic quality, and overall quality of each technique. In the table, ">avg." means "above-average", "exc." means "excellent".

Technique	Pragmatic	Hedonic	Overall
GravityZone+	1.31 (>avg.)	1.38 (>avg.)	1.34 (>avg.)
LassoGrid+	1.83 (exc.)	1.77 (good)	1.80 (exc.)
MagicBall+	1.56 (good)	2.05 (exc.)	1.80 (exc.)

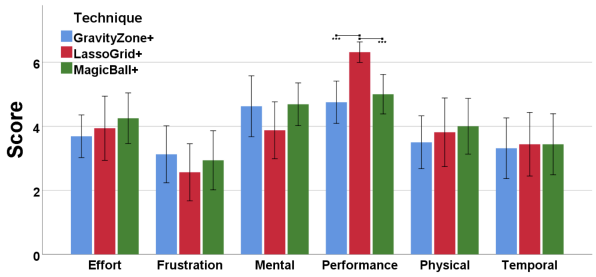


Fig. 7. The results from Raw NASA-TLX Questionnaires. Error bars indicate the 95% confidence interval. Statistical significant effects are marked (*** = p < .001).

The study lasted approximately 35 minutes for each participant. A similar procedure as the first study was used. Participants were compensated with a $10 voucher.

9.3 Results

As in the first study, we discarded the error trials (174 errors, 5.7%) and the outliers (78 trials, 2.5%) to analyze the selection time. We employed a RM-ANOVA with Greenhouse-Geisser correction for analyzing the effect of each factor. Pairwise comparisons with Bonferroni adjustment were used for technique comparison. Error rate data was transformed using ART [75] and was then analyzed through a RM-ANOVA. Regarding user feedback, we summarised the results from the questionnaires in Table 2 and Figure 7.

Since we were interested in how the techniques were affected by different environmental factors, we only present the effects and interactions related to the factor Technique.

9.3.1 Selection Time

A RM-ANOVA indicated that Technique ($F_{1.375,20.627} = 20.039, p < .001, \eta_p^2 = 0.572$) had a significant main effect on selection time, with a large effect size. A post-hoc test revealed that *LassoGrid+* was significantly faster than *GravityZone+* ($p < .001$) and *MagicBall+* ($p < .001$). *GravityZone+* was also indicated to be faster than *MagicBall+* ($p = .048$).

There were interaction effects between Technique $\times$ AreaDensity ($F_{1.356,20.336} = 6.090, p = .015, \eta_p^2 = 0.289$), Technique $\times$ OcclusionLayer ($F_{1.864,27.954} = 7.365, p = .003, \eta_p^2 = 0.329$), and Technique $\times$ TargetDepth ($F_{1.747,26.206} = 14.584, p < .001, \eta_p^2 = 0.493$), all with medium to large effect size. We present these interaction effects in Figure 8. No other interaction effects were found. Although there was no interaction between Technique and OccurrenceArea, OccurrenceArea itself did have a significant main effect on selection time ($F_{1,15} = 61.186, p < .001, \eta_p^2 = 0.803$).

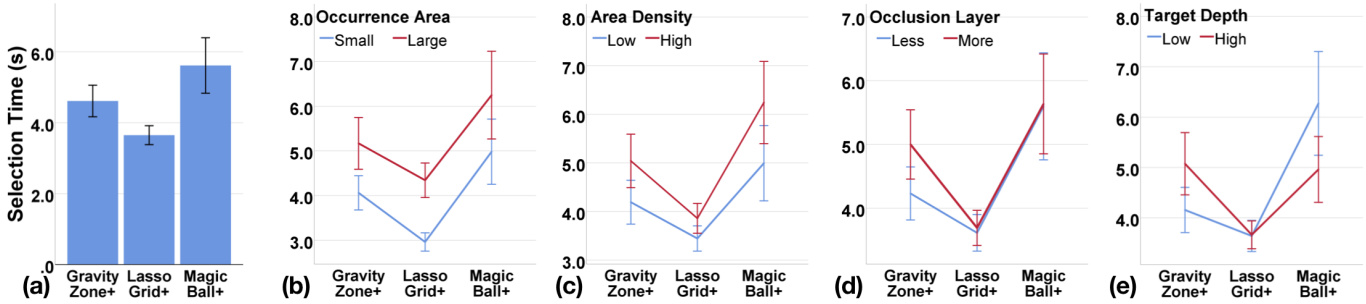


Fig. 8. Plots of selection time for the three improved techniques. These plots include techniques' overall selection time (a) and their selection time in different levels of Occurrence Area (b), Area Density (c), Occlusion Layer (d), and Target Depth (e). Error bars indicate the 95% confidence interval.

9.3.2 Error Rate

TECHNIQUE had a significant main effect on error rate ($F_{2,705} = 8.027, p < .001$). A post-hoc test showed that *GravityZone+* (5.96%, $p = .032$) and *LassoGrid+* (3.51%, $p < .001$) had statistically significantly lower error rates than *MagicBall+* (7.52%), with no significant difference between the two ($p = .490$).

A RM-ANOVA also revealed further interactions among the other factors. However, as the error rate was relatively low for all techniques (according to [11, 66]) and better techniques clearly outperformed the worse ones in terms of performance (lower error rate techniques as well as lower selection time), we do not examine these results here in further detail. For detailed statistics, please refer to our supplementary materials.

9.4 Summary and Discussion

In this section, we first examine the influence of the four environmental factors on the three techniques and then compare the input techniques from different perspectives.

Occurrence area had a similar impact on all techniques—as it increased, the selection time of the three techniques also increased significantly. This was expected as an inaccurate prior estimate of where the target might be located (larger occurrence area) leads to a higher search time, prolonging the selection process.

Area density affected the selection performance of all techniques, but the magnitude of the effect was different, as indicated by the significant interaction effect. As the density increased, search and selection for *GravityZone+* and *MagicBall+* became much more difficult than for *LassoGrid+*. One potential reason for this difference is the fact that organizing the objects on a grid-like 2D layout demanded less effort for searching rather than the original clustered and overlapped 3D arrangements [19].

Occlusion layer only influenced the performance of *GravityZone+*. With *GravityZone+*, it can feel somewhat cumbersome navigating through multi-layers of distractors. However, when using *LassoGrid+*, which arranged objects in the target area on a grid, and *MagicBall+*, which provided a quick overview of all the objects, users were not impeded by these layers at all.

Target depth affected the performance of *GravityZone+* and *MagicBall+*, but not of *LassoGrid+*. *LassoGrid+* was invariant to the change of target depth, as it transformed the 3D region to a grid, regardless of the real depth of the objects. Although we added depth information on *LassoGrid+*, it only changed the visual size of the objects, but not its effective size with the selection enhancement technique [35, 66]. For *GravityZone+*, when the target was located further, participants were required to navigate more to reach it, thus induced longer selection time. However, participants spent more time in selecting the lower depth target using *MagicBall+* than the higher depth ones. This was because participants were observing the whole environment from the outside of the mini-map, lower depth target actually looked farther away. Hence, there might be more distractors on the way of getting the goal target.

After seeing how each environmental factor affected the performance of the techniques, we compared the techniques in terms of different measurements below. The performance data were consistent in terms of selection time and error rates. *LassoGrid+* had the lowest selection time and error rate, while for *MagicBall+* they were the highest. The NASA-TLX results also show a similar trend. Participants were more satisfied with their performance and had lower frustration and mental workload levels when using *LassoGrid+*. Concerning the UEQ-S results, *LassoGrid+* was shown to have excellent pragmatic value, while *MagicBall+* was rated outstanding in the hedonic quality. They both had excellent overall quality. However, *GravityZone+* was rated just above-average on all aspects of the UEQ-S. It seemed to suffer from the "middle children syndrome" [43], where it did not look as innovative as *MagicBall+* and was not as effective as *LassoGrid+*. Therefore, its ratings from the participants were relatively low.

10 Design Recommendations

Based on the results from both studies, we distill design recommendations regarding choosing input techniques for the selection of fully-occluded target in virtual environments.

R1. When the goal of the task is rapid selection, we suggest using grid-based techniques (*GridWall*, *FlowerCone*, and *LassoGrid+*) to ensure optimal user performance. Use *LassoGrid+* when it is difficult to decide which one of them to use, as it allows users to define their estimate of where targets might occur freely, and only one trigger/button will be needed for the whole selection process. Consider adding selection enhancement techniques (like highlighting the closest object) to improve performance further.

R2. When maintaining the object location information is essential (e.g., 3D plots), we recommend using depth-based techniques (*AlphaCursor* and *GravityZone+*) or *MagicBall+*. Our results indicate that *GravityZone+* should be favored if better performance is needed. *AlphaCursor* and *MagicBall+* can be used when it is not desirable to move the objects in the scene.

R3. If the technique is used for recreational purposes (like game applications), consider use *SmashProbe* and *MagicBall+* as they are more exciting or have higher hedonic quality. However, avoid using *SmashProbe* when there are too many objects in the scene, as it could be very distracting.

R4. Be sure to consider the environmental factors (occlusion layers, target depths, object densities, and the estimation of target location) of the application and how they might influence the performance of the technique. If the environment constantly changes, as a rule of thumb, use *LassoGrid+* as it was shown to be relatively robust in terms of performance.

11 Demonstrations

Based on our findings, we have developed two proof-of-concept demonstrations in VR showing the techniques in real application scenarios (see Figure 9). The first demo shows an ocean exploration scenario in VR, which belongs to the case of exploring complex 3D data visualizations. Users are immersed under the ocean and surrounded by a large

number of underwater creatures. With our techniques, they can select an animal of interest that lives in certain areas or is hidden by corals to delve into its detailed information (like name, habitat, life cycles, etc.). A similar scenario would be to explore specific locations occluded by buildings in a 3D city visualization. The second demo mimics a 3D modeling scenario. Users can acquire fully-occluded objects in the scene and perform consequent manipulations like translation and duplication. Both applications are demonstrated in the supplementary video.

Fig. 9. (a) In the sea exploration scenario, a user used *LassoGrid+* to learn about animals (which might be fully-occluded) living within a particular area. (b) *AlphaCursor* reveals the hidden tree in the modeling scene.

12 Limitations

We have identified several limitations in our work. First, for simplicity, we simulated a user's estimated area of where a target might occur only in a circular form. However, in a real-world scenario, multiple occurrence areas can exist, and they can be in any shape, even with some depth.

Second, we did not fine-tune the parameters of all the techniques through user studies, as it was not the primary goal of this work. For example, instead of arranging objects on a grid, other layouts are also possible (e.g., rings [10]), which could further improve the performance of the techniques.

Third, we did not include unselectable objects in the scene, as we envision a superimposed scenario that culls out the unselectable objects for the ease of selection. However, future work might want to investigate how unselectable objects can be embedded into the scene and how various properties related to these objects (like sizes and placements) can affect the selection.

Fourth, our experiments feature more abstract tasks that enabled us to control the variables of interest precisely, however, we did not evaluate technique performance under practical scenarios. To strike a balance between internal and external validity of our findings, though two proof-of-concept demonstrations are provided, more work is necessary to understand how the techniques can perform and how we can adjust them in realistic workflows. For example, future work can explore how the techniques could be applied to disambiguate vertex or edge selection in 3D modeling applications.

13 Conclusion

In this paper, we explored fully-occluded target selection in virtual reality environments. Based on the existing literature on the topic, we highlighted three open challenges within this research topic in terms of problem formulation, combining occlusion visualization with selection techniques, and in-depth evaluation. To address them, we first framed a general problem-solving strategy and, according to that, devised the design space. We then designed seven potential techniques and evaluated them through a user study.

Based on the study results, we derived design implications and refined the most promising techniques. We conducted a second study to analyze how four environmental factors (occlusion layers, target depths, object densities, and the estimation of target locations) influence technique performance. Based on our findings, we offer a set of distilled recommendations for future virtual reality systems that offer fully-occluded target selection. We believe our design approaches and proposed techniques can trigger the creation of exciting user interfaces and applications related to fully-occluded selection. Future work can optimize further the techniques, as well as develop new methods for selecting fully-occluded targets in VR.

Acknowledgments

We thank our participants for their interest in the project and insightful discussions. We also appreciate the reviewers for their professionalism and dedication that helped improve our paper. This research was supported by the Melbourne Research Scholarship provided by The University of Melbourne.

References

[1] D. H. Abelow. Reality alternate, Nov. 10 2015. US Patent 9,183,560.

[2] Agustina and C. Sun. Xpointer: An x-ray telepointer for relaxed-space-time wysiwis and unconstrained collaborative 3d design systems. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pp. 729–740. ACM, New York, NY, USA, 2013. doi: 10.1145/2441776.2441857

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

[4] F. Argelaguet, A. Kunert, A. Kulik, and B. Froehlich. Improving co-located collaboration with show-through techniques. In *Proceedings of the 2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 55–62, March 2010. doi: 10.1109/3DUI.2010.5444719

[5] J. Arora, A. Saini, N. Mehra, V. Jain, S. Shrey, and A. Parnami. Virtualbricks: Exploring a scalable, modular toolkit for enabling physical manipulation in vr. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 56:1–56:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300286

[6] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):457–467, Jan 2018. doi: 10.1109/TVCG.2017.2745941

[7] F. Bacim, R. Kopper, and D. A. Bowman. Design and evaluation of 3d selection techniques based on progressive refinement. *International Journal of Human-Computer Studies*, 71(7):785 – 802, 2013. doi: 10.1016/j.ijhcs.2013.03.003

[8] F. Bacim, M. Nabiyouni, and D. A. Bowman. Slice-n-swipe: A free-hand gesture user interface for 3d point cloud annotation. In *Proceedings of the 2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 185–186. IEEE, 2014. doi: 10.1109/3DUI.2014.6798882

[9] F. Bacim de Araujo e Silva. *Increasing Selection Accuracy and Speed through Progressive Refinement*. PhD thesis, Virginia Tech, 2015.

[10] G. Bailly, E. Lecolinet, and L. Nigay. Visual menu techniques. *ACM Computer Survey*, 49(4):60:1–60:41, Dec. 2016. doi: 10.1145/3002171

[11] M. Baloup, T. Pietrzak, and G. Casiez. Raycursor: A 3d pointing facilitation technique based on raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 101:1–101:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300331

[12] L. Besançon, M. Sereno, L. Yu, M. Ammi, and T. Isenberg. Hybrid touch/tangible spatial 3d data selection. In *Computer Graphics Forum*, vol. 38, pp. 553–567. Wiley Online Library, 2019. doi: 10.1111/cgf.13710

[13] C. Bichlmeier, S. M. Heining, M. Feuerstein, and N. Navab. The virtual mirror: A new interaction paradigm for augmented reality environments. *IEEE Transactions on Medical Imaging*, 28(9):1498–1510, Sep. 2009. doi: 10.1109/TMI.2009.2018622

[14] D. Bowman, C. Wingrave, J. Campbell, and V. Ly. Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments. 2001.

[15] S. Bruckner and M. E. Groller. Exploded views for volume data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1077–1084, Sep. 2006. doi: 10.1109/TVCG.2006.140

[16] M. Burns and A. Finkelstein. Adaptive cutaways for comprehensible rendering of polygonal scenes. In *Proceedings of the ACM SIGGRAPH Asia 2008 Papers*, SIGGRAPH Asia '08, pp. 154:1–154:7. ACM, New York, NY, USA, 2008. doi: 10.1145/1457515.1409107

[17] M. S. T. Carpendale, D. J. Cowperthwaite, and F. D. Fracchia. Distortion viewing techniques for 3-dimensional data. In *Proceedings IEEE Symposium on Information Visualization '96*, pp. 46–53, Oct 1996. doi: 10.1109/INFVIS.1996.559215

[18] J. Cashion. Intelligent selection techniques for virtual environments. 2014.

[19] J. Cashion, C. Wingrave, and J. J. L. Jr. Dense and dynamic 3d selection for game-based virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):634–642, April 2012. doi: 10.1109/TVCG.2012.40

[20] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson. Immersive analytics. In *Proceedings of the 2015 Big Data Visual Analytics (BDVA)*, pp. 1–8, Sep. 2015. doi: 10.1109/BDVA.2015.7314296

[21] O. Chapuis, J.-B. Labrune, and E. Pietriga. Dynaspot: Speed-dependent area cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pp. 1391–1400. ACM, New York, NY, USA, 2009. doi: 10.1145/1518701.1518911

[22] M. C. Chuah, S. F. Roth, J. Mattis, and J. Kolojejchick. Sdm: Selective dynamic manipulation of visualizations. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 61–70, 1995.

[23] C. Coffin and T. Hollerer. Interactive perspective cut-away views for general 3d scenes. In *Proceedings of the 3D User Interfaces (3DUI'06)*, pp. 25–28, March 2006. doi: 10.1109/VR.2006.88

[24] C. Correa, D. Silver, and M. Chen. Illustrative deformation for data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1320–1327, Nov 2007. doi: 10.1109/TVCG.2007.70565

[25] H. G. Debarba, J. G. Grandi, A. Maciel, L. Nedel, and R. Boulic. Disambiguation canvas: A precise selection technique for virtual environments. In *Proceedings of the Human-Computer Interaction – INTERACT 2013*, pp. 388–405. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-40477-1_24

[26] S. Deng, J. Chang, S.-M. Hu, and J. J. Zhang. Gaze modulated disambiguation technique for gesture control in 3d virtual objects selection. In *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, pp. 1–8, June 2017. doi: 10.1109/CYBConf.2017.7985779

[27] J. Diepstraten, D. Weiskopf, and T. Ertl. Transparency in interactive technical illustrations. *Computer Graphics Forum*, 21(3):317–325, 2002. doi: 10.1111/1467-8659.t01-1-00591

[28] J. Diepstraten, D. Weiskopf, and T. Ertl. Interactive cutaway illustrations. *Computer Graphics Forum*, 22(3):523–532, 2003. doi: 10.1111/1467-8659.t01-3-00700

[29] N. Elmqvist. Balloonprobe: Reducing occlusion in 3d using interactive space distortion. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '05, pp. 134–137. ACM, New York, NY, USA, 2005. doi: 10.1145/1101616.1101643

[30] N. Elmqvist, U. Assarsson, and P. Tsigas. Dynamic transparency for 3d visualization: design and evaluation. *The International Journal of Virtual Reality*, 1(8):65–78, 2009.

[31] N. Elmqvist and P. Tsigas. A taxonomy of 3d occlusion management techniques. In *Proceedings of the 2007 IEEE Virtual Reality Conference*, pp. 51–58, March 2007. doi: 10.1109/VR.2007.352463

[32] N. Elmqvist and M. E. Tudoreanu. Occlusion management in immersive and desktop 3d virtual environments: Theory and evaluation. *International Journal of Virtual Reality*, 6(2):21–32, 2007.

[33] T. T. Elvins, D. R. Nadeau, and D. Kirsh. Worldlets—3d thumbnails for wayfinding in virtual environments. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*, UIST '97, p. 21–30. Association for Computing Machinery, New York, NY, USA, 1997. doi: 10.1145/263407.263504

[34] S. K. Feiner and D. D. Seligmann. Cutaways and ghosting: satisfying visibility constraints in dynamic 3d illustrations. *The Visual Computer*, 8(5):292–302, Sep 1992. doi: 10.1007/BF01897116

[35] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954. doi: 10.1037/h0055392

[36] A. Forsberg, K. Herndon, and R. Zeleznik. Aperture based selection for immersive virtual environments. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, UIST '96, pp. 95–96. ACM, New York, NY, USA, 1996. doi: 10.1145/237091.237105

[37] M. Funk, F. Müller, M. Fendrich, M. Shene, M. Kolvenbach, N. Dobbertin, S. Günther, and M. Mühlhäuser. Assessing the accuracy of point & teleport locomotion with orientation indication for virtual reality using curved trajectories. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 147:1–147:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300377

[38] T. Grossman and R. Balakrishnan. The design and evaluation of selection techniques for 3d volumetric displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pp. 3–12. ACM, New York, NY, USA, 2006. doi: 10.1145/1166253.1166257

[39] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, pp. 904–908, 2006. doi: 10.1177/154193120605000909

[40] T. Hirzle, J. Gugenheimer, F. Geiselhart, A. Bulling, and E. Rukzio. A design space for gaze interaction on head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 625:1–625:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300855

[41] D. Kalkofen, E. Veas, S. Zollmann, M. Steinberger, and D. Schmalstieg. Adaptive ghosted views for augmented reality. In *Proceedings of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–9. IEEE, 2013. doi: 10.1109/ISMAR.2013.6671758

[42] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–74, March 2011. doi: 10.1109/3DUI.2011.5759219

[43] J. Kotin. *Getting started: An introduction to dynamic psychotherapy*. Jason Aronson, 2004.

[44] G. Kurtenbach and W. Buxton. The limits of expert performance using hierarchic marking menus. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pp. 482–487. ACM, New York, NY, USA, 1993. doi: 10.1145/169059.169426

[45] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head- and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 81:1–81:14. ACM, New York, NY, USA, 2018. doi: 10.1145/3173574.3173655

[46] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[47] K. Lilija, H. Pohl, S. Boring, and K. Hornbæk. Augmented reality views for occluded interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 446:1–446:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300676

[48] M. A. Livingston, J. E. Swan, J. L. Gabbard, T. H. Hollerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 56–65. IEEE, 2003. doi: 10.1109/ISMAR.2003.1240688

[49] D. Mendes, D. Medeiros, E. Cordeiro, M. Sousa, A. Ferreira, and J. Jorge. Precious! out-of-reach selection using iterative refinement in vr. In *Proceedings of the 2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 237–238, March 2017. doi: 10.1109/3DUI.2017.7893359

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

[51] A. Olwal and S. Feiner. The flexible pointer: An interaction technique for augmented and virtual reality. In *Proc. of ACM Symposium on User Interface Software and Technology (UIST*, pp. 81–82, 2003.

[52] J. Petford, I. Carson, M. A. Nacenta, and C. Gutwin. A comparison of notification techniques for out-of-view objects in full-coverage displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 58:1–58:13. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300288

[53] J. Petford, M. A. Nacenta, and C. Gutwin. Pointing all around you: Selection performance of mouse and ray-cast pointing in full-coverage displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 533:1–533:14. ACM, New York, NY, USA, 2018. doi: 10.1145/3173574.3174107

[54] B. Preim, A. Baer, D. Cunningham, T. Isenberg, and T. Ropinski. A survey of perceptually motivated 3d visualization of medical image data. *Computer Graphics Forum*, 35(3):501–525, 2016. doi: 10.1111/cgf.12927

[55] G. Ramos, G. Robertson, M. Czerwinski, M. Czerwinski, D. Tan, P. Baudisch, K. Hinckley, K. Hinckley, and M. Agrawala. Tumble! splat! helping users access and manipulate occluded content in 2d drawings. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, pp. 428–435. ACM, New York, NY, USA, 2006. doi: 10.1145/1133265.1133351

[56] G. Ren and E. O'Neill. 3d selection with freehand gesture. *Computers & Graphics*, 37(3):101 – 120, 2013. doi: 10.1016/j.cag.2012.12.006

[57] H. Ro, S. Chae, I. Kim, J. Byun, Y. Yang, Y. Park, and T. Han. A dynamic depth-variable ray-casting interface for object manipulation in ar environments. In *Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2873–2878. IEEE, 2017. doi: 10.1109/SMC.2017.8123063

[58] M. Röhlig, M. Luboschik, and H. Schumann. Visibility widgets for unveiling occluded data in 3d terrain visualization. *Journal of Visual Languages & Computing*, 42:86 – 98, 2017. doi: 10.1016/j.jvlc.2017.08.008

[59] R. Rouse III. *Game design: Theory and practice*. Jones & Bartlett Learning, 2010.

[60] J. Sauro and J. S. Dumas. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pp. 1599–1608. ACM, New York, NY, USA, 2009. doi: 10.1145/1518701.1518946

[61] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and evaluation of a short version of the user experience questionnaire (ueq-s). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6):103–108, 2017. doi: 10.9781/ijimai.2017.09.001

[62] V. Schwind, P. Knierim, N. Haas, and N. Henze. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 360:1–360:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300590

[63] B. Shneiderman. Why not make interfaces better than 3d reality? *IEEE Computer Graphics and Applications*, 23(6):12–15, Nov 2003. doi: 10.1109/MCG.2003.1242376

[64] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen. Outline pursuits: Gaze-assisted selection of occluded objects in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376438

[65] R. B. Smith. Experiences with the alternate reality kit: An example of the tension between literalism and magic. *SIGCHI Bulletin*, 18(4):61–67, May 1986. doi: 10.1145/1165387.30861

[66] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International Journal of Human-Computer Studies*, 61(6):751 – 789, 2004. doi: 10.1016/j.ijhcs.2004.09.001

[67] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: Interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pp. 265–272. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1995. doi: 10.1145/223904.223938

[68] J. Sun and W. Stuerzlinger. Selecting and sliding hidden objects in 3d desktop environments. 2019. doi: 10.20380/GI2019.08

[69] R. J. Teather and W. Stuerzlinger. Pointing at 3d target projections with one-eyed and stereo cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pp. 159–168. ACM, New York, NY, USA, 2013. doi: 10.1145/2470654.2470677

[70] R. Trueba, C. Andujar, and F. Argelaguet. Complexity and occlusion management for the world-in-miniature metaphor. In *Smart Graphics*, pp. 155–166. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[71] H. Tu, S. Huang, J. Yuan, X. Ren, and F. Tian. Crossing-based selection with virtual reality head-mounted displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 618:1–618:14. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300848

[72] L. Vanacken, T. Grossman, and K. Coninx. Exploring the effects of environment density and target visibility on object selection in 3d virtual environments. In *Proceedings of the 2007 IEEE Symposium on 3D User Interfaces*, March 2007. doi: 10.1109/3DUI.2007.340783

[73] E. Velloso, M. Carter, J. Newn, A. Esteves, C. Clarke, and H. Gellersen. Motion correlation: Selecting objects by matching their movement. *ACM Transactions on Computer-Human Interaction*, 24(3), Apr. 2017. doi: 10.1145/3064937

[74] A. M. Webb, A. Kerne, Z. Brown, J.-H. Kim, and E. Kellogg. Layerfish: Bimanual layering with a fisheye in-place. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, ISS '16, pp. 189–198. ACM, New York, NY, USA, 2016. doi: 10.1145/2992154.2992171

[75] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963

[76] M.-L. Wu and V. Popescu. Multiperspective focus+context visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(5):1555–1567, May 2016. doi: 10.1109/TVCG.2015.2443804

[77] H. P. Wyss, R. Blach, and M. Bues. isith - intersection-based spatial interaction for two hands. In *Proceedings of the 3D User Interfaces*, 3DUI '06, pp. 59–61. IEEE Computer Society, Washington, DC, USA, 2006. doi: 10.1109/VR.2006.93

[78] Y. Yan, C. Yu, X. Ma, X. Yi, K. Sun, and Y. Shi. Virtualgrasp: Leveraging experience of interacting with physical objects to facilitate digital object retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 78:1–78:13. ACM, New York, NY, USA, 2018. doi: 10.1145/3173574.3173652

[79] D. Yu, H.-N. Liang, K. Fan, H. Zhang, C. Fleming, and K. Papangelis. Design and evaluation of visualization techniques of off-screen and occluded targets in virtual reality environments. *IEEE Transactions on Visualization and Computer Graphics*, 2019. doi: 10.1109/TVCG.2019.2905580

[80] D. Yu, H.-N. Liang, X. Lu, K. Fan, and B. Ens. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Transactions on Graphics*, 38(6):13, November 2019. doi: 10.1145/3355089.3356544

[81] D. Yu, Q. Zhou, B. Tag, T. Dingler, E. Velloso, and J. Goncalves. Engaging participants during selection studies in virtual reality. In *Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 500–509. IEEE, 2020. doi: 10.1109/VR46266.2020.00071

[82] S. Zhai, W. Buxton, and P. Milgram. The partial-occlusion effect: Utilizing semitransparency in 3d human-computer interaction. *ACM Transaction on Computer-Human Interaction*, 3(3):254–284, Sept. 1996. doi: 10.1145/234526.234532

Chapter 5

GAZE-SUPPORTED 3D OBJECT MANIPULATION

5.1 Summary

In this work, we propose interaction techniques incorporating eye gaze input into the object manipulation process based on mid-air input (i.e., Virtual Hand and Raycasting). While mid-air input is probably more suitable for 3D controls, gaze has been identified as a lightweight and fast input method, which can be a helpful complementary modality. Our techniques considered integration, coordination, and transition strategies of gaze and hand input and were evaluated in two user studies. The user studies covered a controlled working space with all objects within arm-reach distance and a larger virtual environment with distant objects and realistic workflows (i.e., reconstructing a virtual room).

The proposed gaze-supported 3D object manipulation techniques can handle small and distant objects. They were demonstrated to be more efficient in interacting with out-of-reach objects, induce less arm fatigue, and provide more desirable user experiences. The techniques can be applied to various applications containing objects of different sizes and distances.

Env.			**Task**				
Small	*Distant*	*Occluded*	*Effectiveness*	*Efficiency*	*Ergonomics*	*Experience*	*Expressivity*
✓	✓			✓	✓	✓	✓

5.2 Article II

Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. "Gaze-Supported 3D Object Manipulation in Virtual Reality." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1-13. 2021. https://doi.org/10.1145/3411764.3445343

Gaze-Supported 3D Object Manipulation in Virtual Reality

Difeng Yu*
The University of Melbourne
difeng.yu@student.unimelb.edu.au

Xueshi Lu
Xi'an Jiaotong-Liverpool University
xueshi.lu17@student.xjtlu.edu.cn

Rongkai Shi
Xi'an Jiaotong-Liverpool University
rongkai.shi19@student.xjtlu.edu.cn

Hai-Ning Liang*
Xi'an Jiaotong-Liverpool University
haining.liang@xjtlu.edu.cn

Tilman Dingler
The University of Melbourne
tilman.dingler@unimelb.edu.au

Eduardo Velloso
The University of Melbourne
eduardo.velloso@unimelb.edu.au

Jorge Goncalves
The University of Melbourne
jorge.goncalves@unimelb.edu.au

ABSTRACT

This paper investigates integration, coordination, and transition strategies of gaze and hand input for 3D object manipulation in VR. Specifically, this work aims to understand whether incorporating gaze input can benefit VR object manipulation tasks, and how it should be combined with hand input for improved usability and efficiency. We designed four gaze-supported techniques that leverage different combination strategies for object manipulation and evaluated them in two user studies. Overall, we show that gaze did not offer significant performance benefits for transforming objects in the primary working space, where all objects were located in front of the user and within the arm-reach distance, but can be useful for a larger environment with distant targets. We further offer insights regarding combination strategies of gaze and hand input, and derive implications that can help guide the design of future VR systems that incorporate gaze input for 3D object manipulation.

CCS CONCEPTS

• **Human-centered computing** → **User interface design**; *User studies*; *Virtual reality*; *Interaction techniques.*

KEYWORDS

3D object manipulation, gaze input, multimodal interface

ACM Reference Format:
Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-Supported 3D Object Manipulation in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445343

*The corresponding authors from each institution are Difeng Yu and Hai-Ning Liang.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445343

1 INTRODUCTION

As one of the primary tasks in virtual reality (VR) systems, object manipulation is used in many different application domains such as 3D modeling [14, 20], game development [18], online collaboration [25, 38], and immersive data exploration [2, 8]. However, its primary input modality, which uses virtual hands to "direct manipulate" an object, has long been criticized for being inefficient and imprecise [7, 36], and likely to induce arm-fatigue in longer interaction scenarios [16, 29, 35].

Alternatively, gaze has been identified as a light-weight and fast input method, and has shown its potential for assisting with object manipulation tasks (e.g., [33, 40, 59]). However, previous work in VR mostly focused on the use of gaze for target selection [39, 45], which is only a sub-phase of the whole manipulation process, while how gaze input can be incorporated into the "manipulate" phase (translation, rotation, and scaling [31]) is still underexplored. Thus, this research aims to understand *whether* the incorporation of gaze input can benefit the hand manipulation process in VR, and *how* gaze input should be combined with hand input for convenient and efficient 3D object manipulation.

This research investigates different integration, coordination, and transition strategies when incorporating gaze into current systems with mid-air hand input for 3D object manipulation in VR. Specially, we examine a design space that considers how gaze and hand input are *integrated* into different phases of the manipulation task, how they *coordinate* with each other when starting the manipulation, and how to *transition* from one to the other during manipulation. Based on this design space, we developed four gaze-supported manipulation techniques and evaluated them through two user studies. In the first study, we focused on the *primary working space*, where all objects located in front of the user and were within arm-reach distance, and assessed the techniques in terms of user performance and experience. In the second study, we further evaluated our techniques in a larger virtual environment with distant objects and embedded the designed techniques into realistic workflows.

Our findings show that gaze might not offer significant performance benefits for transforming objects in the primary working space, but can be useful in a larger environment with distant targets, while also mitigating the arm fatigue issue. We further derived a set of design implications that reveal the usefulness of different strategies, including hand-only *vs.* eye-hand manipulation, direct *vs.*

remote hand mappings, and implicit *vs.* explicit eye-hand transitions. Our findings and implications provide a helpful guide for the design of future gaze-supported object manipulation techniques in VR.

To summarize, the main contributions of the paper include:

- The design space of how to incorporate gaze into the traditional hand-based object manipulation workflow.
- A novel implicit transition-based approach (called *ImplicitGaze*).
- The evaluation of the techniques, which has led to useful findings and design implications (whether it is beneficial to incorporate gaze and what can be done to improve interaction).

2 RELATED WORK

Here we introduce the most commonly used approaches and recent advances regarding VR object manipulation (also see more thorough recent reviews [31, 36]). We further discuss gaze-supported techniques used in VR and other domains.

2.1 Object Manipulation in VR

Mid-air interaction based on *Virtual Hand* is one of the primary input paradigm for modern VR systems [36]. With spatially tracked hand positions, typically with 6 degrees-of-freedom (DoF), users are able to directly translate and rotate objects in virtual environments in a similar way as they manipulate them in the physical world [44]. Although it has been criticized to be inefficient and imprecise [7, 36], due to its simplicity and intuitiveness of the control, Virtual Hand has been widely applied in various VR applications [14, 20, 27].

Further approaches have been used to enhance Virtual Hand. For example,*Go-Go* [43] and its recent extension [69], which scale up the speed of the virtual hand, enable users to reach distant targets, even at a potentially infinite distance [5]. *Raycasting* also provides an easy solution for acquiring distant objects, but users may not be able to rotate the object precisely with one single hand as they are attached to the end of the ray [5]. Other methods [41, 57, 73] scale-down the whole virtual world to enable the interaction with out-of-reach objects.

To offer fine-grained manipulation control, several interaction techniques decrease the control-display ratio of the hand movement based on hand velocity [17, 70]. Degree-of-freedom (DoF) separation [37, 65] is another promising way to increase the accuracy of mid-air object manipulation—that is, rather than manipulating all the six DoF simultaneously, only one or two of them are controlled each time. For instance, in a recent work, researchers tried to reduce the DoF during object manipulation by constraining it to the shape of a point, ray, or plane, thereby increasing precision [22].

Nevertheless, many mid-air interaction techniques fall short in supporting prolonged manipulation due to cumulative arm muscle fatigue (the so-called "gorilla arm" effect) [29]. This is especially detrimental to interaction scenarios such as 3D modeling in VR, which require fine-grained, focused, and prolonged usage of mid-air interfaces. To address these challenges, providing indirect mappings [35] or integrating other less effort-demanding input modalities such as gaze into object manipulation techniques in VR can be potentially helpful.

2.2 Gaze-Supported Manipulation

Gaze-supported object manipulation has been widely explored in contexts outside VR. In general, while gaze offers fast and natural pointing, it suffers from the lack of precision and the difficulty of confirming a selection. To overcome these challenges, many techniques combine gaze with an additional modality, such as the principle of "gaze select, hands manipulate" [9, 39, 56, 66]. For example, Pfeuffer et al. proposed *Gaze-touch* [39], which enabled users to control gaze-selected targets indirectly using multi-touch gestures on interactive surfaces. Another example is the method proposed by Turner et al. [62], which casts the object being looked at by the user to the touch/cursor position to allow further manipulation. In contrast, other approaches [48, 55, 60–63, 67] for content-transfer between different displays, have embedded gaze movement into the translation process. These prototypes typically require the use of a hand trigger to "attach" the object to the gaze direction and then release the hand trigger to "drop" it. In a follow up research, Turner et al. [59] pushed this concept further by developing techniques that maintain concurrent rotation and scaling operations when performing translation tasks using gaze and touch.

Limited work has investigated gaze input for object manipulation in VR or 3D virtual space. Simeone et al. [52] combined bi-manual touch gestures with gaze input to allow the scaling of objects on the XYZ-axis inside a touchscreen. Liu et al. have presented *OrthoGaze* [34], in which gaze is issued to move an object along three orthogonal planes in VR. Other researchers have used eye gaze to select objects, and leveraged indirect freehand gestures to manipulate them [40, 42, 45, 54]. All of them still followed the idea of "gaze select, hands manipulate". In contrast to these approaches, the gaze input in our work was not only used for the selection of objects but also was involved in the whole target manipulation process, which requires continuous actions rather than the discrete selection operation [59]. Our aim is to understand how different methods of hand-eye integration, coordination, and transition can result in improved user performance and their suitability to be applied to a variety of scenarios.

2.3 Transition Between Gaze and Hand Input

Different collaboration strategies have been explored to combine gaze and other input modalities, such as hands or head [49, 50], and the transition between different modalities can be classified according to whether they are explicit and implicit. Explicit transitions rely on specifically issued commands to switch between gaze and other forms of input. The "switch" orders include actuating the input device or pressing a trigger. In contrast, implicit transitions do not rely on distinct commands to switch between multiple input mechanisms; all modalities always have an effect on the cursor/object that users interact with, and users do not need to concern about the transition during the interaction.

An example of the explicit transition is *Pinpointing* [30], which starts with a fast but imprecise modality like gaze, and then refining it with a slower but more precise input modality such as hand gestures with an explicit button click or a finger gesture for mode transition. An example of the implicit transition is "liberal" *MAGIC* pointing [76], where users can always move the cursor with manual or gaze input once they have decided to do so, without activating any trigger. While explicit transitions offer more robust control in many cases [30], implicit transitions fade the boundary between the input mechanisms and smooth the "flow" of interaction.

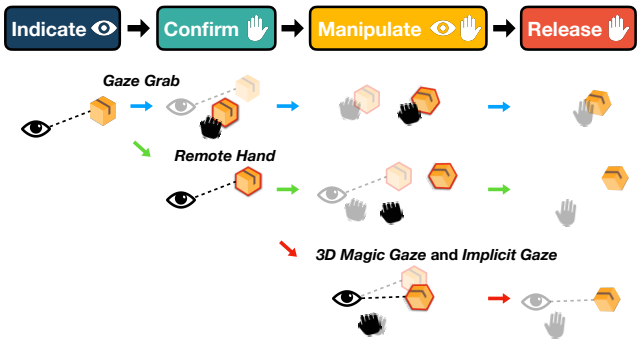


Figure 1: An illustration of the target manipulation process, where gaze is used for indicate, hands are used for trigger confirm and release, and both inputs are applied (as hand only, or gaze and hand collaboratively) for object manipulation.

It is important to note that we distinguish "implicit" from other names presented in the literature like "seamless" [51, 55], the smoothness of the transition, and "concurrent" [59], the ability to manipulate multiple degrees-of-freedom simultaneously. Seamlessness and concurrency do not ensure an implicit transition; we only consider if an explicit triggering mechanism is used. As discussed, previous works on gaze-supported VR object manipulation mainly used gaze as a selection technique, rather than incorporated it into the manipulate phase which includes translation, rotation, and scaling. Therefore, how to transition between gaze and hand input for manipulation tasks is still underexplored.

3 DESIGN SPACE

3D object manipulation techniques can be broken down into an initial selection of object and the later manipulation steps including translation, rotation, and scaling [6, 40]. We first introduce this process and propose corresponding input modalities for each sub-phase. We then formulate a design space that considers how gaze and hand input are integrated into different phases of the manipulation task, coordinate with each other when starting the manipulation, and transition from one to the other during manipulation. Based on the design space, we further point out several gaps in the existing literature, and use this knowledge to design our proposed techniques.

3.1 Target Manipulation Process

We first introduce a target manipulation process that is based on previous works [6, 40]. The whole task can be decomposed into four phases: indicate, confirm, manipulate, and release (see Figure 1). We identified suitable input modalities for each phase, which is then useful to structure and narrow down our exploration space.

3.1.1 Indicate. Indicating is the action of determining the target of interest with an input device. The literature suggests that gaze-based pointing requires less effort and can be faster than manual input [39, 40, 55]. Further, gaze tracking has become more accurate with recent advances in the field [15]. Therefore, we consider gaze as our input mechanism in the indicate phase.

3.1.2 Confirm. Confirming the selection allows users to "pick up" and start manipulating the indicated target. Because gaze-based confirming techniques, such as dwell, can be inefficient and may induce unwanted selection [26], we decided to use a hand-based method, specifically, pressing the trigger on the hand-held controller for a robust control of the confirm phase.

3.1.3 Manipulate. Manipulation of objects, including translation, rotation, and scaling, can be achieved by hand input alone, or by gaze and hand input together. Gaze can be treated as a 2 degrees-of-freedom (DoF) modality as an estimated gaze point normally moves on a 2D spherical plane, while accurately predicting the depth of the gaze point can be challenging [24]. In contrast, hand-based mechanisms typically feature 6 DoF motion input (both translation and rotation along the 3 axes). Based on its properties, gaze offers more opportunities for rapid translating objects in the lateral direction [59]. As for hands, they are likely to be better in positioning objects in the depth dimension (the third DoF), and rotating or scaling them (as they either require the rotation of the input device or need multiple control points). To distinguish this phase from the whole target manipulation process, we call it *the manipulate phase* in this paper.

3.1.4 Release. Releasing the trigger signals the completion of one operation. Similar to the confirm phase, we use the trigger on the controller for the robust control of the release phase.

3.2 Design Dimensions

We considered the following three-dimensional design space by emphasizing the integration, coordination, and transition of gaze and hand input for the manipulate phase. While we acknowledge that exploring other design dimensions such as target properties and input techniques can be useful, this research focuses on exploring how to incorporate gaze-input into the traditional hand-based workflow.

- **D1. Integration**: which input mechanism(s) of gaze and hand has (have) been integrated into the manipulate phase.
- **D2. Coordination**: when starting the manipulate phase, if the indicated target will snap to the hand position or remain in its original place. This further corresponds to whether the object is directly mapped onto the hand position (direct mapping) or manipulated by hands remotely (remote mapping).
- **D3. Transition**: if both input mechanisms are involved in the manipulate phase, whether the transition between gaze and hand input is explicit or implicit (with or without specifically issued triggering commands like button pressing).

3.2.1 Synthesis of Prior Work. We further summarized how existing gaze-supported manipulation techniques fit into each dimension of the design space (see Table 1). We have focused on the ones that involve hand input in the manipulate phase, rather than relying on the gaze input alone. That is, the approaches that use gaze input only as a supporting mechanism for manipulation.

3.2.2 Research Gaps and Design Opportunities. The design space and the synthesis of prior work reveal some research gaps that are essential for framing the design of gaze-supported object manipulation techniques but are still underexplored in the literature and thus create new design opportunities.

	Techniques	Integration		Coordination		Transition		
		Gaze	Hand	Direct	Remote	Implicit	Explicit	None
2D	Eye drop [61, 62]		✓	✓				✓
	TouchGP [55]	✓	✓		✓		✓	
	Gaze-Touch [39]		✓		✓			✓
	TouchT [59]		✓		✓			✓
	GazeT [59]	✓	✓		✓		✓	
	MagicT [59]	✓	✓		✓		✓	
	Gaze [66]		✓		✓			✓
3D	Gaze + Non-touch [42]		✓		✓			✓
	Three-point [52]	✓	✓	✓				✓
	Gaze + pinch [40]		✓		✓			✓
	GG interaction [45]		✓		✓			✓
	Gaze + Gesture [9, 10, 54]		✓		✓			✓
	Gaze Grab		✓	✓				✓
	Remote Hand		✓		✓			✓
	3D Magic Gaze	✓	✓		✓		✓	
	Implicit Gaze	✓	✓		✓	✓		

Table 1: Summary of how existing gaze-supported manipulation solutions and ours (the bottom four) fit into the design space. Our techniques enabled us to explore explicit and implicit transitions, which have not been well-covered by previous research in 3D, and how different design dimensions may influence user performance and experiences in VR manipulation.

G1. *Transition mechanisms between gaze and hand input have not been investigated in the manipulate phase in VR*; most of the previous work focused on the rationale of "gaze select, touch manipulate". However, gaze can not only support discrete pointing tasks but can also be beneficial for target manipulation, which requires continuous actions [55, 59]. Further exploration is needed to understand how gaze input supports manipulation in VR environments, which offers 3D spatial input and stereo vision [31].

G2. *Implicit transition is still under-explored for target manipulation tasks in general.* According to Table 1, there is lack of implicit transition techniques in the manipulate phase. All transitions are based on either releasing a pressed trigger [55] or exceeding a hand movement threshold [59] to switch from gaze input to hand input.

G3. *Techniques that leverage different elements of the design space have not been compared in terms of their efficiency and usability.* For example, it is unclear how gaze-supported methods that allow remote (indirect) manipulation compare to direct manipulation-based solutions in terms of performance and user experiences, although they have been applied in different applications [68]. Furthermore, it is still unclear if gaze-supported techniques can provide more benefits than hand-only techniques in the manipulate phase in VR.

4 TECHNIQUE DESIGN

Based on the identified research gaps and design opportunities, we developed the following four techniques to (1) explore transition mechanisms (**G1 - 3DMagicGaze**), especially implicit transition (**G2 - ImplicitGaze**), for target manipulation in VR and (2) evaluate and compare approaches that leverage different elements of the design space in terms of user performance, experiences, and their suitability to be applied to a variety of scenarios (**G3**). Table 1 shows how each technique fits within the design space.

4.1 Gaze Grab

With *GazeGrab*, the gaze-indicated target snaps to the hand position once the selection is confirmed. Next, the hand takes full control of the selected target during the manipulation phase until the trigger is released. This technique allows the direct manipulation of objects and represents a VR-enhanced version of previous research on content transfer [61]. Similar techniques have also been demoed in VR applications [68], though it has not been empirically evaluated or compared with other techniques. In our design, the gaze-grabbed object is located slightly above the virtual hand position, to avoid visual occlusion.

4.2 Remote Hand

To manipulate an object through *RemoteHand*, a user first points at it with eye gaze and then confirms the selection with a hand trigger. The target then follows the rotation and translation of the hand, without snapping to the hand location. This technique enables the indirect manipulation of targets with hand movement. It can be seen as a 3D extension of existing approaches in 2D [39, 40, 66], which follow the underlying rationale of "gaze selects, hand manipulates".

4.3 3D Magic Gaze

3DMagicGaze establishes a circular safe region (10° radius, invisible to users) around the target once the initial eye-based selection is confirmed. If the gaze point is within the safe region, only the hand can control the transformation of the object. Otherwise, when the gaze point is outside of the safe region and if the hand movement distance exceeds a threshold (0.08m), the object snaps to the gaze point direction (without changing its depth to the user). A new safe region appears around the target after the snapping takes place. The design of this technique follows *MagicT* [59] in 2D, which requires an explicit command (hand movement) to switch from gaze input to manual input.

4.4 Implicit Gaze

ImplicitGaze also forms a circular safe region around the target once the eye-based selection is confirmed. If the gaze point is inside the safe region, hand input will control the object's transformation. Otherwise, if the gaze point is outside the safe region, the object snaps to the gaze point direction (without changing its depth to the user). A new safe region appears around the target after snapping. Unlike *3DMagicGaze*, this technique does not rely on any trigger mechanism to switch between gaze and hand input, thus features "implicit" transition between input modalities. To prevent the gaze cursor from being "over-active" [76], we introduced a dynamically-resized safe region, which resizes automatically based on the user's gaze behavior. It then increases its size from the original radius (6°) with a constant speed (10°/s) if the gaze point stays within the region until the maximum size (20°). This is to simulate users' search behavior, in which the longer the gaze point is fixed within specific regions, the more likely the user is approaching the target location [19, 39, 53]. Therefore, increasing the safe region's size can avoid unwanted snapping and allow robust, fine-grained hand translation.

The parameter values were obtained from our pilot tests. We kept the following design aspects consistent among the techniques: (1) the gaze pointer is invisible to users so as not to distract them; (2)

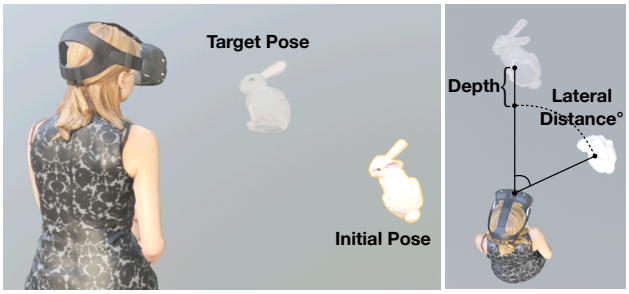


Figure 2: (left) Task illustration: participants were required to transform an object from its initial configuration to a target pose; (right) An illustration of Lateral Distance and Depth which were independent variables of the first study.

all techniques had the same control-display mapping (1:1) for hand manipulation; and (3) all techniques used the trigger button of the hand-held controller to confirm and release the selection.

Next, we present two user studies where we evaluated and compared the four gaze-supported manipulation techniques that employed different integration, coordination, and transition strategies. In the first study, we focused on the *primary working space*, where all objects located in front of the user and were within arm-reach distance, and assessed the techniques in terms of user performance and experience. In the second study, we further evaluated our techniques in a larger virtual environment with distant objects and embedded the designed techniques into realistic workflows.

5 STUDY 1: CONTROLLED EVALUATION

In this study, our goal was to evaluate and compare the four gaze-supported manipulation techniques (*GazeGrab*, *RemoteHand*, *3DMagicGaze*, and *ImplicitGaze*) that leveraged different design features from the presented design space in a controlled working space. By doing so, we aimed to better understand whether gaze input should be incorporated into hand manipulation process, and how gaze input could be combined with hand input for convenient and efficient 3D object manipulation in VR. The study mainly focused on the *primary working space*, where all targets of interest are located in front of the user (less than 90° horizontal offset when the user is looking forward) and are within arm-reach distance. Most of the work in VR is likely to happen within this area, so there is no need for users to frequently turn back or move around the virtual environment [1, 13, 72].

5.1 Participants and Apparatus

We recruited 12 university students (3 women, 9 men) between the age of 18 to 29 years (mean = 22.5) for this first study. All participants reported to be right-handed.

We developed the system using the Pico Neo 2 Eye, a standalone VR headset with 6 DOF tracking and Tobii eye-tracking features. The headset has 1920 × 2160 pixels screen resolution per eye and 101° field-of-view (FoV). The embedded eye tracker has 90Hz data output frequency, 0.5° estimated accuracy, and 25° left/right/down and 20° up trackable FoV. The software was implemented in C# in Unity3D.

5.2 Task

The task required participants to transform a 3D model from its initial configuration to a new target pose (see Figure 2 left). The target location was randomly selected within 30° of angle distance when the participant was looking straightforward along the z-axis (the depth axis) of his/her local space. The initial position was then calculated according to the target position based on our independent variables—lateral distance (the angular distance between the start and target location) and depth (the differences in the depth dimension). The target position was to be expected by participants. In other words, they knew where the object should be translated to when starting the manipulation task, even when the initial target was located outside the user's field-of-view (but within the primary working space). This allowed us to minimize the search time, which may confound with the manipulation time. Another factor, which is the object orientation, was adjusted according to the experiment requirement.

5.3 Evaluation Metrics

5.3.1 Performance Measures. To evaluate technique performance, we controlled transformation errors to be under a threshold (smaller than 0.015m and 3.5°) while comparing task completion time.

- *Manipulation Time*: the time elapsed between when object selection is confirmed and when both of the following conditions are satisfied: (1) the target is correctly placed with errors under the pre-determined threshold; and (2) the trigger is released.
- *Coarse Translation Time*: the time elapsed between the selection confirmation and the first time when the distance between the acquired object and target position is smaller than 0.05m. The rationale for including this variable was that, during our pilot studies, we found users took a long time to re-adjust the object orientation and fine-tune its position after reaching an approximate target location.
- *Re-position Time*: the elapsed time for fine-grained manipulation (= Manipulation Time - Coarse Translation Time).

5.3.2 Hand Manipulation Measures. We were also interested in investigating how techniques may influence hand movement and rotation for manipulation tasks, which may correlate to the arm fatigue measures, based on the simple rationale that more hand motion is likely to induce more arm fatigue [23].

- *Hand Movement Distance*: the accumulated distance (by accumulating the displacement of hand per frame) that the hand has travelled during the manipulation process.
- *Hand Rotation Angles*: the accumulated angle that the hand has rotated during the manipulation process.

5.3.3 Subjective Measures. We also compared the techniques based on subjective measures, including arm fatigue, ease of use, required workload, and individual rankings.

- *Borg CR10* [4, 29]: a categorical rating (0-10 points) which can be used to assess perceived arm exertion/fatigue. It has been shown to correlate well with objective measures from, for example, EMG data [58]. We adopted the same format and verbal description as previous works [29] in this experiment.
- *Single Easement Questionnaire* [46]: to measure the ease-of-use of the techniques with a 7-point scale.

- *Raw NASA-TLX* [21]: to measure the task load induced by the techniques with 7-point scales.
- *Subjective Ranking*: a rank of all the techniques according to participants' overall preference.

5.4 Design and Procedure

The study employed a $4 \times 3 \times 2$ within-subjects design with three independent variables: Technique (*RemoteHand*, *GazeGrab*, *ImplicitGaze*, and *3DMagicGaze*), Lateral Distance (35° and 55°), and Depth (0.05m, 0.10m, and 0.15m). Lateral distance represents the angular distance between the start and target location, whereas the depth factor looks at the differences in the depth dimension along the user's line of sight (see Figure 2 right). The current level and task setting made all objects to be located within the primary working space (from 0° to 85° horizontal offset and within arm-reach distance). The presentation order of Technique was counterbalanced using the Latin Square approach, whereas Lateral Distance and Depth were presented in random order. Additionally, the rotation factor (20°, 50°, 80°, 110°, and 140°), which is the required rotation (in angles) from the initial to the target transform, was pre-determined for each repetition and the same set of values was used across all conditions (though appeared with a randomized order). Exploring the effect of rotation was not our primary focus, as all techniques used a similar method to achieve that purpose. In the experiment, each condition was repeated 5 times which resulted in 1440 (= 12 participants × 4 techniques × 3 lateral distances × 2 depths × 5 repetitions) trials of data.

The whole experiment lasted approximately 50 minutes in total. Participants first completed a questionnaire to collect their demographic information. They were then introduced to the experiment task and the VR device, and instructed to complete the trials as fast and as accurately as possible. Next, we asked participants to put on the headset and start the experience in VR. The VR experience consisted of four sessions corresponding to four manipulation techniques. Each session began with ten warm-up trials for participants to get familiar with the input method, followed by the formal test trials. After each session, we collected user feedback with the Borg CR10, Single Easement, NASA-TLX, and Subjective Ranking questionnaires. Participants were required to have a rest between each session.

5.5 Results

To analyze the collected data, we first discarded the outliers that deviated more than three standard deviations from the mean value ($mean \pm 3std.$) in each condition (20 trials, 1.3%). Furthermore, a Shapiro-Wilk test indicated that the data is non-normally distributed. Therefore, all data underwent pre-processing through Aligned Rank Transform (ART) [71]. Next, we performed repeated-measures ANOVAs (RM-ANOVA) and Bonferroni-adjusted pairwise comparisons for each measurement. We also computed effect size (the non-parametric estimator for CL, symboled A_w [32, 64]) to accompany the pairwise tests based on unranked (non-normal) data. The results from performance measures, hand manipulation measures, and Borg CR10 are summarized in Figure 3.

5.5.1 Performance Measures. A RM-ANOVA indicated that Technique ($F_{3,253} = 4.141, p = .007$) and Lateral Distance ($F_{1,253} = 5.414, p = .021$) had significant main effects on Manipulation Time, but not Depth ($F_{2,253} = 0.186, p = .831$). No interaction between these variables was found. A post-hoc test indicated that *GazeGrab* (13.7*s*) was significantly slower ($p = .004$, $A_w = 0.64$) than *RemoteHand* (12.3*s*).

Another RM-ANOVA showed that both Technique ($F_{3,253} = 3.084, p = .030$) and Lateral Distance ($F_{1,253} = 25.024, p < .001$) had significant main effects on Coarse Manipulation Time, but not Depth ($F_{2,253} = 0.610, p = .544$). An interaction effect between Technique and Depth was also identified ($F_{6,253} = 3.396, p = .003$). When Depth increased, while *RemoteHand*, *ImplicitGaze*, and *3DMagicGaze* led to larger Coarse Manipulation Time, *GazeGrab* required less time. A post-hoc test indicated that *ImplicitGaze* (4.1*s*) was significantly faster ($p = .036$, $A_w = 0.61$) than *GazeGrab* (4.7*s*).

Finally, Technique ($F_{3,253} = 3.861, p = .010$) had a significant main effect on Re-position Time, but not Lateral Distance ($F_{1,253} = 1.377, p = .242$) or Depth ($F_{2,253} = 0.452, p = .637$). No interaction effects were found. According to a post-hoc test, *GazeGrab* (9.1*s*) was significantly slower ($p = .007$, $A_w = 0.64$) than *RemoteHand* (7.6*s*).

5.5.2 Hand Manipulation Measures. A RM-ANOVA showed that both Technique ($F_{3,253} = 13.559, p < .001$) and Lateral Distance ($F_{1,253} = 55.681, p < .001$) had significant main effects on Hand Movement Distance, but not Depth ($F_{2,253} = 0.126, p = .881$). No interaction effects were found. A post-hoc test indicated that *ImplicitGaze* required much smaller hand movement than *3DMagicGaze* ($p = .047$, $A_w = 0.38$), *GazeGrab* ($p < .001$, $A_w = 0.30$), and *RemoteHand* ($p < .001$, $A_w = 0.31$). Furthermore, *3DMagicGaze* required significantly less hand movement than *GazeGrab* ($p = .008$, $A_w = 0.38$).

Furthermore, Technique ($F_{3,253} = 15.663, p < .001$) and Lateral Distance ($F_{1,253} = 26.569, p < .001$) had significant main effects on Hand Rotation Angles, but not Depth ($F_{2,253} = 0.924, p = .398$). Additionally, no interaction effects between the variables were found. *ImplicitGaze* led to significantly less hand rotation than *3DMagicGaze* ($p = .003$, $A_w = 0.37$) and *RemoteHand* ($p = .008$, $A_w = 0.39$). Additionally, *GazeGrab* also resulted in significantly less hand rotation than *3DMagicGaze* ($p < .001$, $A_w = 0.29$) and *RemoteHand* ($p < .001$, $A_w = 0.30$).

5.5.3 Subjective Measures. A RM-ANOVA test indicated that *GazeGrab* induced more arm fatigue and higher physical workload than all other techniques (for all pairwise comparison, $p < .001$). It also led to higher (mental and physical) effort and created more frustration than *ImplicitGaze* and *RemoteHand* (all $p < .024$), and higher mental demand than *3DMagicGaze* ($p = .034$). Subjective ranking data indicated that participants significantly preferred *ImplicitGaze*, *RemoteHand*, and *3DMagicGaze* over *GazeGrab* (all $p < .001$). No other statistically significant effect was identified.

5.6 Discussion

5.6.1 Hand-Only vs. Eye-Hand Manipulation. When comparing the hand-only (during the manipulate phase) technique (*RemoteHand*) and the eye-hand techniques (*3DMagicGaze* and *ImplicitGaze*), we did not observe significant performance differences. This extends findings from previous research on 2D screens [59], where gaze was not be able to enhance the performance of manipulation tasks in the primary working space. On the other hand, adding transitions between gaze and hand also did not deteriorate performance compared to hand-only techniques; participants quickly learned/adapted to

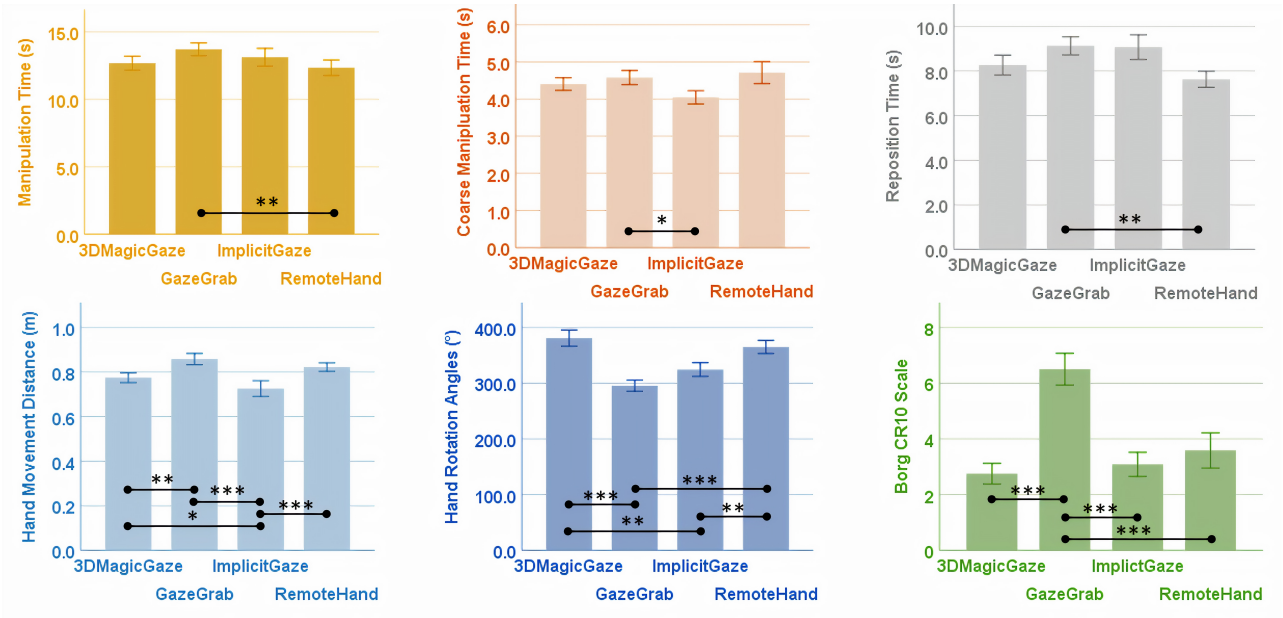


Figure 3: Plots of techniques' performance under different measurements. Error bars indicate the standard error. Statistical significant effects are marked (* = p < .05, ** = p < .01, and *** = p < .001).

these new input methods. As expected, *RemoteHand* required more hand movement and rotations to achieve the same manipulation task. However, the results from the Borg CR10 and NASA-TLX questionnaires did not show significant benefits of eye-hand transitions over hand-only techniques regarding arm fatigue and perceived workload.

5.6.2 Direct vs. Remote Hand Mappings. When comparing *GazeGrab* to other techniques that allowed remote hand mappings (specifically *RemoteHand*), we observed substantial differences in performance measures and subjective feedback. *GazeGrab* required a much longer time frame to re-position an object than *RemoteHand* and caused significantly higher perceived arm fatigue. This was mostly because, as indicated in previous research, "direct manipulation" techniques are imprecise in nature [36]. Participants found it difficult to place the object in the correct position by holding it with their arms. Further, participants suggested that *GazeGrab* was cumbersome as it required them to "suspend" their arms in the air to perform the manipulation (in contrast with the techniques based on indirect mapping which allowed them to manipulate the target with their arms down). Interestingly, *GazeGrab* induced less hand rotation than *RemoteHand*. In fact, according to the mean value shown in Figure 3, *GazeGrab* had the smallest hand rotation angles. This is likely due to the presence of direct mappings, which leads to users finding it easier to determine how to optimally rotate an object to the target configuration.

5.6.3 Implicit vs. Explicit Eye-Hand Transitions. When comparing *ImplicitGaze* and *3DMagicGaze*, we found that they led to similar empirical performance, while *ImplicitGaze* required less hand movement and rotation to complete the manipulation task. This difference was most likely due to the transition mechanism we chose for *3DMagicGaze*, which entailed the use of hand movement to snap the target to the hand position. Our choice was based on Turner et al. [59] work, where they thought such input structure would demonstrate "some form of integrity" (as we usually use a final hand manipulation to fine-grain the translation made by gaze input). However, according to our study results, we found this explicit hand movement can have side effects. It required participants' hands to move for longer periods of time and rotate more to achieve the same task compared to the implicit approach. Even if we change to other mode switching mechanisms, like trigger tapping [55], it is likely that such extra efforts would still be needed for methods based on explicit transitions. In contrast, implicit transition techniques can be an ideal solution as they require minimum effort for mode switching. Our results also showed no issues regarding unwanted snapping (in other words, not inducing the Midas touch problem [28]) by using a dynamically-resized safe region.

5.6.4 Effect of Lateral Distance and Depth. As expected, our results showed that lateral distance influenced the technique performance in coarse manipulation time, but not re-position time (mostly orientation adjustment). Depth did not have a definite impact on selection performance, likely due to the differences between the levels not being substantial (as all of them were within arm-reach distance).

5.6.5 Summary of Study 1's Key Findings. Based on the discussion, we summarize the following key findings from the first study.

- Our results show no evidence that manipulating objects (mainly translation) based on both eye and hand input (*3DMagicGaze* and *ImplicitGaze*) can offer significant performance benefits in VR manipulation tasks over the hand-only approach (*RemoteHand*) in the primary working space.
- Direct hand mapping (*GazeGrab*) is less precise and can lead to more arm fatigue than remote hand mappings (like *RemoteHand*).

However, it might help users to determine how to optimally rotate an object to the target configuration.
- Implicit transition (*ImplicitGaze*) and explicit transition (*3DMagicGaze*) led to similar task performance, while implicit transition required less effort (e.g., hand movement) than explicit transition. In particular, a dynamically-resized safe region was shown to be useful as there was little evidence of the Midas touch issue [28].

After assessing technical performance and initial user feedback in Study 1, we further extended the evaluation to a larger space which requires the use of locomotions in Study 2.

6 STUDY 2: APPLICATION

In this second study, we aimed to assess how gaze-supported manipulation techniques perform under a larger environment and when applied to realistic workflows. We also wanted to compare our techniques with Virtual Hand (hand input only for selection and manipulation), which is currently the most common method for manipulating objects. We also measured user experience and collected user feedback, which can help adapt the gaze-supported techniques to real use cases.

6.1 Participants and Apparatus

We recruited eight university students (3 women, 5 men) with previous experience in 3D modeling (1-8 years, mean = 2.75, using software like SolidWorks, 3DS MAX, CAD, Rhino, and Unity). We hope that more fruitful discussions could be triggered with experienced/expert users in the relevant domain. Their ages were between 21-29 years (mean = 24.4). All of them were right-handed. We used the same device as in the previous study.

6.2 Interaction Scenario

Participants were instructed to reconstruct an empty room following a miniature, as shown in Figure 4, using the manipulation techniques. However, they were not required to follow how the miniature looked like precisely; rather, it was used as a guide for them to make their own creations. Participants could move around the room using the teleportation mechanism, choose desired objects from a prefab list (see Figure 4), and manipulate (translate, rotate, and scale) the selected item. This differed from the first study, which controlled the participants in a static position (within primary working space) and had specific time-controlled task requirements. In this interaction scenario, we emphasized the "design-by-yourself" concept, where the techniques were integrated into users' own workflow and creative experiences [75]. Similar applications include Mozilla Hubs [25] or Minecraft VR [38], where users/players can decorate/build virtual space with different objects/building blocks.

6.3 Procedure

The whole experiment lasted approximately 60 minutes in total. Participants first completed a demographic questionnaire. Then, participants were briefed about the task and program functionalities, and were asked to put on the headset on and started interacting with the virtual space. The whole interaction experience was divided into five sessions (four gaze-supported techniques and virtual hand were presented in a randomized order). During each session, they learned about a manipulation technique and performed the task as described in the previous section. At the end of each session, they completed a short version of the User Experience Questionnaire (UEQ-S) [47] and answered a set of structured questions to provide their overall feedback towards the technique. The structured questions asked about the strengths and weaknesses of each method. After finishing the five sessions, they were also invited to provide their opinions regarding the different design features employed in the techniques (hand-only *vs.* hand-eye, direct *vs.* remote mappings, and implicit *vs.* explicit transitions). Responses were recorded for further analysis.

Figure 4: Participants were instructed to construct an empty room following a miniature (left) with the gaze-supported manipulation techniques. They were able to teleport around the room, select objects from a prefab list (right), and manipulate (translate, rotate, and scale) the selected item.

Technique	Pragmatic	Hedonic	Overall
GazeGrab	0.66	0.84	0.75
RemoteHand	0.21	-0.25	-0.02
3DMagicGaze	0.31	0.94	0.63
ImplicitGaze	0.68	1.10	0.89
Virtual Hand	-0.13	-0.63	-0.38

Table 2: The results from the short version of User Experience Questionnaires (UEQ-S) which outline the pragmatic quality, hedonic quality, and overall quality of each Technique (higher scores are better).

6.4 Results

The results from UEQ-S are summarized in Table 2, which indicates that the gaze-supported techniques performed better comparing to Virtual Hand in terms of pragmatic, hedonic, and overall quality. Next, we provide a summary of participant interview responses grouped by technique.

6.4.1 Gaze Grab. As a way of hand-eye coordination, *GazeGrab* has a unique feature of snapping the object to the hand position when starting the manipulate phase. A number of participants (N=5) commented that it was "*efficient*" and "*convenient*" way of achieving this; "*I normally moved to the destination first, and then brought the object to me with the technique. It was very quick.*" (P2). However, a couple of participants mentioned that "*the efficiency of GazeGrab was highly dependent on the accuracy of teleportation method, which was sometimes not very accurate.*" (P3). The inaccurate teleportation might require users to re-adjust their standing position when using *GazeGrab*. Two participants also said that the technique "*required some learning*". Notably, P5 noticed that "*when object flew to me,*

especially big objects like a sofa, I was afraid that it might hit me.", and P5 also found it challenging to fine-grain the position of an object as "*the object would fly to my hand again when pressing the trigger, and my previous effort was wasted*".

6.4.2 Remote Hand. Although this technique has the ability to manipulate objects remotely, almost all participants (N=7) noted that *RemoteHand* was inconvenient when moving objects that were at a far distance; "*It seemed that the object only moved a little bit when I moved my arms.*" (P2). "*This was fatiguing.*" (P1). Moreover, P5 mentioned that "*when I tried to move the object for a large distance, my arm's movement might also cause the rotation of the object. So I had to rotate it back.*" Despite these limitations, most participants (N=7) felt *RemoteHand* was accurate for manipulation. In addition, P5 commented on the agency provided by the technique "*manipulating objects remotely made me feel that I was taking control of the whole space*".

6.4.3 3D Magic Gaze. Half of the participants (N=4) explicitly mentioned that, with the help of their eyes, *3DMagicGaze* was quick for long-distance object translation. However, a few participants (N=5) mentioned some flaws in the hand confirmation mechanism: "*I needed time to get used to this (hand movement for confirmation).*" (P6) "*For small or medium movement, it was sometimes hard for me to decide whether using hand or gaze.*" (P5). Additionally, some participants (N=5) thought the switching between eye and hand input was confusing at times: "*I often forgot using hand to bring (snap) the object.*" (P7) "*I found sometimes waving my arms did not make the quick transformation (snap). For example, I wanted to put a bed adjacent to the wall, but it was hard to achieve—the movement was either too small or too large*" (P4). The later was because the gaze cursor was still inside the safe region, so the hand snapping did not happen. In contrast, P8 said that "*I did not feel any big difference comparing to ImplicitGaze.*" and indicated that hand movement was natural for confirmation. P6 further said *3DMagicGaze* felt more "*stable*" than *ImplicitGaze*, since the selected object would not frequently snap to the gaze direction.

6.4.4 Implicit Gaze. Participants (N=7) felt that *ImplicitGaze* was "*novel*" and "*efficient*"; "*I can just stand still and manipulate the objects quickly.*" (P4). However, several participants (N=4) also commented about the difficulty of using eyes to achieve precise manipulation. "*When I was searching for the places, the object, especially the big ones, would block my view. Also, there were some unwanted movements caused by eyes.*" (P3). On the positive side, P7 commented that "*I thought eye movement might cause some random movements before using it, but it actually didn't when trying.*" Noticeably, some participants (N=3) thought it was not as easy to move the object in the depth dimension with *ImplicitGaze*, as the movement in that dimension is particularly slower than lateral directions.

6.4.5 Virtual Hand. Almost all participants (N=7) thought *VirtualHand* was natural and realistic; "*I always know how to do it (the manipulation), as that's what we do in everyday life.*" (P3). The technique also felt more "*controllable*" due to these characteristics. However, all participants (N=8) acknowledged that *VirtualHand* was "*fatiguing*" and "*not efficient enough for long-distance translation*".

6.5 Discussion

In this section, we discuss and summarize the results and provide solutions for the identified limitations and design implications that can help future implementation of gaze-supported manipulation techniques in VR.

6.5.1 Hand-Only vs. Eye-Hand Manipulation. While the benefit of rapid eye movement for object translation is not salient in the primary working space (as shown in the first study), for manipulating faraway objects in a larger environment, participants clearly preferred the efficiency and convenience of gaze-hand combination for coarse translation. Indeed, theoretically, an exact control-display mapping (1:1) of hand movement has little effect (visually) on objects located in a far distance from a user's perspective. In such a situation, it is thus more ideal for translating the target according to visual angles (as what gaze input does), rather than exact distance mapping (as what hand input does in this research). Another solution, which can enhance hand-only approaches (e.g., *RemoteHand*) in the manipulate phase is to provide hand amplification (e.g., [69]), where the hand movement is amplified using specific functions, so the object appears to move a larger distance.

However, eye-hand manipulation became less useful for close and large objects, as it might occlude the user's line-of-sight (since the target follows gaze), which made location searching difficult. A quick fix could entail making the target under manipulation semi-transparent [11], so that the user's view is not fully-blocked. Some participants also found hand-only manipulation to be more manageable, as they reported being more used to this type of input.

6.5.2 Direct vs. Remote Hand Mappings. With the feature of bringing faraway objects to users' hands (turns a remote object to direct hand mapping), *GazeGrab* shifted how participants interacted with objects when compared to the other three gaze-supported techniques. With remote-mapping based methods like *ImplicitGaze*, participants tended to remain in the same standing position and transferred the items remotely. In contrast, with *GazeGrab*, they were likely to first move to a new target position and then bring the object to their location. As reported by the participants, this transformation was efficient in transporting distant targets but can be cumbersome for close ones. Repetitive snapping close objects to hands can make the adjustment difficult, and it is likely better to disable this function when the target is within arm-reach distance. Furthermore, users needed to re-adjust their standing position if there was any inaccuracy caused by the locomotion technique. If the VR locomotion/teleportation [3] is sufficiently smooth, efficient, and accurate, the snap-to-hand function could be useful by translating distant objects along the depth dimension.

Another issue brought by direct hand mappings is that if the object under manipulation is quite large, participants found it difficult to transform the object into a satisfiable configuration as a significant part of their view is occupied by the item. Some participants also reported that it made them feel unsafe as they thought the object might collide with their body. Potential solutions to these issues could entail providing a mini-map [12, 57] as an overlay to give an non-occluded vision to support fine-grained transformation and making the oversized object semi-transparent to minimize its intrusiveness.

6.5.3 Implicit vs. Explicit Eye-Hand Transitions. Participants' opinions differed in whether it would be more beneficial to apply explicit or implicit transitions between eye and hand input. The advocates of the explicit transition mechanism most appreciated its robustness; the rapid eye movement would not frequently bring the object to the user's facing direction. Although the dynamically-resized safe region was reported as being useful (*ImplicitGaze* did not produce random gaze-like movement for objects), it was not able to handle rapid, long-distance searching actions, and could occlude participants' view by snapping the target to the gaze location. As mentioned previously, making the target semi-transparent would mitigate this issue.

On the other hand, some participants found that using hand movement to confirm the gaze action was somewhat redundant. Moreover, because of the separate nature of gaze and hand input, participants noticed that it was sometimes challenging to determine whether they were using hand or gaze input. Additionally, it can be confusing for users when they actually want to use gaze to translate an object, but because the gaze point is still located inside the safe region, only the hand movement (which was meant to be a trigger action) affected objects' location. In these scenarios, it would be helpful to provide a small widget to indicate which input modality is taking control of the object for explicit transition based techniques.

Also, as suggested by participants, it would be beneficial to provide hand amplification [69] for both *ImplicitGaze* and *3DMagicGaze* in the depth dimension to speed up the translation along the z-axis.

6.5.4 Gaze-Supported Techniques vs. Virtual Hand. As indicated in Table 2, the results from the user experiences questionnaire suggest that the current market-available solution (Virtual Hand) was not sufficient for target manipulation tasks in VR, while gaze-supported techniques lead to pragmatic and hedonic improvements. Despite being "natural" and "realistic", Virtual Hand was seen as not being an efficient, convenient, and comfortable solution for long-term object manipulation in VR.

7 DESIGN IMPLICATIONS

We derived a set of design implications for future gaze-supported manipulation techniques in VR. We do not advocate a one-size-fits-all technique, as different design features can be useful for different environments and task proposes. Instead, we summarize their strengths, possible applications, and provide potential compensation for their weaknesses.

- While embedding gaze input (like *ImplicitGaze* and *3DMagicGaze*) might not offer significant performance benefits for manipulating (translating) objects that are within the primary working space (that is, all targets are located in front of the user and within arm-reach distance), it can be useful for a larger environment with distant objects.
- If gaze input is used for object selection and only hand input is used for manipulation, consider adding hand amplification (e.g., [69]) when users need to manipulate objects that are outside of the primary working space. Otherwise, it can feel tiresome to manipulate remotely gaze-selected objects.
- The hand-eye coordination strategy which snaps the target to the hand position when selection is triggered is efficient for bringing distant objects to the user. However, this function may require the user to teleport to different places frequently when working in a large environment. Therefore, a complementary precise and convenient teleportation mechanism is needed. Additionally, we suggest disabling the snap-to-hand function for objects within arm-reach distance, as repetitive snapping close objects to hands can cause confusion and make the fine-grained adjustment difficult.
- While manipulating an object directly via hands is intuitive, it may lead to more arm fatigue as users need to hold their arms in the air. One could consider minimizing the duration of using such direct-mapping and use indirect-mapping techniques (like *RemoteHand*) which allow users to rest their arms under a comfortable position. Also, large objects can easily occlude users' view and pose difficulties for accurate manipulation. Therefore, providing an accompanying mini-map (e.g., [57, 73, 74]) as an overlay would provide an overview of the environment, while making the oversized object transparent to reduce intrusiveness.
- Providing an implicit transition between gaze and hand input (such as *ImplicitGaze*) can enable the smooth and concurrent transformation. It would be useful to consider applying a dynamically-resized safe region (as used in this research) to reduce the random movement of objects caused by eye saccades. Note there are also other design opportunities to enable implicit transitions. For example, designers may choose to use a probabilistic/heuristic model to implicitly determine whether gaze or hand should take control of the target. Also, we suggest making the object under manipulation semi-transparent to avoid visual occlusion while searching.
- Explicit transition (like *3DMagicGaze*) enables robust control over the effect of gaze on objects. However, some effort is required in performing the 'switch' command and users may be unsure about whether to make a 'switch' or not. We recommend adding a small widget to indicate which input modality is currently taking control of the manipulation.
- For techniques that use both gaze and hand for manipulation (e.g., *ImplicitGaze* and *3DMagicGaze*), hand amplification in the depth dimension would be beneficial to speed up the translation along the z-axis when interacting with objects outside of the primary working space.

8 LIMITATIONS AND FUTURE WORK

We have identified several limitations in this research. First, we did not embed techniques that enable non-linear mapping of hand input [69], as our primary focus was gaze input. Hand amplification can interplay with or enhance gaze input, and it would be interesting to investigate how they influence one another. Second, we did not explore the long-term usage of gaze-supported manipulation techniques. For instance, if 3D modelers used gaze input every day, they would probably find even more efficient ways of using them. Third, we did not test the methods alongside more complex sculpturing and modeling tools/functions (like smoothing and inflating an object). Further research can extend the gaze input modality to accompany more advanced manipulation functions. Fourth, we treated gaze as a 2 DoF modality and thus explored more of its usage for translating objects in the lateral direction. However, we acknowledge that there is a potential of using gaze for rotation and scaling with novel approaches. Lastly, as head gaze can be a cheaper solution than eye gaze for current VR systems, it is worth exploring if head gaze possesses similar features as eye gaze for object manipulation.

9 CONCLUSION

In this research, we explore gaze-supported 3D object manipulation in VR. Specifically, we investigate how different ways of integrating, coordinating, and transitioning gaze and hand input can aid the existing approach based on the virtual hand. Results from two user studies evaluating and comparing four techniques regarding their usability and efficiency show that gaze input does not offer significant performance benefits for object manipulation in the primary working space (when all targets are located in front of the user and within arm-reach distance), but can be useful for larger spaces with distant objects. Gaze input was also shown to mitigate the arm fatigue issue, and different integration, coordination, and transition strategies can provide benefits for building more usable and efficient object manipulation techniques. Our work contributes novel insights regarding multimodal interfaces with gaze and hand input that can enhance existing and future 3D object manipulation solutions in VR.

ACKNOWLEDGMENTS

This research is partially funded by the Melbourne Research Scholarship provided by The University of Melbourne. It is also supported in part by Xi'an Jiaotong-Liverpool University (XJTLU) Key Special Fund (KSF-A-03). Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Award (Project Number: DE180100315) funded by the Australian Government.

REFERENCES

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

[2] Benjamin Bach, Ronell Sicat, Johanna Beyer, Maxime Cordeil, and Hanspeter Pfister. 2017. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 457–467. https://doi.org/10.1109/TVCG.2017.2745941

[3] Costas Boletsis. 2017. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction* 1, 4 (2017), 24. https://doi.org/10.3390/mti1040024

[4] Gunnar AV Borg. 1982. Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise* (1982).

[5] Doug A. Bowman and Larry F. Hodges. 1997. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 35–ff. https://doi.org/10.1145/253284.253301

[6] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. 2001. Testbed Evaluation of Virtual Environment Interaction Techniques. *Presence: Teleoperators and Virtual Environments* 10, 1 (2001), 75–95. https://doi.org/10.1162/105474601750182333

[7] Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. 2012. Questioning Naturalism in 3D User Interfaces. *Commun. ACM* 55, 9 (Sept. 2012), 78–88. https://doi.org/10.1145/2330667.2330687

[8] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. 2019. Investigating Smartphone-Based Pan and Zoom in 3D Data Spaces in Augmented Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 13 pages. https://doi.org/10.1145/3338286.3340113

[9] Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2818346.2820752

[10] Shujie Deng, Nan Jiang, Jian Chang, Shihui Guo, and Jian J Zhang. 2017. Understanding the impact of multimodal interaction using gaze informed mid-air gesture control in 3D virtual objects manipulation. *International Journal of Human-Computer Studies* 105 (2017), 68–80. https://doi.org/10.1016/j.ijhcs.2017.04.002

[11] Joachim Diepstraten, Daniel Weiskopf, and Thomas Ertl. 2002. Transparency in interactive technical illustrations. In *Computer Graphics Forum*, Vol. 21. Wiley Online Library, 317–325. https://doi.org/10.1111/1467-8659.t01-1-00591

[12] Niklas Elmqvist and Philippas Tsigas. 2007. A taxonomy of 3D occlusion management techniques. In *2007 IEEE Virtual Reality Conference*. IEEE, 51–58. https://doi.org/10.1109/VR.2007.352463

[13] Barrett M. Ens, Rory Finnegan, and Pourang P. Irani. 2014. The Personal Cockpit: A Spatial Interface for Effective Task Switching on Head-Worn Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3171–3180. https://doi.org/10.1145/2556288.2557058

[14] Facebook. [n.d.]. *Oculus Medium.* Retrieved September 6, 2020 from https://www.oculus.com/medium/

[15] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1118–1130. https://doi.org/10.1145/3025453.3025599

[16] Tiare Feuchtner and Jörg Müller. 2018. Ownershift: Facilitating Overhead Interaction in Virtual Reality with an Ownership-Preserving Hand Space Shift. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 31–43. https://doi.org/10.1145/3242587.3242594

[17] S. Frees and G. D. Kessler. 2005. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* 99–106. https://doi.org/10.1109/VR.2005.1492759

[18] Epic Games. [n.d.]. *Unreal Engine VR Mode.* Retrieved September 6, 2020 from https://docs.unrealengine.com/en-US/Engine/Editor/VR/index.html/

[19] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. 2019. Learning Cooperative Personalized Policies from Gaze Data. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 197–208. https://doi.org/10.1145/3332165.3347933

[20] Google. [n.d.]. *Introducing Blocks.* Retrieved September 6, 2020 from https://arvr.google.com/blocks/

[21] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908. https://doi.org/10.1177/154193120605000909

[22] Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. Plane, Ray, and Point: Enabling Precise Spatial Manipulations with Shape Constraints. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1185–1195. https://doi.org/10.1145/3332165.3347916

[23] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[24] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 625, 12 pages. https://doi.org/10.1145/3290605.3300855

[25] Mozilla Hubs. [n.d.]. *Hubs - Private social VR in your web browser.* Retrieved Augest 19, 2020 from https://hubs.mozilla.com/

[26] Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze Gestures or Dwell-Based Interaction?. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/2168556.2168602

[27] VRChat Inc. [n.d.]. *VRChat.* Retrieved September 6, 2020 from https://www.vrchat.com/

[28] Robert J. K. Jacob. 1990. 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* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[29] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3328–3339. https://doi.org/10.1145/3025453.3025523

[30] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article 81, 14 pages.

https://doi.org/10.1145/3173574.3173655

[31] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.

[32] Johnson Ching-Hong Li. 2016. Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data. *Behavior research methods* 48, 4 (2016), 1560–1574. https://doi.org/10.3758/s13428-015-0667-z

[33] Zhenxing Li, Deepak Akkil, and Roope Raisamo. 2019. Gaze Augmented Hand-Based Kinesthetic Interaction: What You See is What You Feel. *IEEE transactions on haptics* 12, 2 (2019), 114–127. https://doi.org/10.1109/TOH.2019.2896027

[34] Chang Liu, Alexander Plopski, and Jason Orlosky. 2020. OrthoGaze: Gaze-based Three-dimensional Object Manipulation using Orthogonal Planes. *Computers & Graphics* (2020). https://doi.org/10.1016/j.cag.2020.04.005

[35] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-Air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 63–71. https://doi.org/10.1145/2807442.2807489

[36] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. 2019. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum* 38, 1 (2019), 21–45. https://doi.org/10.1111/cgf.13390

[37] Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. 2016. The Benefits of DOF Separation in Mid-Air 3D Object Manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/2993369.2993396

[38] Minecraft. [n.d.]. *Minecraft Official Site*. Retrieved Augest 19, 2020 from https://www.minecraft.net/

[39] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-Touch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[40] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[41] Jeffrey S. Pierce, Brian C. Stearns, and Randy Pausch. 1999. Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments. In *Proceedings of the 1999 Symposium on Interactive 3D Graphics* (Atlanta, Georgia, USA) *(I3D '99)*. Association for Computing Machinery, New York, NY, USA, 141–145. https://doi.org/10.1145/300523.300540

[42] Matti Pouke, Antti Karhu, Seamus Hickey, and Leena Arhippainen. 2012. Gaze Tracking and Non-Touch Gesture Based Interaction Method for Mobile 3D Virtual Spaces. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (Melbourne, Australia) *(OzCHI '12)*. Association for Computing Machinery, New York, NY, USA, 505–512. https://doi.org/10.1145/2414536.2414614

[43] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 79–80. https://doi.org/10.1145/237091.237102

[44] Warren Robinett and Richard Holloway. 1992. Implementation of Flying, Scaling and Grabbing in Virtual Worlds. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics* (Cambridge, Massachusetts, USA) *(I3D '92)*. Association for Computing Machinery, New York, NY, USA, 189–192. https://doi.org/10.1145/147156.147201

[45] Kunhee Ryu, Joong-Jae Lee, and Jung-Min Park. 2019. GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection. *Journal on Multimodal User Interfaces* 13, 4 (2019), 383–393. https://doi.org/10.1007/s12193-019-00305-y

[46] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of Three One-Question, Post-Task Usability Questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1599–1608. https://doi.org/10.1145/1518701.1518946

[47] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *IJIMAI* 4, 6 (2017), 103–108. https://doi.org/10.9781/ijimai.2017.09.001

[48] Marcos Serrano, Barrett Ens, Xing-Dong Yang, and Pourang Irani. 2015. Gluey: Developing a Head-Worn Display Interface to Unify the Interaction Experience in Distributed Display Environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 161–171. https://doi.org/10.1145/2785830.2785838

[49] Ludwig Sidenmark and Hans Gellersen. 2019. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (Dec. 2019), 40 pages. https://doi.org/10.1145/3361218

[50] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

[51] Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. 2020. BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/3379155.3391312

[52] Adalberto L. Simeone, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2016. Three-Point Interaction: Combining Bi-Manual Direct Touch with Gaze. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (Bari, Italy) *(AVI '16)*. Association for Computing Machinery, New York, NY, USA, 168–175. https://doi.org/10.1145/2909132.2909251

[53] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How do people explore virtual environments? *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1633–1642. https://doi.org/10.1109/TVCG.2018.2793599

[54] Dana Slambekova, Reynold Bailey, and Joe Geigel. 2012. Gaze and Gesture Based Object Manipulation in Virtual Worlds. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology* (Toronto, Ontario, Canada) *(VRST '12)*. Association for Computing Machinery, New York, NY, USA, 203–204. https://doi.org/10.1145/2407336.2407380

[55] Sophie Stellmach and Raimund Dachselt. 2013. Still Looking: Investigating Seamless Gaze-Supported Selection, Positioning, and Manipulation of Distant Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/2470654.2470695

[56] Sophie Stellmach, Sebastian Stober, Andreas Nürnberger, and Raimund Dachselt. 2011. Designing Gaze-Supported Multimodal Interactions for the Exploration of Large Image Collections. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications* (Karlskrona, Sweden) *(NGCA '11)*. Association for Computing Machinery, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/1983302.1983303

[57] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 265–272. https://doi.org/10.1145/223904.223938

[58] Amedeo Troiano, Francesco Naddeo, Erik Sosso, Gianfranco Camarota, Roberto Merletti, and Luca Mesin. 2008. Assessment of force and fatigue in isometric contractions of the upper trapezius muscle by surface EMG signal and perceived exertion scale. *Gait & Posture* 28, 2 (2008), 179 – 186. https://doi.org/10.1016/j.gaitpost.2008.04.002

[59] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4179–4188. https://doi.org/10.1145/2702123.2702355

[60] Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. 2013. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *IFIP Conference on Human-Computer Interaction*. Springer, 170–186. https://doi.org/10.1007/978-3-642-40480-1_11

[61] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2013. Eye Drop: An Interaction Concept for Gaze-Supported Point-to-Point Content Transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) *(MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 37, 4 pages. https://doi.org/10.1145/2541831.2541868

[62] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-Device Gaze-Supported Point-to-Point Content Transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2578153.2578155

[63] Jayson Turner, Andreas Bulling, and Hans Gellersen. 2011. Combining Gaze with Manual Interaction to Extend Physical Reach. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction* (Beijing, China) *(PETMEI '11)*. Association for Computing Machinery, New York, NY, USA, 33–36. https://doi.org/10.1145/2029956.2029966

[64] András Vargha and Harold D Delaney. 2000. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics* 25, 2 (2000), 101–132. https://doi.org/10.3102/10769986025002101

[65] Manuel Veit, Antonio Capobianco, and Dominique Bechmann. 2009. Influence of Degrees of Freedom's Manipulation on Performances during Orientation Tasks in Virtual Reality Environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (Kyoto, Japan) *(VRST*

'09). Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/1643928.1643942

[66] Eduardo Velloso, Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. In *Human-Computer Interaction – INTERACT 2015*. Springer International Publishing, Cham, 315–330. https://doi.org/10.1007/978-3-319-22668-2_25

[67] Simon Voelker, Sebastian Hueber, Christian Holz, Christian Remy, and Nicolai Marquardt. 2020. GazeConduits: Calibration-Free Cross-Device Collaboration through Gaze and Touch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376578

[68] Tobii VR. [n.d.]. *Hand-Eye Coordination*. Retrieved Augest 12, 2020 from https://vr.tobii.com/sdk/develop/unity/unity-examples/hand-eye-coordination/

[69] Johann Wentzel, Greg d'Eon, and Daniel Vogel. 2020. Improving Virtual Reality Ergonomics Through Reach-Bounded Non-Linear Input Amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376687

[70] Curtis Wilkes and Doug A. Bowman. 2008. Advantages of Velocity-Based Scaling for Distant 3D Manipulation. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology* (Bordeaux, France) *(VRST '08)*. Association for Computing Machinery, New York, NY, USA, 23–29. https://doi.org/10.1145/1450579.1450585

[71] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[72] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173616

[73] Difeng Yu, Hai-Ning Liang, Kaixuan Fan, Heng Zhang, Charles Fleming, and Konstantinos Papangelis. 2020. 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 (2020), 2762–2774. https://doi.org/10.1109/TVCG.2019.2905580

[74] Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Fully-Occluded Target Selection in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3402–3413. https://doi.org/10.1109/TVCG.2020.3023606

[75] Difeng Yu, Qiushi Zhou, Benjamin Tag, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Engaging Participants during Selection Studies in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 500–509. https://doi.org/10.1109/VR46266.2020.00071

[76] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

Chapter 6

BLENDING ON-BODY AND MID-AIR INTERACTION

6.1 Summary

In this work, we propose design patterns and interaction techniques that leverage combined on-body and mid-air interfaces for object selection and manipulation in VR. The on-body space (i.e., body surfaces) offers new interaction possibilities: it is always available, allows eyes-free targeting, and provides a support surface for input. Like gaze, on-body space has great potential to complement and augment Raycasting and Virtual Hand. With our designs, a user may use thumb-on-finger gestures, finger-on-arm gestures, or on-body displays with mid-air input to complete a 3D interaction task. We probed into the design space by developing various techniques for different selection and manipulation tasks (e.g., occluded selection, group selection, and one degree-of-freedom transformation) and conducted an expert evaluation study to elicit immediate design issues with the novel combination.

The study results and our implementations demonstrated that the proposed solution could be used for small, distant, and occluded target selection. The techniques could enable faster and more precise manipulation by changing the control-display ratio and isolating the transformations. They also created novel and fulfilling user experiences by empowering a multitude of helpful functionalities for selecting and manipulating objects in a sample 3D modeling application.

Env.			**Task**				
Small	*Distant*	*Occluded*	*Effectiveness*	*Efficiency*	*Ergonomics*	*Experience*	*Expressivity*
✓	✓	✓	✓	✓		✓	✓

6.2 Article III

Difeng Yu, Qiushi Zhou, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. "Blending On-Body and Mid-Air Interaction in Virtual Reality." In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 637-646. IEEE, 2022. https://doi.org/10.1109/ISMAR55827.2022.00081

Blending On-Body and Mid-Air Interaction in Virtual Reality

Difeng Yu* Qiushi Zhou† Tilman Dingler‡ Eduardo Velloso§ Jorge Goncalves¶

University of Melbourne, Melbourne, Australia

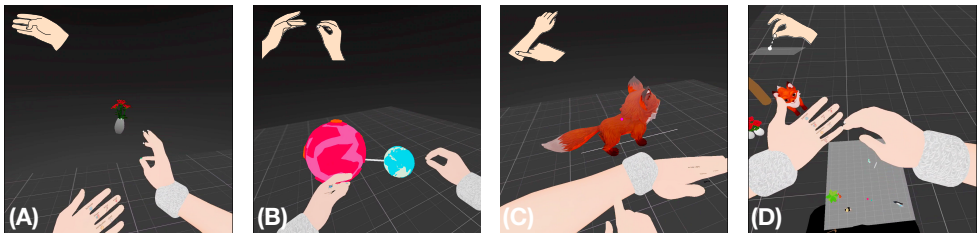


Figure 1: Sample interaction techniques based on BodyOn. (A) A user is scaling a vase towards a specific direction by performing thumb-on-finger gestures and mid-air movements. (B) A user is rotating a blue planet around and/or moving it towards a red planet by combining bimanual thumb-on-finger gestures with mid-air input. (C) Finger-on-arm gestures and mid-air input enable users to translate a fox with one degree of freedom. (D) Users can teleport to different locations by manipulating an on-body minimap display.

Abstract

On-body interfaces, which leverage the human body's surface as an input or output platform, can provide new opportunities for designing VR interaction. However, it remains unclear how on-body interfaces can best support current VR systems that mainly rely on mid-air interaction. We propose BodyOn, a collection of six design patterns that leverage combined on-body and mid-air interfaces to achieve more effective 3D interaction. Specifically, a user may use thumb-on-finger gestures, finger-on-arm gestures, or on-body displays with mid-air input, including hand movement and orientation, to complete an interaction task. To test our design concepts, we implemented example interaction techniques based on BodyOn that can assist users in various 3D interaction tasks. We further conducted an expert evaluation using the techniques as probes to elicit immediate design issues that emerge from the novel combination of on-body and mid-air interaction. We provide insights that can inspire and inform the design of future 3D user interfaces.

Index Terms: Human-centered computing—Human Computer Interaction (HCI)—Interaction Paradigms—Virtual Reality;

1 Introduction

Virtual reality (VR) technologies, or immersive technologies in general, represent a significant paradigm shift from the traditional PC-based interaction by putting users "into" the digital content. Whereas a large number of VR techniques enable users to interact with content located within a virtual environment through mid-air input (like hand movement or orientation) [1, 41], interfaces that leverage users' *on-body spaces*—the virtual representation of the human body's surfaces—are often overlooked.

The on-body space offers new interaction possibilities for VR systems: it is always available [28,29], allows eyes-free targeting [26,53], and provides a supporting surface for input [25, 28]. However, the design space of how on-body interaction can be incorporated into current mid-air interaction workflows in VR systems is largely underexplored [8]. While on-body interfaces can be appealing, they cannot fully replace the current paradigm based on mid-air interaction. For instance, mid-air techniques are more appropriate than on-body ones to enable 3D translation and movement of objects in VR. Therefore, it is critical to explore the synergies across these input modalities to best leverage their strengths and overcome their limitations.

To explore this opportunity, we propose *BodyOn*, a design space consisting of six design patterns for integrating on-body interfaces into current mid-air interaction workflows in VR headsets (Figure 1). In contrast to previous work that considered the on-body space as a standalone input and output modality [5, 7, 22], BodyOn takes a unique perspective by combining both on-body and mid-air interfaces to expand the design space of VR interaction techniques. Within this design space, a user may use thumb-on-finger gestures, finger-on-arm gestures, or on-body display in combination with mid-air input, including hand movement and orientation, to accomplish various VR interaction tasks (see Figure 1 for examples).

We instantiate this design space through a set of example interaction techniques based on BodyOn to accomplish canonical interaction tasks in a 3D modelling system, including selection, manipulation, navigation, and system control (e.g., menu control and mode switching). These techniques served as probes to showcase possible designs with BodyOn, and allowed us to form a testbed to verify the feasibility and applicability of the high-level design concepts. We then conducted an expert evaluation to gather feedback about the implemented interaction techniques. The study allowed us to identify immediate design issues with the new combination of on-body and mid-air interactions. For example, we found that when users focus on manipulating objects in the mid-air space, they can ignore on-body visual feedback. We discuss the lessons learned from our experience regarding future systems that may benefit from BodyOn.

The main contributions of our work are:

- BodyOn: a collection of six design patterns for inspiring new 3D UI designs that combine on-body and mid-air interactions in immersive VR space.
- Example interaction techniques to explore the design space and showcase how to solve 3D interaction tasks at various complexity levels with BodyOn.
- Insights based on an expert evaluation for future systems that leverage both on-body and mid-air interactions.

*e-mail: difeng.yu@student.unimelb.edu.au
†e-mail: qiushi.zhou@unimelb.edu.au
‡e-mail: tilman.dingler@unimelb.edu.au
§e-mail: eduardo.velloso@unimelb.edu.au
¶e-mail: jorge.goncalves@unimelb.edu.au

2 Related Work

BodyOn enhances current mid-air interaction techniques in VR systems by incorporating on-body interaction.

2.1 Mid-Air Interaction

Mid-air interaction is the most common form of interaction in contemporary headset-based VR systems. It allows users to control and manipulate digital content in VR through mid-air gestures and movements, typically using game controllers or bare hands [14, 35, 49]. Previous research has identified mid-air interaction as being natural, straightforward, and particularly suitable for manipulating virtual contents in 3D space given its high degree-of-freedom input [36]. However, it has also long been criticized for being imprecise [4, 41], fatiguing [32], and for lacking tactile feedback [20].

To further improve the usability and increase the interaction vocabulary of mid-air interaction, researchers have explored low-effort approaches with indirect mapping of input (e.g., a relaxed arms-down position [12, 40]) and employed computational models (e.g., based on selection distribution [56]) to improve its accuracy. Others have leveraged the potential benefit provided by multi-modal input and have incorporated other modalities (such as eye gaze [57], smartphones, and tablets) into the interaction [11]. For example, *BISHARE* [59] investigated joint interaction paradigms between smartphones and AR headsets to enrich AR interaction experiences by distributing system input and virtual content across both platforms. Other recent research including *SymbiosisSketch* [3], *TabletInVR* [48], and *VRSketchIn* [16] contributed new design spaces using on-tablet input to assist mid-air input in sketching and modelling in VR. In this work, we focus on using on-body interfaces to enhance and augment mid-air bare-hand interaction in VR headsets.

2.2 On-Body Interaction

On-body interfaces leverage the human body as an input/output platform [8, 28, 29]. Compared with smartphones and tablets, previous studies have identified that on-body interfaces provide the following unique benefits: they are always available for interaction [28, 29], afford a higher sense of agency [9, 15], and enable more accurate eyes-free targeting [26, 53]. Additionally, they support additional haptic feedback [25, 28], which has the potential to enable more precise and less physical demanding input than mid-air input due to the direct physical contact with the user's own body [4, 30]. For these reasons, on-body interaction holds a lot of potential for supporting mid-air interaction in VR headsets. However, on-body interfaces usually lack direct support for providing 3D input.

Existing literature has proposed several on-body interaction techniques [23, 34, 44]. For example, *Armura* [28] explored a set of possible interactions like menu navigation, page-turning, and peephole display using hands and arms as projection surfaces. *PalmGesture* [52], *PalmType* [51], and *DigiTouch* [54] all considered the use of on-palm input for text entry and widget-based interaction in AR/VR headsets. *SkinWidget* [5] demonstrated on-forearm touch, drag, slide, and rotation gestures for interacting with an on-arm menu in VR. *BodyLoci* [22] and *Tap-Tap Menu* [7] further used tapping gestures to interact with menus and buttons located on the whole body in VR. *DigiGlo* [13] proposed palm surfaces as a display in VR. Body-referenced input (interfaces that are attached close to a user's body surface) has also been explored in VR [6, 38, 55].

More relevant to our work are interaction techniques that consider combining both on-body and mid-air interfaces. *BodyScape* [50] evaluated a technique that employs mid-air gestures for pointing and on-arm tapping for selection confirmation. This work opened up new opportunities for combining the two interaction modalities. *WatchSense* [46] leveraged smartwatch-based fingertip tracking to enable combined mid-air and touch interaction by using the thumb as a base for touch input and the index finger for mid-air input. Ens et al. [18] integrated mini-scale on-finger input (for example, on a ring device) with mid-air gestures to allow 3D content manipulation by varying the temporal relationship of the input.

In summary, existing research has shown great promise of on-body interfaces, but few works have demonstrated their use for supporting mid-air interactions. Our research takes these ideas further by exploring how on-body interfaces should be incorporated into the mid-air workflow.

3 BodyOn

BodyOn is a collection of six design patterns that integrate on-body interfaces into current mid-air interaction workflows in VR headsets. In this section, we first present a design space that leverages on-body and mid-air interfaces as input and output modalities. We then identify design opportunities in the literature that motivate the design of BodyOn and detail the six design patterns which are templates of design that can be adopted to solve a multitude of interaction tasks.

3.1 Design Space

Both on-body and mid-air gestures can serve as modalities to capture user input or display output. We present a design space that connects on-body and mid-air interfaces in different input and output forms for interaction (see Figure 2 left).

The design space has two dimensions. One dimension is *input*: on-body, mid-air, and the combined on-body + mid-air information can all be used as input. In the scope of this research, on-body input leverages body contact information (on-body touch, gestures, or deformations [8]) as an input modality for interaction, while mid-air input employs mid-air gestures including hand translation, rotation, and relation as an input modality. The combination of on-body and mid-air input means that the interaction is a result of inputs from both modalities. For example, a user can achieve this by performing mid-air gestures with one hand and on-body gestures with the other hand for input. The other dimension of the design space is the *output*: both on-body and mid-air can be used as output. That is, virtual contents can be either attached to body surfaces or to the mid-air space as displayed output.

Based on the design space, we identify input $\rightarrow$ output mappings that combine on-body and mid-air interfaces, including On-Body $\rightarrow$ Mid-Air, Mid-Air $\rightarrow$ On-Body, and On-Body + Mid-Air $\rightarrow$ *Output* (on-body or mid-air). Additionally, we envision virtual content to be transferred between on-body and mid-air space for leveraging unique properties of the displays (*Output*: On-Body $\rightleftharpoons$ Mid-Air). Because this work focuses on blending on-body and mid-air techniques, we exclude conditions where there is only one single input and output modality (i.e., On-Body $\rightarrow$ On-Body and Mid-Air $\rightarrow$ Mid-Air). We scrutinize the relevant mappings in the next section.

3.2 Synthesis of Prior Work and Design Opportunities

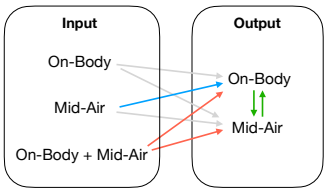
We have identified possible mappings between on-body and mid-air interfaces for input and output. We further explore new design opportunities by examining how existing research fits into our design space.

3.2.1 Manipulating Mid-Air Content with On-Body Input

For the On-Body $\rightarrow$ Mid-Air mapping, prior research has proposed techniques that leverage finger-on-palm gestures for text entry or application control (e.g., sliding fingers to increase the volume of an application) [51, 52, 54]. However, little work has employed on-body input for manipulating objects in 3D mid-air space. This is understandable if we consider the affordance of on-body input—body surfaces naturally afford 1D, 2D, but only limited 3D input based on their geometry and how they are positioned and stretched [10, 43]. Therefore, we deem existing applications that mainly use on-body input for 2D content manipulation appropriate and sufficient for this mapping.

3.2.2 Manipulating On-Body Content with Mid-Air Input

A few works have explored the Mid-Air $\rightarrow$ On-Body mapping [37]. For example, *Armura* [28] allows users to flip a page displayed on the



Mapping	Key Literature
On-Body → On-Body	*Armura* [28], *SkinWidget* [5], *Haptic Hand* [34], *Tap-Tap Menu* [7]
On-Body → Mid-Air	*PalmGesture* [52], *PalmType* [51], *DigiTouch* [54]
Mid-Air → Mid-Air	A common VR interaction paradigm [36]
Mid-Air → On-Body	*Armura* [28], *DigiGlo* [13], body-referenced input [37, 38, 55]
On-Body + Mid-Air → *Output*	*BodyScape* [50], *WatchSense* [46], Ring-based interaction [18]
Output: On-Body ⇌ Mid-Air	Not available

Figure 2: Design space and key literature summarization.

hand with swiping gestures. *DigiGlo* [13] enables users to interact with games displayed on their hands through various hand gestures in VR. Wrist-referenced interfaces [38, 55] allow users to interact with UIs displayed on or close to their wrist. While these works focus on using hands or arms as displays, we argue that body surfaces afford larger display areas if considering other body parts like the torso, legs, feet, etc. Different body parts can be designed to convey different semantic meanings of an interaction. Thus, one underexplored space is to use mid-air input to interact with virtual content displayed on body surfaces other than on hands and arms.

3.2.3 Combining On-Body and Mid-Air Input

Existing works have considered combining on-body and mid-air input for interaction (On-Body + Mid-Air → *Output*). *BodyScape* [50] uses one hand for mid-air pointing and the other hand performing on-arm tapping for selection confirmation. *WatchSense* [46] uses a thumb for on-hand touch (which creates a stable base) and an index finger for mid-air controls like zooming in/out an image. Ens et al. [18] use thumb-on-index finger tapping and swiping gestures to provide additional capabilities for mid-air input. While these works demonstrate the potential usefulness of combining on-body and mid-air input, there is still no cohesive view on how on-body and mid-air input should be combined, especially considering the bimanual input capability of hands [24]. Leveraging the feature that each hand can perform separate or combined on-body and mid-air actions, a user interface may create a richer set of interaction vocabularies to afford more complex interaction tasks in VR systems. Therefore, one design opportunity here is to scrutinize how on-body and mid-air input can be combined, considering the bimanual input property of hands.

3.2.4 Content Transfer Between On-Body and Mid-Air Space

Little research has explored content transfer between on-body and mid-air space (*Output*: On-Body ⇌ Mid-Air). However, mid-air and on-body spaces have unique display affordances. The mid-air space provides an extensive area for displaying 2D or 3D virtual content [19]. However, because virtual objects and user interfaces are anchored to the world space, unwanted occlusions may occur if users change their viewpoint (e.g., an element of interest is occluded by a wall [58]). In contrast, when a UI display is attached to the body surface, it follows the user's movement when travelling inside virtual environments and can be accessed once the user pays attention to it. For example, when a user is walking, the on-body displays attached to their wrists, belly, or feet will always be available for interactions when the user looks at them. Therefore, one design opportunity is to enable content transfer between the two displays to better leverage their strengths.

3.2.5 Summary

In sum, we conclude with three design opportunities. (1) On-Body + Mid-Air → *Output*: combining on-body and mid-air input for interaction, especially considering the bimanual input property for a rich set of interaction vocabularies, (2) Mid-Air → On-Body: extending the display of virtual contents to body parts other than arms and hands, and (3) *Output*: On-Body ⇌ Mid-Air: enabling content transfer between the two interfaces to better leverage their unique display properties.

3.3 Design Patterns

Based on the identified design opportunities, we propose BodyOn, a set of six design patterns that combine on-body (*OB*) and mid-air (*MA*) interfaces for interactions in VR headsets (see Figure 3). The design patterns leverage combined OB and MA input (**P1**-**P4**), MA input for OB content manipulation (**P5**), and content transfer between OB and MA space (**P6**).

3.3.1 Combining On-Body and Mid-Air Input

We envision that OB and MA input can be combined in various ways for interaction, especially considering the bimanual input property of hands. In this research, we restrict the input area of OB interfaces to hands and arms because they are more comfortable and socially acceptable by users across multiple poses [10, 27, 50].

Under this constraint, we identify two types of OB inputs that are suitable for combined OB and MA input: thumb-on-finger (*TOF*) input and finger-on-arm (*FOA*) input. TOF input leverages contact information between a thumb and other fingers on the same hand to issue an input. FOA input uses contact information between the fingers of one hand and the arm of the other hand to command input. Users can perform a diverse range of gestures including tapping, sliding, and drawing shapes, and information like contact locations, hardness, and gestures can be employed to construct input signals.

TOF input can be performed with one hand or both hands, and, concurrently, MA information of one or both hands can be leveraged for input. FOA input require the involvement of both hands, and the hand that does not perform FOA input can be used to provide MA input. These combinations result in the following four patterns.

Pattern 1 - Single Hand: MA + TOF. Users perform single hand TOF input and MA input together to interact with virtual objects. While previous research on TOF gestures mostly focused on gesture recognition [33, 45] or utilizing these gestures for interactions like text entry [54], our work emphasizes the incorporation of the TOF input into the MA input flow. In this case, MA information (i.e., hand position and/or orientation) is combined with TOF input to enable a richer set of interactions. For example, when manipulating an object with MA input, TOF input can provide another layer of control to adjust the object's movement speed.

Pattern 2 - Both Hands: MA (One Hand) + TOF. Users perform MA input with one hand and TOF input with the other hand or both hands. The pattern involves both hands, while only one hand's MA information (position and orientation) is used for input. The hand that issues MA input can work on a primary 3D interaction task, and the TOF input can act as background support for the primary task. For example, a user is drawing 3D curves with one hand in a virtual space, and the user can perform TOF input on the other hand to quickly change the drawing colours in an eyes-free manner without disturbing the workflow of the drawing hand.

Pattern 3 - Both Hands: MA (Both Hands) + TOF. Users carry out MA input with both hands and use TOF input on one or both of the hands. In this pattern, the MA information from both hands, including their locations, orientations, and relations, is used. Simultaneously, TOF input comes into play (performed by one or both hands) to uncover more complex interactions that are possible in 3D VR environ-

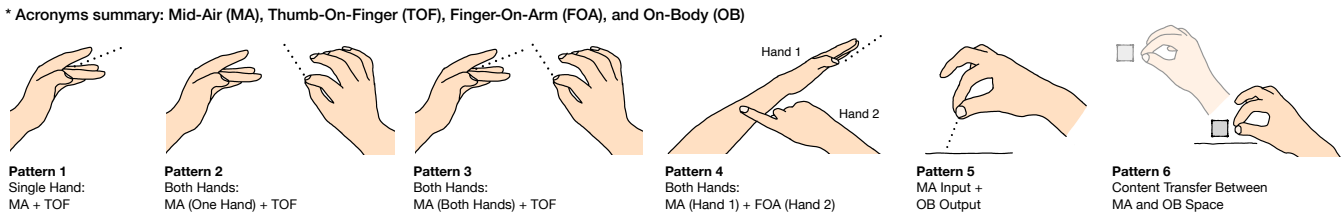


Figure 3: BodyOn is a collection of six design patterns that combine on-body and mid-air interfaces for new VR interactions. **P1** leverages single-handed thumb-on-finger (TOF) input and mid-air input (i.e., translation and orientation) for user input. **P2** involves both hands and uses TOF input to support mid-air input performed by the primary hand. **P3** employs TOF to support mid-air input performed by both hands. **P4** uses finger-on-arm input with one hand on the other arm, while the latter is used for mid-air input at the same time. **P5** utilizes mid-air input for interacting with on-body displays. **P6** enables content transfer between on-body and mid-air space.

ments. The underlying concept is similar to many asymmetric bimanual techniques where one hand acts as a spatial reference and the other is used for manipulation [24]. For example, using the MA information from both hands may allow users to rotate an object (holding by one hand) around a point (attached to the other hand) or move an object towards a particular point. TOF input can act as a mode switching trigger to allow the transformation to happen between those two possibilities.

Pattern 4 - Both Hands: MA (Hand 1) + FOA (Hand 2). Users perform FOA input with one hand on the other arm, while the latter is used for MA input at the same time. In this pattern, the arm that performs MA input also serves as a place for FOA input. This is a novel approach as previous works that use FOA gestures use them as a sole input modality [5, 39]. As an example of where this pattern would be useful, users may want to translate a 3D cursor [58] to select objects with different depths by sliding fingers on the arm and pointing in the target direction.

3.3.2 Manipulating On-Body Content with Mid-Air Input

While previous works have explored OB displays mainly on hands and arms, we want to expand the design space to consider content display on other body parts such as the torso and feet. Therefore, we summarize the following pattern.

Pattern 5 - MA Input + OB Display. Users use MA input techniques (like Raycasting, remote virtual hand, or distant triggering) to interact with OB displays. While the appropriate areas for direct OB input are restricted to hands and arms, OB displays can be extended to other body parts which can benefit users with their unique features (e.g., inherently following the user's movement). Thus, an alternative solution can be to use MA input to interact with such OB interfaces remotely. For example, users can point and select a virtual OB widget and move them across different body parts. They can also trigger certain actions remotely by putting one hand close to OB widgets.

3.3.3 Content Transfer Between On-Body and Mid-Air Space

We envision that enabling content transfer between OB and MA space can better leverage the display properties of the two interfaces. Therefore, we derive the following pattern.

Pattern 6 - Content Transfer. Users can transfer objects between MA and OB space. For example, users may want to store a model inside a 3D virtual space as a prefab for later use. In this case, they can transfer the object from the MA space to their OB space and put it back to the MA space at a different location.

4 Example Interaction Techniques based on BodyOn

To examine the *feasibility* and *applicability* of the design patterns, we developed a set of example interaction techniques based on BodyOn to solve various VR interaction tasks in a 3D modelling system[1]. Our goal was to use these interaction techniques as probes

[1]Open source: `https://github.com/Davin-Yu/BodyOn-ISMAR22`

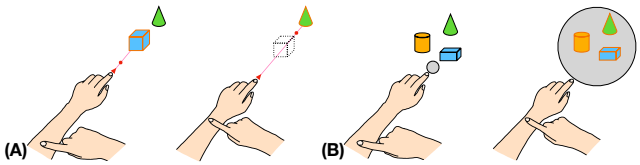


Figure 4: A user can select an occluded target (A) or a group of targets (B) with finger-on-arm gestures and mid-air pointing.

to test the high-level design concepts from BodyOn. By developing the techniques, our intention was to sketch "what is possible" with the new design patterns and map out possible design boundaries. These example techniques further allowed us to conduct an expert evaluation to elicit immediate design issues with the new combination of on-body and mid-air interactions.

For demonstration purposes, we used 3D modelling as a testbed because it involves canonical interactions (select, manipulate, travel, and system control [36]) with various complexities in 3D UI design. For each interaction task, we considered how on-body interfaces can enhance the current form of mid-air interaction or achieve additional functionalities by leveraging BodyOn. Table 1 provides an overview of the techniques and how they fit into the design patterns. Please also refer to our supplementary video for technique demonstrations.

4.1 Selection

Object selection is a fundamental task in interactive VR systems [2, 36]. Our interaction techniques based on BodyOn enable single object selection, occluded object selection, and group selection.

4.1.1 Simple Raycasting Selection

A user can select a target with Raycasting. When the pointer is "on" the object, the object will flicker to indicate that it is available for selection. The user can use the same mechanism to select objects attached to the on-body space (**P5**).

4.1.2 Occluded Target Selection

We developed a BodyOn-based occluded target selection technique inspired by *AlphaCursor* [58]. A user can control a movable cursor on the virtual ray attached to the index finger of their non-dominant hand (NDH) with finger-on-arm sliding gestures performed by their dominant hand (DH) (**P4**) to reveal occluded objects as the cursor goes deeper into the environment (see Figure 4A). The object is selected if a pinch gesture is performed with the DH. The object flickers once the selection ray hits it and gleams golden colour once selected.

4.1.3 Group Selection

A user can select a group of objects by controlling a resizable cursor attached to the index finger of their NDH. As shown in Figure 4B, the

Table 1: A summary of the implemented interaction techniques and how they fit into the design patterns. Acronyms: TOF (thumb-on-finger), OB (on-body), FOA (finger-on-arm), and MA (mid-air).

Design Patterns	Implemented Interaction Techniques
P1 - Single Hand: MA + TOF	Simple object manipulation, adjustable CD ratio
P2 - Both Hands: MA (One Hand) + TOF	Stroking, coloring, menu control, object creation and removal
P3 - Both Hands: MA (Both Hands) + TOF	Plane, ray, and point techniques
P4 - Both Hands: MA (Hand 1) + FOA (Hand 2)	Occluded target selection, group selection, 1 DOF transformation, teleportation
P5 - MA Input + OB Output	On-body object selection, travel through minimap
P6 - Content Transfer	Object storage and retrieval

user can make the cursor larger or smaller by sliding the DH index finger on the arm of the NDH (**P4**). The selection is triggered once a pinch gesture is performed with the DH, and all the flickering objects inside the cursor are selected.

4.2 Manipulation

Object manipulation tasks commonly include translation, rotation, and scaling of objects [36, 41]. Other tasks in relevant applications (e.g., Google Blocks and Tilt Brush) include stroking, colouring, object creation or removal, and object storage or retrieval.

4.2.1 Simple Object Manipulation

A common way of manipulating a selected object is to move or rotate the DH by holding the index finger pinch gesture. The object then follows the hand movement and rotation with 1:1 control-display mapping (CD Ratio = 1), as if the object is grabbed by the DH. A user can manipulate an on-body object in the same way (**P5**). Users hear a click sound once they select an object, and the facets of the selected object then start blinking.

Alternatively, a user can pinch their middle finger for object translation, pinch their ring finger for object rotation, and pinch their pinky finger for object scaling (**P1**) (see Figure 5). The additional three functionalities isolate the 6 degrees-of-freedom (DOF) virtual hand manipulation to 3 DOF for translation, rotation, and scaling. It does not require normal operations of going through multiple stages like using a DOF-separation widget, which may slow down the performance [31]. The quick access may give more control (object scaling) and precision (by separating the DOF [42]) for manipulation tasks.

4.2.2 Precise Object Manipulation

The techniques also enable precise object manipulation.

- *Adjustable CD Ratio*. When sliding the thumb from the fingertip to the root (**P1**), the control display mapping will change for each transformation. The CD Ratio changes to 2 when the thumb is on the second segment of the finger and changes to 1/3 when the thumb is on the third segment of the finger. This type of control may allow both precise (with a higher CD Ratio) and rapid (with a lower CD Ratio) manipulation [21]. Color indicators at fingertips turn to green, heavy green, or light green from their original state (gray) if normal, slow, or fast manipulations are enabled (see Figure 5B-E).

- *One DOF translation*. A user slides the DH index finger on the arm of the NDH (**P4**) to control a target moving along a line, which is defined by the pointing direction of the NDH (see Figure 6A). By isolating the movement to 1 DOF, the user may have more precise control over the manipulated target [41].

- *Plane, Ray, and Point* [31]. This technique uses shapes including planes, rays, and points to constrain object movement with multiple hand gestures [31] . Our method leverages the combination of on-body and mid-air bimanual input to achieve those functions (see Figure 6B-E). A user uses the NDH thumb to select the Plane, Ray, or Point technique with icons displayed on the NDH middle finger. A shape (plane, ray, or point) is generated once an index finger pinch is detected on DH, and the position and orientation of both hands are then used as references for the techniques (**P3**). The user may move the DH to rotate the selected object around a point, around a line, or along a plane. Alternatively, the user can quickly switch between different techniques by tapping their NDH thumb on the middle finger. If a middle finger pinch is detected when using the Point technique, the selected object moves towards the point rather than rotates around it. The design demonstrates that BodyOn allows more complex object control via both mid-air and on-body interfaces. Importantly, the menus displayed on-body make the functions fully discoverable and do not require remembering new gestures.

4.2.3 Stroking and Coloring

A user can produce a line stroke by holding DH index finger pinch. Meanwhile, the user can quickly access a colour palette displayed on NDH fingers and switch between different stroking colours with thumb-on-finger gestures (**P2**) (see Figure 7A). In this case, switching the colour may not disrupt the main workflow of the DH. A similar process can be followed to recolour an object.

4.2.4 Object Creation and Removal

A user can create an object (sphere, cube, cone, or cylinder) at the location of the DH by selecting a target shape icon on the NDH and pinching the DH index finger (**P2**) (see Figure 7B). The user can also use the DH index finger pinch to remove an object.

4.2.5 Object Storage and Retrieval

One interaction technique uses the on-body space as a container for storing and retrieving prefabs (**P6**). As shown in Figure 7C, a user can put a group of objects close to a pocket of the virtual avatar and release the DH index finger pinch to put them "into" the pocket. The saved prefab (on feet) can then be retrieved via Raycasting (**P5**).

4.3 Navigation

Teleportation and on-body minimap can be used for navigation.

4.3.1 Teleportation

A user can travel to a target location by teleportation with a parabolic curve. The initial curve has a take-off angle of 45°, a horizontal speed of 2m/s, with a vertical gravity acceleration. The user can perform a sliding gesture on the arm of the NDH (**P4**) to adjust the horizontal speed to maximize or minimize the furthest distance the user can travel through the teleportation technique.

4.3.2 Travel Through Minimap

A minimap [47] will pop up if a user puts the DH close to their abdomen. The minimap travel is triggered when the user moves the DH above the destination and performs a mid-air pinch gesture. The minimap can be closed if a user puts the DH close to their abdomen again. The manipulation of the on-body minimap relies on mid-air input for on-body displays (**P5**).

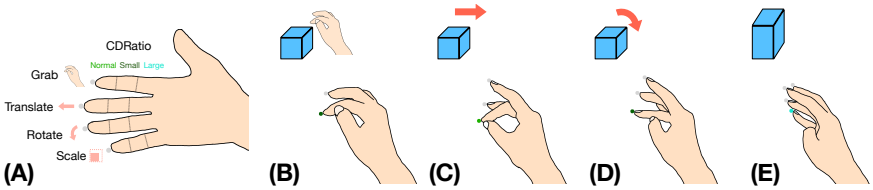


Figure 5: A user can manipulate an object using grabbing (B), translating (C), rotating (D), and scaling (E) by tapping the thumb on the index, middle, ring, and pinky fingers. The user can further adjust the movement/rotation speed (CD Ratio) to normal, slow, and fast by tapping on the first, second, and third segments of the finger.

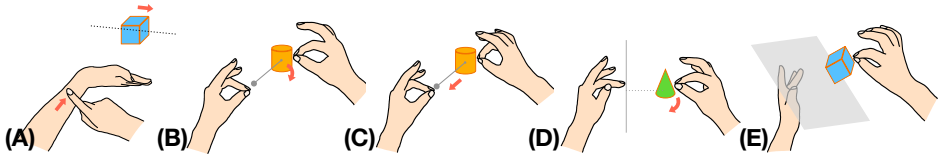


Figure 6: A user can translate an object in one DOF to enable more precise control by pointing at a movement direction through NDH and performing finger-on-arm gestures with DH (A). By combining various bimanual thumb-on-finger gestures and mid-air motions, the user can move an object around a point (B), towards a point (C), around a line (D), and along a plane (E).

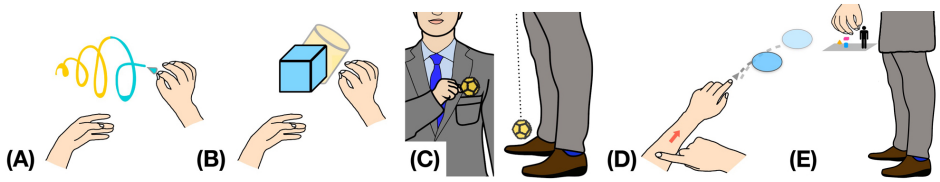


Figure 7: A user can use DH to draw lines with different colours (A) and create various shapes (B) by performing thumb-on-finger gestures on the NDH without disrupting the main workflow in the DH. The user can also store a group of objects by putting them into the pocket and later retrieving them from the feet (C). Moreover, the user can adjust the target destination of teleportation through finger-on-arm gestures (D) and travel to different locations by manipulating an on-body minimap (E).

4.4 System Control

We use the menu structure to navigate between the aforementioned functionalities or modes. The menu items are selected when the NDH thumb is tapped on the corresponding icon located on NDH fingers (**P2**). A user can quickly switch between different system functions without disturbing the main workflow. Furthermore, the eyes-free capability offered by on-body input may allow expert users to access different modes without looking at the icons.

4.5 Implementation

The interaction techniques based on BodyOn were developed with an Oculus Quest 2 headset (1832×1920 pixel resolution per eye). Hand tracking is enabled by its inside-out cameras, and the hand keypoints data are streamed from the OVR Plugin version 1.55.1. The software was developed using C# in Unity (version: 2020.1.17f1).

The arm and leg postures were approximated with two bone inverse kinematics (IK) constraints in the Animation Rigging package (version: 0.3.4). The feet would not go through a virtual floor, and the animated character's body rotation was constantly linearly interpolated to the horizontal orientation of the users' eyes.

The current vision-based hand-tracking in the headset still has limited tracking accuracy. They can suffer from occlusion and noise (e.g., lighting conditions), which may lead to inaccurate results when users' hands move around. Therefore, we implemented the thumb-on-finger and finger-on-arm gestures with the following compensations in our program to make the techniques more robust.

A thumb-on-finger gesture is detected once the distance between the thumb tip and other fingers' bones is smaller than 0.02m for index fingers or 0.03m for the middle, ring, and pinky fingers (as we found the tracking to be more accurate on index fingers). We determined the area of touch by calculating the distance from the thumb tip to the joints (proximal interphalangeal joints, intermediate interphalangeal joints, and distal interphalangeal joints) and the tips of each finger. We further increased the robustness of menu selection by picking up the closest menu icon to the thumb tip once the thumb-on-finger gestures are observed. When the hand movement exceeds a threshold (0.002m displacement and 0.5° rotation in 25 frames), the thumb tip is "locked" onto the finger to prevent unexpected clicking during the movement.

Similarly, a finger-on-arm gesture is detected once the distance between the index fingertip and forearm is smaller than 0.05m. The touch location is determined by calculating the distance between the index fingertip to the elbow and wrist.

5 Expert Evaluation

Our interaction techniques represent different design possibilities based on the six design patterns of BodyOn. Therefore, the primary goal of our evaluation is not to fully validate the design space, but instead to use the techniques as probes to elicit immediate design issues with the novel combination of on-body and mid-air interactions.

5.1 Participants and Apparatus

Six experts, including one woman and five men, aged between 26 and 36, were recruited. All of them frequently use desktop-based 3D modelling software like Blender, Maya, AutoCAD, and Fusion 360 or game development applications such as Unity and Unreal. Three reported using VR/AR devices 3-5 times per week, while one reported using these devices almost every day. We hoped that domain experts would give us more insightful feedback on the interaction

techniques and tools as they've already had previous experiences dealing with similar software (like modelling tools on PC). They were compensated $20 for participating in the study.

The study was conducted in a 3m × 4m tracking space. An Oculus Quest 2 headset, which is a standalone VR headset, was used in the study. The user's view was streamed to a laptop through Wi-Fi for observation and instruction.

5.2 Procedure

The walkthrough experience took about 60 minutes for each expert and consisted of the following three phases.

5.2.1 Welcome and Briefing (10 minutes)

The experts first filled in a consent form and a demographics questionnaire. We then introduced them to the purpose of the walkthrough, the overview of the six design patterns, and the interaction types that the techniques support (selection, manipulation, navigation, and system control).

5.2.2 Guided and Free-Form Exploration (30 minutes)

During the walkthrough experience, the experts were guided through all the techniques that corresponded to the six design patterns and were asked to complete specific tasks like constructing a door on its frame and rotating it around (detailed in the supplementary material). After completing all the required tasks, they were asked to perform free-from exploration while providing their thoughts on the interaction.

5.2.3 Interview (20 minutes)

After the exploration, we conducted a semi-structured interview with the experts where we asked them to (i) illustrate the advantages and disadvantages of the techniques over previous tools they had used in desktop software and VR applications; (ii) give their overall impression about the usability and learnability of the interaction techniques; (iii) describe what they liked and disliked; (iv) provide opinions on how we should further improve the techniques; (v) any other comments about the techniques or patterns that they had not covered.

5.3 Results

Overall, the experts (*E* in short) enjoyed the walkthrough experience and were positive about the combination of on-body and mid-air interaction. For example, *E1* commented "*The interactions are really intuitive, and the concepts behind the system are amazing!*" By combining on-body and mid-air interfaces, the system certainly brought "*a lot of new functionalities*" (*E2*, *E3*, and *E6*) as compared to existing software.

Using both on-body and mid-air gestures as input, users found many clever and helpful features were enabled. For example, the manipulation techniques of changing CD Ratio and isolating transformation enabled by single hand thumb-on-finger and mid-air gestures (**P1**) were mentioned to allow "*more accurate manipulation*" (*E6*) and could "*speed up the transformation for a large room*" (*E1*). experts also noticed that the gestures and techniques were "*easy to learn*" and they could control an object or switch between different modes with on-body gestures without looking at their hands or arms (eyes-free input). All experts particularly liked **P4**, with which they performed mid-air gestures with one hand and finger-on-arm gestures with the other hand to achieve operations like occluded object selection. For example, *E1* said that "sliding on arms was not tiring." *E2* mentioned that "*it enables a lot more functions and is less fatiguing (than mid-air input alone).*"

Several interesting comments pointed out potential issues with the current implementation of combined on-body and mid-air input. One main issue was related to how the feedback of on-body input should be displayed. *E3* noticed that it was hard to perceive the visual feedback provided on-body while focusing on the mid-air input. While using thumb-on-finger gestures to change CD Ratio, *E3* commented that "*because the (visual) feedback is on fingertips, when I am focusing on an object, I cannot see the feedback.*" Similarly, when performing mid-air tasks with one hand and on-body gestures with the other hand as support (**P2**), *E3* felt that when focusing on the mid-air input (e.g., painting) the current visual feedback provided on the non-dominant hand (which might be moved outside of the user's view) was not enough. *E3* mentioned that "*I need to see the feedback (of which mode the system is in).*" These comments resonated with the experience of some experts like *E6* who encountered unintentional misclicks from the thumb-on-finger input with the supporting hand (maybe due to system recognition error) and got confused about the unexpected mode switching event through the on-body input. *E6*, therefore, suggested that "*it would be better to sometimes detach the control panel on the body surface and put it in mid-air or disable it (to avoid misclicks).*"

In addition, the users also had various opinions on the input regions of thumb-to-finger gestures. While *E1* and *E5* found no problem performing all the gestures, others felt uncomfortable holding the thumb on the root of other fingers. Therefore, *E2* and *E3* suggested using thumb sliding and holding gestures only on the index and middle fingers, and *E3* further recommended using the pinky finger as a display rather than as an input region.

Another interesting finding from our observation is that although mid-air and on-body information is leveraged by the design patterns at the same time, users may not perform the mid-air and on-body input simultaneously. For example, while a user is performing mid-air pointing, finger-on-arm sliding often happens after the user has already pointed at the desired direction (e.g., for one DOF translation).

Regarding interaction techniques that allowed mid-air gestures to interact with on-body displays, all the users liked the minimap attached to the abdomen. They said that, for example, "*taking out a minimap from my body is cool.*" (*E1*) and described minimap as "*my favourite feature*" (*E4*). *E6* mentioned that it provided "*a nice top-down view (of the virtual environment)*". Users also found on-body and mid-air content transfer (**P6**) to be helpful and "*is the shortcut for copy and paste*" (*E2*). However, the placement of the on-body visualization may need to be carefully considered. *E5* mentioned that the minimap was placed "*too close to the body*". To retrieve an object from the foot, *E3* mentioned that "*I have to bend my body (to see the objects on my foot).*"

6 Discussion

This paper introduces BodyOn, a collection of six design patterns that leverage both on-body and mid-air interfaces to achieve better interactions in VR. The patterns were designed for (1) combining on-body and mid-air input, especially considering the bimanual input property (2) extending the display area of virtual contents to body parts other than arms and hands, and (3) enabling content transfer between on-body and mid-air space. We ground our design concepts on a set of example interaction techniques to solve tasks at various complexities in a 3D modelling system. We further use these techniques as probes to elicit immediate design issues with the novel combination of on-body and mid-air interfaces in an expert evaluation study. In this section, we reflect on the lessons learned from our experience, and discuss limitations and future work.

6.1 Combining On-Body and Mid-Air Interaction

By instantiating the high-level design concepts through the interaction techniques, we confirm that BodyOn can provide versatile interaction vocabularies to support the current VR workflow based on mid-air interaction. On-body interfaces can provide quick controls to adjust the control-display ratio and isolate transformation with a simple combination of single-handed thumb-on-finger clicks/swipes and mid-air movements (**P1**). They also offer quick access to different tools with thumb-to-finger gestures as background support (**P2**). More complex interactions can be enabled by leveraging the mid-air relationship between two hands and combining it with thumb-on-finger input

(**P3**). Mid-air gestures can also be combined with 1D/2D sliding input on the arms to achieve additional useful and effective functions like selecting an occluded object (**P4**). Furthermore, using mid-air input to interact with on-body displays (**P5**) and transferring contents between mid-air and on-body space (**P6**) leverage the unique property of on-body display to make the content/information accessible while a user is moving inside virtual environments. The virtual menus displayed on body surfaces also make the interaction discoverable.

Our expert evaluation has demonstrated a great potential of combining on-body and mid-air interfaces. It showed that the interactions based on BodyOn could be quickly integrated into the mid-air interaction-based workflow and support the desired functionalities. The expert evaluation study also points out valuable lessons (*L*s) to further improve the designs.

6.1.1 Cognitive Bandwidth of On-Body and Mid-Air Interfaces

While BodyOn leverages on-body and mid-air input information simultaneously for interaction, users seem to have limited cognitive bandwidth in processing the information of two interfaces at the same time. For example, users were found to tend to perform finger-on-arm input after the hand that performed mid-air input has already pointed in the desired direction. Designers may need to *consider the additional cognitive load when combining these two interfaces and allow users to perform the actions sequentially (L1)*.

Furthermore, when users were focusing on manipulating objects located in the mid-air space, it was sometimes difficult for them to notice on-body visual feedback, such as small indicators on a fingertip or highlighted icons on a hand. The later issue may result in user confusion with the unintentional misclicks caused by on-body input because the input feedback is not perceived by the user. Therefore, it is essential to *present the feedback of on-body input within users' attention regions (L2)*. For example, it can be beneficial to provide a flashing icon on HUD or distinguished sound feedback when on-body input is detected to avoid user confusion. Such solutions aim to communicate the on-body input event that is being triggered while may introduce an additional cognitive burden in practical use.

Additionally, because unwanted on-body events can be caused by touching a trigger unintentionally when users are interacting with objects in the mid-air space, we recommend *providing a centralized button/gesture to switch on-body interfaces on and off as needed (L3)*. Another potential strategy is to implicitly determine users' current intention and determine whether an on-body click/touch should trigger a new event to mitigate the effect of misclicks [57]. For example, a designer can check the direction of gaze (on either body surfaces or mid-air interfaces) as an indicator of whether the user intends to perform on-body input. While these approaches may automatically filter out a large number of unintentional clicks, they can induce false-positive classifications (i.e., misclassifying a user's true intention).

6.1.2 On-Body Input and Output Location

Because previous research suggests that restricting the input area of on-body interfaces to hands and arms can be more comfortable and socially acceptable by users [10,27,50], we chose to employ thumb-to-finger and finger-to-arm gestures for on-body input. While all experts liked finger-to-arm gestures, we found it would have been beneficial to *enable thumb-to-finger input region customization (L4)*, because users have different preferences for the thumb-to-finger input regions. It will be useful to consider results from previous research by constraining the touching area to the first and second segments of the index and middle fingers to satisfy a larger population [33]. It may be further helpful to allow users to customize their own comfort regions and assign different functionalities on different finger segments by themselves (like personalizing their input control on a game controller).

Placing user interfaces on body surfaces like torso and feet can utilize previously unused on-body space for virtual content display. Interfaces presented on different body parts may convey different semantic meanings of interaction (e.g., putting a virtual object close to the heart means saving the object) and offer different viewing perspectives (e.g., top-down view of an on-body minimap attached close to the abdomen). Through the evaluation, we learned that *the location of on-body displays still needs to be carefully designed (L5)*. Due to the weight of current head-mounted displays, placing an object at locations that require users to heavily bend their body/neck (e.g., close to the chest) can induce discomfort.

6.2 Applications

BodyOn encompasses high-level design concepts that integrate on-body and mid-air interfaces. We envision the design patterns to be generalizable to other interactive applications in addition to 3D modelling. For example, in the emerging field of immersive analytics [17, 23], where users apply immersive technologies for data understanding and sense-making, new interactions are required for more challenging task scenarios like data manipulation and transformation. In addition, designers should have more choices to map out more complex interactions with the vocabularies enabled by BodyOn. Moreover, BodyOn can also inspire more fruitful interaction experiences in VR games. We also envision the design concepts of BodyOn to be adaptable to other displays like AR if carefully considering the affordance of the platform.

6.3 Limitations and Future Work

While the results of our study are encouraging, we have also identified several limitations regarding our current design and evaluation for future work. Our prototype was based on visual trackers from the headset (to make the system self-contained), and the tracking was not always accurate. Thus, the experts needed to adjust their postures periodically (e.g., rotating their hands or moving the hands back to the tracking area) to let the system recognize their postures, which might have affected their interaction experiences. Furthermore, the virtual character's body posture was approximated by inverse kinematics, and the torso and foot postures were not accurately captured. There are some more interesting design opportunities if the virtual character could follow the movement of users' feet and legs. Therefore, future work can incorporate tracking technologies with higher precision to explore these opportunities.

We also acknowledge the importance of quantitatively evaluating the techniques' performance in terms of, for example, user completion time and learning time. However, we did not conduct such studies because our designs were not implemented on a highly-accurate motion capture system (e.g., OptiTrack). Performing quantitative performance evaluation on our current prototype may introduce noises from the tracking system, thus producing misleading results. Therefore, we pursued a qualitative expert evaluation where our goal was to help elicit immediate design issues regarding the new combination of on-body and mid-air interfaces. We would like to include a quantitative evaluation of a more accurate system in a future study.

7 Conclusion

We present BodyOn, a collection of six design patterns that leverage both on-body and mid-air interfaces collaboratively for better VR interactions. Interactive techniques based on BodyOn were developed to showcase the possible designs with the patterns. We found that techniques based on BodyOn could provide flexible control, offer quick access to different tools, and bring additional useful and effective functions. They were easy to learn and could be quickly integrated into the mid-air interaction workflow. Our study also revealed some issues with our current implementation, such as users ignoring on-body visual feedback when focusing on mid-air tasks. Finally, we discussed the lessons learned from the implementation and evaluation, which can inform the design of future systems that blend both on-body and mid-air interactions. We envision BodyOn inspiring new interactions in a multitude 3D interaction scenarios in the future.

REFERENCES

[1] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S. Zhao, and J. Koh. 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, 2012.

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

[3] R. Arora, R. Habib Kazi, T. Grossman, G. Fitzmaurice, and K. Singh. Symbiosissketch: Combining 2d & 3d sketching for designing detailed 3d objects in situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–15. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173759

[4] R. Arora, R. H. Kazi, F. Anderson, T. Grossman, K. Singh, and G. Fitzmaurice. Experimental evaluation of sketching on surfaces in vr. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, p. 5643–5654. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025474

[5] T. Azai, S. Ogawa, M. Otsuki, F. Shibata, and A. Kimura. Selection and manipulation methods for a menu widget on the human forearm. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, p. 357–360. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3027063.3052959

[6] T. Azai, M. Otsuki, F. Shibata, and A. Kimura. Open palm menu: A virtual menu placed in front of the palm. In *Proceedings of the 9th Augmented Human International Conference*, AH '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3174910.3174929

[7] T. Azai, S. Ushiro, J. Li, M. Otsuki, F. Shibata, and A. Kimura. Tap-tap menu: Body touching for virtual interactive menus. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, VRST '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3281505.3281561

[8] J. Bergström and K. Hornbæk. Human–computer interaction on the skin. *ACM Comput. Surv.*, 52(4), Aug. 2019. doi: 10.1145/3332166

[9] J. Bergstrom-Lehtovirta, D. Coyle, J. Knibbe, and K. Hornbæk. I really did that: Sense of agency with touchpad, keyboard, and on-skin interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–8. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173952

[10] J. Bergstrom-Lehtovirta, K. Hornbæk, and S. Boring. *It's a Wrap: Mapping On-Skin Input to Off-Skin Displays*, p. 1–11. Association for Computing Machinery, New York, NY, USA, 2018.

[11] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg. The state of the art of spatial interfaces for 3d visualization. In *Computer Graphics Forum*, vol. 40, pp. 293–326. Wiley Online Library, 2021. doi: 10.1111/cgf.14189

[12] E. Brasier, O. Chapuis, N. Ferey, J. Vezien, and C. Appert. Arpads: Mid-air indirect input for augmented reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 332–343. IEEE, 2020. doi: 10.1109/ISMAR50242.2020.00060

[13] J. Chatain, D. M. Sisserman, L. Reichardt, V. Fayolle, M. Kapur, R. W. Sumner, F. Zünd, and A. H. Bermano. Digiglo: Exploring the palm as an input and display mechanism through digital gloves. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '20, p. 374–385. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3410404.3414260

[14] P. I. Cornelio Martinez, S. De Pirro, C. T. Vi, and S. Subramanian. Agency in mid-air interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, p. 2426–2439. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025457

[15] D. Coyle, J. Moore, P. O. Kristensson, P. Fletcher, and A. Blackwell. I did that! measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2025–2034. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208350

[16] T. Drey, J. Gugenheimer, J. Karlbauer, M. Milo, and E. Rukzio. Vrsketchin: Exploring the design space of pen and tablet interaction for 3d sketching in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376628

[17] B. Ens, B. Bach, M. Cordeil, U. Engelke, M. Serrano, W. Willett, A. Prouzeau, C. Anthes, W. Büschel, C. Dunne, T. Dwyer, J. Grubert, J. H. Haga, N. Kirshenbaum, D. Kobayashi, T. Lin, M. Olaosebikan, F. Pointecker, D. Saffo, N. Saquib, D. Schmalstieg, D. A. Szafir, M. Whitlock, and Y. Yang. Grand challenges in immersive analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021.

[18] B. Ens, A. Byagowi, T. Han, J. D. Hincapié-Ramos, and P. Irani. Combining ring input with hand tracking for precise, natural interaction with spatial analytic interfaces. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, SUI '16, p. 99–102. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2983310.2985757

[19] B. M. Ens, R. Finnegan, and P. P. Irani. The personal cockpit: A spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 3171–3180. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557058

[20] C. Fang, Y. Zhang, M. Dworman, and C. Harrison. Wireality: Enabling complex tangible geometries in virtual reality with worn multi-string haptics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–10. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376470

[21] S. Frees, G. D. Kessler, and E. Kay. Prism interaction for enhancing control in immersive virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 14(1):2–es, May 2007. doi: 10.1145/1229855.1229857

[22] B. Fruchard, E. Lecolinet, and O. Chapuis. Impact of semantic aids on command memorization for on-body interaction and directional gestures. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, AVI '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3206505.3206524

[23] B. Fruchard, A. Prouzeau, O. Chapuis, and E. Lecolinet. Leveraging body interactions to support immersive analytics. In *The ACM CHI Conference on Human Factors in Computing Systems-Workshop on Interaction Design & Prototyping for Immersive Analytics*, pp. 10–pages, 2019.

[24] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior*, 19(4):486–517, 1987. doi: 10.1080/00222895.1987.10735426

[25] S. Gustafson, C. Holz, and P. Baudisch. Imaginary phone: Learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, p. 283–292. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/2047196.2047233

[26] S. G. Gustafson, B. Rabe, and P. M. Baudisch. Understanding palm-based imaginary interfaces: The role of visual and tactile cues when browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 889–898. Association for Computing Machinery, New York, NY, USA, 2013.

[27] C. Harrison and H. Faste. Implications of location and touch for on-body projected interfaces. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, p. 543–552. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2598510.2598587

[28] C. Harrison, S. Ramamurthy, and S. E. Hudson. On-body interaction: Armed and dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, p. 69–76. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2148131.2148148

[29] C. Harrison, D. Tan, and D. Morris. Skinput: Appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, p. 453–462. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1753326.1753394

[30] H. Havlucu, M. Y. Ergin, I. Bostan, O. T. Buruk, T. Göksun, and O. Özcan. It made more sense: Comparison of user-elicited on-skin touch and freehand gesture sets. In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pp. 159–171. Springer, 2017. doi: 10.1007/978-3-319-58697-7_11

[31] D. Hayatpur, S. Heo, H. Xia, W. Stuerzlinger, and D. Wigdor. Plane, ray, and point: Enabling precise spatial manipulations with shape constraints. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 1185–1195. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3332165.3347916

[32] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. A survey of design issues in spatial input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, UIST '94, p. 213–222. Association for Computing Machinery, New York, NY, USA, 1994. doi: 10.1145/192426.192501

[33] D.-Y. Huang, L. Chan, S. Yang, F. Wang, R.-H. Liang, D.-N. Yang, Y.-P. Hung, and B.-Y. Chen. Digitspace: Designing thumb-to-fingers touch interfaces for one-handed and eyes-free interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 1526–1537. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858483

[34] L. Kohli and M. Whitton. The haptic hand: providing user interface feedback with the non-dominant hand in virtual environments. In *Proceedings of Graphics Interface 2005*, pp. 1–8, 2005. doi: 10.5555/1089508.1089510

[35] P. Koutsabasis and P. Vogiatzidakis. Empirical research in mid-air interaction: A systematic review. *International Journal of Human–Computer Interaction*, 35(18):1747–1768, 2019. doi: 10.1080/10447318.2019.1572352

[36] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. 2017.

[37] I. Lediaeva and J. LaViola. Evaluation of body-referenced graphical menus in virtual environments. In *Proceedings of Graphics Interface 2020*, GI 2020, pp. 308 – 316, 2020. doi: 10.20380/GI2020.31

[38] Z. Li, J. Chan, J. Walton, H. Benko, D. Wigdor, and M. Glueck. Armstrong: An empirical examination of pointing at non-dominant arm-anchored uis in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445064

[39] S.-Y. Lin, C.-H. Su, K.-Y. Cheng, R.-H. Liang, T.-H. Kuo, and B.-Y. Chen. Pub - point upon body: Exploring eyes-free interaction and methods on an arm. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, p. 481–488. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/2047196.2047259

[40] M. Liu, M. Nancel, and D. Vogel. Gunslinger: Subtle arms-down mid-air interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software Technology*, UIST '15, p. 63–71. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2807442.2807489

[41] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, vol. 38, pp. 21–45. Wiley Online Library, 2019. doi: 10.1111/cgf.13390

[42] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge. The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, p. 261–268. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2993369.2993396

[43] H. Saidi, M. Serrano, P. Irani, C. Hurter, and E. Dubois. On-body tangible interaction: Using the body to support tangible manipulations for immersive environments. In *IFIP Conference on Human-Computer Interaction*, pp. 471–492. Springer, 2019. doi: 10.1007/978-3-030-29390-1_26

[44] M. Serrano, B. M. Ens, and P. P. Irani. Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 3181–3190. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2556984

[45] M. Soliman, F. Mueller, L. Hegemann, J. S. Roo, C. Theobalt, and J. Steimle. Fingerinput: Capturing expressive single-hand thumb-to-finger microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ISS '18, p. 177–187. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3279778.3279799

[46] S. Sridhar, A. Markussen, A. Oulasvirta, C. Theobalt, and S. Boring. Watchsense: On- and above-skin input sensing through a wearable depth sensor. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 3891–3902. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3026005

[47] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: Interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, p. 265–272. ACM Press/Addison-Wesley Publishing Co., USA, 1995. doi: 10.1145/223904.223938

[48] H. B. Surale, A. Gupta, M. Hancock, and D. Vogel. Tabletinvr: Exploring the design space for using a multi-touch tablet in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300243

[49] P. Vogiatzidakis and P. Koutsabasis. Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies and Interaction*, 2(4):65, 2018. doi: 10.3390/mti2040065

[50] J. Wagner, M. Nancel, S. G. Gustafson, S. Huot, and W. E. Mackay. Body-centric design space for multi-surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 1299–1308. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2466170

[51] C.-Y. Wang, W.-C. Chu, P.-T. Chiu, M.-C. Hsiu, Y.-H. Chiang, and M. Y. Chen. Palmtype: Using palms as keyboards for smart glasses. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, p. 153–160, 2015. doi: 10.1145/2785830.2785886

[52] C.-Y. Wang, M.-C. Hsiu, P.-T. Chiu, C.-H. Chang, L. Chan, B.-Y. Chen, and M. Y. Chen. Palmgesture: Using palms as gesture interfaces for eyes-free input. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, p. 217–226, 2015. doi: 10.1145/2785830.2785885

[53] M. Weigel, A. S. Nittala, A. Olwal, and J. Steimle. *SkinMarks: Enabling Interactions on Body Landmarks Using Conformal Skin Electronics*, p. 3095–3105. Association for Computing Machinery, New York, NY, USA, 2017.

[54] E. Whitmire, M. Jain, D. Jain, G. Nelson, R. Karkar, S. Patel, and M. Goel. Digitouch: Reconfigurable thumb-to-finger input and text entry on head-mounted displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), Sept. 2017. doi: 10.1145/3130978

[55] X. Xu, A. Dancu, P. Maes, and S. Nanayakkara. Hand range interface: Information always at hand with a body-centric mid-air input surface. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3229434.3229449

[56] D. Yu, H.-N. Liang, X. Lu, K. Fan, and B. Ens. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Trans. Graph.*, 38(6), Nov. 2019. doi: 10.1145/3355089.3356544

[57] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445343

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

[59] F. Zhu and T. Grossman. Bishare: Exploring bidirectional interactions between smartphones and head-mounted augmented reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376233

Chapter 7

OPTIMIZING INTELLIGENT SUGGESTION TIMING

7.1 Summary

In this work, we optimize the timing of displaying an intelligent suggestion to provide timely support for users in an interactive task. Intelligent suggestion techniques leverage probability estimates from a target prediction model to provide users with an easy-to-use method (e.g., a button click) to interact with the most probable target in an interaction scenario. Such techniques alleviate the need to manually point at targets or conduct a full visual search of an environment. Through a series of three experiments, we showed that our framework was both theoretically and empirically effective for providing intelligent suggestions at optimal timing.

Our experiments demonstrated that the solution suited small and distant target acquisition. The solution was effective in preventing errors in a dense target selection task and efficient in offering suggestions that could shorten the task completion time. It could mitigate the need for precise pointing and could improve user experience. The framework can be adapted to various interaction scenarios (e.g., a cluttered environment or a mentally-demanding task).

Env.			**Task**				
Small	*Distant*	*Occluded*	*Effectiveness*	*Efficiency*	*Ergonomics*	*Experience*	*Expressivity*
✓	✓		✓	✓	✓	✓	✓

7.2 Article IV

Difeng Yu, Ruta Desai, Ting Zhang, Hrvoje Benko, Tanya R. Jonker, and Aakar Gupta. "Optimizing the Timing of Intelligent Suggestion in Virtual Reality." In Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, pp. 1-20. 2022. https://doi.org/10.1145/3526113.3545632

Optimizing the Timing of Intelligent Suggestion in Virtual Reality

Difeng Yu
University of Melbourne
Melbourne, VIC, Australia

Ruta Desai
Reality Labs Research, Meta Inc
Redmond, WA, USA

Ting Zhang
Reality Labs Research, Meta Inc
Redmond, WA, USA

Hrvoje Benko
Reality Labs Research, Meta Inc
Redmond, WA, USA

Tanya R. Jonker
Reality Labs Research, Meta Inc
Redmond, WA, USA

Aakar Gupta
Reality Labs Research, Meta Inc
Redmond, WA, USA

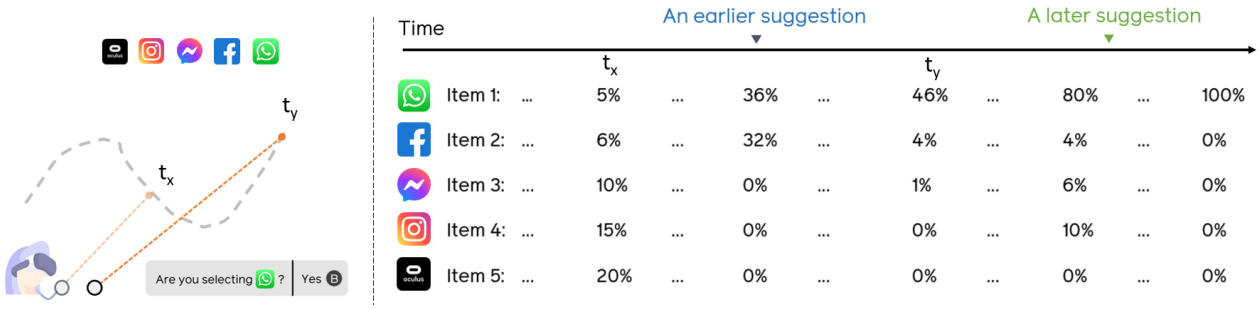


Figure 1: An overview of the intelligent suggestion timing problem. While a user is attempting to select an icon in virtual reality, a target prediction model could be continuously estimating the likelihood that the user will select each icon (e.g., at timestamp t_x and t_y). Depending on the results of these estimations, a system could then display an intelligent suggestion to the user that highlights the most probable icon for them to select. This suggestion, for example, could enable them to select an icon using a simple click, so that the user does not need to manually point towards the icon. While such suggestions could improve the usability of intelligent user interfaces, it is currently unknown whether early suggestions, which could save the user time and effort but may be less accurate, or later suggestions, which could save less time and effort but may be more accurate, are more beneficial for users.

ABSTRACT

Intelligent suggestion techniques can enable low-friction selection-based input within virtual or augmented reality (VR/AR) systems. Such techniques leverage probability estimates from a target prediction model to provide users with an easy-to-use method to select the most probable target in an environment. For example, a system could highlight the predicted target and enable a user to select it with a simple click. However, as the probability estimates can be made at any time, it is unclear *when* an intelligent suggestion should be presented. Earlier suggestions could save a user time and effort but be less accurate. Later suggestions, on the other hand, could be more accurate but save less time and effort. This paper thus proposes a computational framework that can be used to determine the optimal timing of intelligent suggestions based on user-centric costs and benefits. A series of studies demonstrated the value of the framework for minimizing task completion time and maximizing suggestion usage and showed that it was both theoretically and empirically effective at determining the optimal timing for intelligent suggestions.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; *Mixed / augmented reality*; *Virtual reality*.

KEYWORDS

intention prediction, intelligent interfaces, optimization framework, reinforcement learning

ACM Reference Format:
Difeng Yu, Ruta Desai, Ting Zhang, Hrvoje Benko, Tanya R. Jonker, and Aakar Gupta. 2022. Optimizing the Timing of Intelligent Suggestion in Virtual Reality. In *The 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3526113.3545632

UIST '22, October 29-November 2, 2022, Bend, OR, USA

ACM ISBN 978-1-4503-9320-1/22/10…$15.00
https://doi.org/10.1145/3526113.3545632

1 INTRODUCTION

Target selection in virtual and augmented reality (VR/AR) systems is difficult, especially when interaction scenarios are complex (e.g., with small, faraway, cluttered objects) and input techniques are cumbersome to use (e.g., mid-air hand pointing). Recent research has utilized statistical or machine learning models to estimate the

likelihood of a user selecting different items or objects of interest [20, 25, 65]. Based on the estimated probabilities computed by these models, an interaction system may then use visual highlighting or display a notification to draw the user's attention towards the most probable target. Next, a user may select the predicted target with a shortcut (such as a simple click) [1, 30, 70]. Such techniques can alleviate the need to manually point at targets or conduct a full visual search of an environment, potentially leading to quicker, easier, and more comfortable interactions. They can also be useful within VR/AR systems that employ noisy, high-friction input modalities [1, 22, 70] or support scenarios that require users to complete manually-intensive or mental-demanding tasks, such as selecting objects in a cluttered environment or navigating through a complex hierarchical menu [17, 25, 38, 71].

While current target prediction models can determine *which* target a user may select, they cannot determine *when* intelligent suggestions should be provided to users. While an earlier suggestion could save a user time and effort, such suggestions have a higher chance of being incorrect, which could cause users frustration, break their trust, or decrease their performance [12, 40]. On the other hand, later suggestions are likely to be more accurate but less beneficial because users have already spent ample time and effort to complete their task. By the time a model has accumulated enough evidence to be certain of a user's intended target, the user may have almost completed their action, thus rendering the late-breaking intelligent suggestion useless or disruptive (refer to Figure 1 for a problem overview).

Despite this important nuance, existing target prediction models have not scrutinized *when* to offer a suggestion and instead used a heuristically proposed probability threshold. For example, prior work on forecasting which target a user might reach towards with their hands used a threshold of 85% because the model seemed accurate enough at that point based on their evaluation of the model confidence value over time [20]. In contrast, Huang et al. used a threshold of 43% when predicting which sandwich ingredient a user might choose via gazing [36]. They used this threshold because it was based on the average model confidence value for a correct prediction. The mixture of design intuitions and model performance observations used in this prior work may not lead to optimal suggestion timings—one may wonder if a better threshold could be chosen. Furthermore, this prior research did not consider the user-centric costs and benefits of intelligent suggestions (e.g., the exact time saved by a suggestion). Thus, this research introduces the *COBO* (cost-benefit optimization) framework, which determines the optimal timing of intelligent suggestions by considering user-centric costs and benefits. Specifically, COBO uses the probability estimates computed by a target prediction model over time as input and quantifies the cost and benefit of a suggestion to produce a final gain function. The obtained gain function then enables the determination of the most beneficial timing for suggestions either through optimization of this function or through designer's intuition.

To study how users would respond to an intelligent suggestion displayed at different timings, a dense target selection task and a text matching task were implemented in VR. VR was chosen as the testbed because VR input techniques such as mid-air pointing are effortful and are likely to benefit from intelligent suggestions. Based on the study results, cost and benefit functions were developed and simulations were run under two optimization strategies – Optimal Thresholding and Reinforcement Learning – to minimize user task completion time and maximize intelligent suggestion usage. The efficacy of these strategies was then verified in two validation experiments, which showed that COBO was helpful for determining the optimal timing of intelligent suggestions both theoretically and empirically.

The primary contributions of this research are:

- A framework (i.e., *COBO*) to optimize the timing of intelligent suggestions through a computational approach that considers user-centric costs and benefits.
- Study outcomes that demonstrate the effectiveness of COBO for intelligent suggestion timing optimization on two objectives: minimizing user task completion time and maximizing intelligent suggestion usage.

2 BACKGROUND AND RELATED WORK

This research was informed by facilitation techniques that aim to improve user performance and save user efforts in object selection tasks. It also took inspiration from works that applied probabilistic models to estimate user-intended target(s) and research that leveraged Reinforcement Learning for objective optimization in interactive applications.

2.1 Selection Facilitation Techniques

Selection facilitation techniques have been used as a method to improve interaction since the introduction of early graphical user interfaces. While numerous techniques have been proposed, the majority decrease the movement distance required to reach a target and/or increase the effective size of the target [28]. To shorten the movement distance, techniques may snap the cursor to the target (e.g., [10, 73]). To increase the target size, techniques may expand the target [44] or resize the cursor [28, 46]. A visual indicator (e.g., visual highlighting) may also provide feedback when a technique has selected a candidate object. The user can then use an explicit action (e.g., a button press) to confirm that the object that is currently selected is the one they desired to select.

Selection facilitation techniques have also been explored in VR/AR scenarios (see surveys such as [6, 42]). For example, Schjerlund et al. applied multiple virtual hands to shorten the selection distance [60] and Baloup et al. compared various raycasting-based methods that enlarged the objects' effective size in VR [11]. Selection facilitation techniques have been applied to VR/AR systems because mid-air pointing, which is a commonly used input modality in these systems for 3D input, can be inefficient and imprecise [7, 70].

More relevant to the present research are selection techniques that *predict* user-intended targets [1, 70]. In addition to decreasing target distances and increasing target sizes, prediction-based methods have also been found to reduce search time [15]. While a user may have trouble finding the intended target in more complex environments (e.g., those with lots of visual clutter), an intelligent suggestion can present a potential target to users, thus minimizing the time spent searching and manually pointing. We describe these techniques in the next section.

2.2 Target Prediction

Users' intended selection targets can be sensed through behavioral cues, such as body and eye movements. Much existing research focuses on building models that appropriate gaze traces or scanpaths to predict selection intentions [21, 37, 39, 59, 61, 72]. For example, Borji et al. [15] built models that predicted search targets based on gaze fixations on a large random-dot array. Their modeling rationale was that attention and gaze are guided toward visual features that are similar to a search target. Using this approach, they demonstrated that their models outperformed a random baseline, especially when a larger number of fixations was considered. Huang et al. [36] used a support vector machine model to predict a customer's intended target in a sandwich-making scenario and made correct estimations approximately 1.8 seconds before a customer's spoken request. Sattar et al. [58] proposed a model to predict the categories and attributes of user intended objects from gaze data, which were then used to reconstruct plausible targets. Researchers have also explored target forecasting in VR (e.g., [35]), with some taking advantage of gaze fixations to anticipate users' hand movements while reaching for objects [19, 26].

Hand and input device trajectories have also been used in selection tasks to infer user-intended targets [13, 14, 47, 74]. For example, Ahmad et al. [1–4] investigated probabilistic intent prediction approaches for in-vehicle touchscreen input based on pointing gestures. Yu et al. [70] examined the selection distribution of VR input controllers and used this information to predict the likelihood of a user selecting a candidate object. Clarence et al. [20] used long short-term memory (LSTM) models to predict the probability of selecting candidate objects using hand-reach features such as position and orientation. Researchers have also predicted future cursor positions in target-agnostic manners (e.g., [10, 30–32, 41, 43, 51, 68]).

In addition to user behaviour, models can also make use of users' preceding actions or contextual information to infer their next selection intent [27, 66, 67]. For instance, Goodman et al. [27] applied a language model for text entry to estimate the most likely selected key based on an entered sequence and the current input distribution. White et al. [67] leveraged interaction contexts such as previous search queries and clicks to predict users' short-term interests.

Although target prediction models can be effective at determining *which* object a user intends to select previous work has not examined *when* intelligent suggestion should be enabled to maximize its benefits. Some researchers have used design intuitions to trade-off between successful early predictions and the possibility of introducing false positives [1, 20, 36]. Others chose to always display a predicted target (e.g., typing predictions). However, intuitions may not lead to optimized performance and always-on, constantly changing suggestions during cursor navigation or visual search might lead to user costs that were not anticipated, especially in VR/AR scenarios where screen space is limited and distraction may be costlier. As such, our research introduces a method for optimizing the timing of intelligent suggestions that was designed to be extensible to any of these aforementioned prediction models.

2.3 Reinforcement Learning

Recently, reinforcement learning (RL) has been used in the development of adaptive user interfaces [25, 62] and human behavior simulations [18, 33]. In a typical training setting, an RL agent interacts with its environment using a set of actions and receives corresponding feedback (i.e., rewards or penalties) to help it learn from the environment [8]. Through this trial-and-error process, the agent can discover an action policy that leads to a maximized reward. Such a learning paradigm may be particularly suitable for interactive settings that incorporate human-in-the-loop [9].

HCI researchers have applied both model-based and model-free RL for interface optimization. For example, Todi et al. [62] leveraged model-based RL that utilized predictive HCI models to estimate a potential reward of an agent's action. Their model-based agent learned to adapt menu interfaces through order changing or grouping to improve user performance. In contrast, Gebhardt et al. [25] applied model-free RL to support users in a visual search task by showing and hiding object labels (e.g., price tags). Their RL agent observed user behavior (i.e., gaze trajectories) and received rewards or penalties depending on whether a label was shown when the user's gaze point was fixated on the object. Compared to model-based approaches, the model-free agent did not make predictions about the next state and reward before it took an action.

The present work employs model-free RL to discover an optimal policy of suggestion timing. Model-free RL was chosen because it does not require a transition dynamics model to derive a useful policy. The reward function integrated user-centric costs and benefits in terms of, for example, the exact time saved in seconds.

3 RESEARCH OVERVIEW

Our framework relies on quantifying user-centric costs and benefits of a suggestion over time (e.g, the exact time saved by a suggestion) to produce a final gain function for optimal suggestion timing determination. In the following sections, we introduce our framework and present three studies that aimed to demonstrate and validate the proposed framework.

The first is a user study to collect data to approximate the cost and benefit functions related to two optimization objectives (i.e., time saved and suggestion usage percentage) in a manually-intensive task and a mentally-demanding task. This is essential to complete the cost and benefit quantification step in the framework.

The second is a simulation study where simulations were run with two optimization strategies (Optimal Thresholding and Reinforcement Learning) for single- and multi-objective optimization. These simulations aimed to optimize the gain functions related to the objectives and theoretically evaluate the optimization strategies.

In the third study, the optimization findings were empirically validated by running user studies that compared the optimal timing of intelligent suggestions produced by our framework against two baselines—heuristic-based thresholding and no suggestion. The baselines help contextualize the impact of our solution relative to a literature baseline and interfaces that offer no suggestions.

4 COBO FRAMEWORK

COBO (cost-benefit optimization) is a framework to optimize *when* to display intelligent suggestions by considering the costs and benefits that an intelligent suggestion may provide to the user (e.g., the exact time saved) given specific timing and model probabilities. More precisely, COBO takes input probability estimations from a

target prediction model and user-centric costs and benefits of a suggestion over time to form a final gain function. The optimized suggestion timing is then determined by finding the maximum gain on this gain function curve (Figure 2). To apply the COBO framework, three components are needed: a target prediction model, a method for cost and benefit quantification, and a strategy for gain function optimization.

4.1 Target Prediction Model

Target prediction models are probabilistic models that infer a user's intended target of interest. A model typically produces a probability distribution $\{p_t^k\}$ among N potential candidates, which indicates the likelihood of a user selecting each candidate $k \in \mathcal{K} = \{1, ..., N\}$ at timestamp t (Figure 2 left). It may then output the most likely target and its corresponding probability value q_t (also called the model confidence). In the model, timestamp $t \in \{1, ..., T\}$, where T is the total number of timestamps that the model produces estimations since the onset of the selection until the user manually selects a target. In the present work, the target prediction models produce output at a constant frequency f. Therefore, timestamp t can be converted to time in seconds t_s using $t_s = t/f$.

The target prediction models can be trained using data collected from various information channels (e.g., user hand movement [1, 20], eye gaze information [21, 36], prior selection information [27], etc.). While the output of the target prediction model (i.e., probability estimates over time) is used as input to the COBO framework, the model itself is not a part of the framework. For simplicity, this research only displays intelligent suggestions for the most probable object. Thus, only the model confidence q_t is used as input to the COBO framework rather than the whole probability distribution. It is also assumed that model confidence is a reasonable approximation of the ground truth prediction accuracy [29, 49].

4.2 Cost and Benefit Quantification

COBO requires a quantification of the user-centric costs and benefits of displaying an intelligent suggestion over time based on the optimization objective. For example, if the objective is to minimize user task completion time, the cost and benefit quantification can use an estimation on how long it takes users to respond to suggestions, how much time a correct suggestion may save, and how much of a time delay an incorrect suggestion may cause. Such quantification can be specified from the results of empirical user studies or through literature-informed assumptions. The obtained cost function $\texttt{Cost}(t)$ and benefit function $\texttt{Benefit}(t)$ can then be used to build a final gain function.

The total gain of displaying an intelligent suggestion for the most probable object at a particular timestamp t is shown in Equation 1. The gain function is equivalent to the benefit obtained, multiplied by the probability that the predicted object is the true target minus the cost, multiplied by the probability of the object not being the real target.

$$\texttt{Gain}(t) = \texttt{Benefit}(t) \cdot q_t - \texttt{Cost}(t) \cdot (1 - q_t) \tag{1}$$

When applying the COBO framework, the gain objective can vary in different applications according to a designer's needs (e.g., minimizing completion time, minimizing induced errors, maximizing user satisfaction, etc.). This research demonstrates the optimization of two gain objectives, i.e., the time saved by users and the suggestion usage percentage.

4.2.1 Time Saved by Users. Task completion time is an obvious metric of user task performance. Ideally, an effective user interface shortens task completion time, while maintaining accuracy to increase user efficiency. To maximize time savings for users, the following three variables were considered when displaying an intelligent suggestion at timestamp t:

- Response time $\texttt{RT}(t)$: the time elapsed between the first appearance of a correct suggestion and the time when the user applies the suggestion (e.g., through a simple click).
- Response rate $\texttt{RR}(t)$: the overall user response rate to a correct suggestion.
- Delayed time $\texttt{DT}(t)$: the average time delay caused by displaying an incorrect suggestion.

For simplicity, we assume that there are minimal effects of i) the delayed time of a correct suggestion if a user does not apply it and ii) the response time of an incorrect suggestion if a user assumes it is correct.

For a given trial with total timestamps T, the potential benefit of displaying a suggestion at t is represented in Equation 2. The equation can be interpreted as the estimated timestamps saved if a correct suggestion is given at t, multiplied by their rate of response. The max function ensures the benefit value is no smaller than 0.

$$\texttt{Benefit}(t) = \texttt{max}(0, T - (t + \texttt{RT}(t))) \cdot \texttt{RR}(t) \tag{2}$$

The potential cost is the time delay caused by an incorrect prediction (Equation 3).

$$\texttt{Cost}(t) = \texttt{DT}(t) \tag{3}$$

Inserting Equation 2 and 3 into Equation 1, results in an estimated gain function that considers the timestamps saved for users (Equation 4). It can be converted to the time saved in seconds by dividing it by the model output frequency f.

$$\texttt{Gain}(t) = \texttt{max}(0, T - (t + \texttt{RT}(t))) \cdot \texttt{RR}(t) \cdot q_t - \texttt{DT}(t) \cdot (1 - q_t) \tag{4}$$

4.2.2 Suggestion Usage Percentage. Although time savings is a useful objective for performance improvement, it may not necessarily be valuable to the user experience. For example, previous work has shown that even when word prediction may impair average text entry speeds on mobile devices, users still prefer to use them [50, 54]. As such, we also sought to optimize for intelligent suggestion usage percentage. It was assumed that as long as a user applies an intelligent suggestion, it leads to a preferred user experience.

Based on this, the gain function can be written as Equation 5. The benefit function is approximated by the likelihood of users responding to a correct suggestion. For simplicity, the probability of users applying an incorrect suggestion is ignored so the cost function is omitted.

$$\texttt{Gain}(t) = \texttt{RR}(t) \cdot q_t \tag{5}$$

4.3 Gain Optimization

The value of the gain function Gain(t) changes over time such that the model confidence value q_t, the user-centric cost Cost(t), and

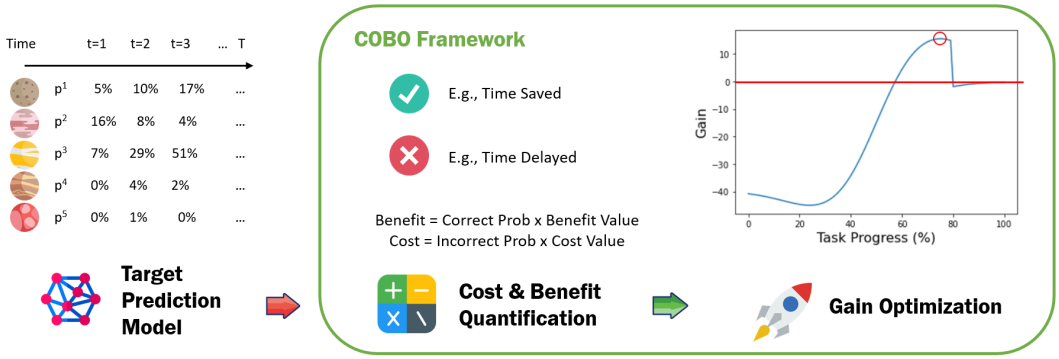


Figure 2: An overview of the COBO framework. COBO uses the probability estimates of a target prediction model as input and quantifies the cost and benefit of the suggestion over time to produce a final gain function. The gain function is computed using the benefit of displaying a suggestion minus its cost across the time axis. By applying an optimization strategy, the framework determines when displaying a suggestion will be useful (gain > 0) and when the gain value (max(gain)) will be maximized.

benefit Benefit(t) will be different as the task progresses and t increases. In real applications, the target selection model does not infer when a user starts the task ($t = 0$) or when the user finishes the task, so the task progress is unknown to the prediction model. One solution is thus to infer t from the the real-time model confidence value of the target prediction model q_t because the model tends to become more confident in its predictions as the user reaches the end of their task. Several prior studies have indicated that the relationship between t and q_t may follow a sigmoid function [20, 36], thus the implicit relationship between t and q_t can be modelled as $t = g(q_t)$. By doing this, the final objective function (Equation 6) only depends on the real-time confidence output q_t. The objective function returns the q_t that leads to the maximum gain. The returned q_t can be directly applied to determine a suggestion timing. For example, if the optimized $q_t = 0.6$, the system should display an intelligent suggestion when the model confidence reaches 0.6.

$$\underset{q_t \in [0,1]}{\texttt{argmax}} \left[\texttt{Benefit}(g(q_t)) \cdot q_t - \texttt{Cost}(g(q_t)) \cdot (1 - q_t)\right] \quad (6)$$

In practice, we obtain the mapping function $t = g(q_t)$ from a training dataset D_{train}. The purpose of D_{train} is to provide known relationship between t and q_t so that an optimization strategy can learn how to handle new real-time q_t values. In this work, we created a dataset, D_{train}, wherein each data trial consisted of known q_t values for all $t \in \{1, ..., T\}$. Such a dataset can also be generated by using a trained prediction model to produce q_t for each $t \in \{1, ..., T\}$ of the feature data (e.g., hand movement [20] or gaze information [36] over time). Once D_{train} and the cost and benefit functions are available, an optimization strategy can calculate the expected gain by simulating the effect of enabling intelligent suggestions at different q_t (which correspond to a known t) on the trials in D_{train}, to consequently compute an optimal solution over the training set. With the hypothesis that the training data is a reasonable approximation of the unseen testing data, the optimized solution can be generalized to real applications.

Since the objective is to find a q_t or a set of q_ts that can lead to the maximum gain, various optimization methods can be applied to solve this problem. In this work, two optimization strategies (i.e., Optimal Thresholding and Reinforcement Learning) were explored.

4.3.1 Optimal Thresholding (OT). The Optimal Thresholding strategy aimed to obtain a single optimized model confidence threshold that worked best on D_{train}. To achieve this aim, different confidence values $q_t \in [0, 1]$ were tested and the q_t that lead to the highest expected gain on D_{train} was selected.

4.3.2 Reinforcement Learning (RL). Rather than relying on a single threshold for all trials, RL-based optimization strategies can provide "dynamic thresholds" based on the profile of each trial (e.g., the speed of increase of the model confidence value). This has the potential to further boost the optimization performance compared to Optimal Thresholding. Therefore, RL was applied to derive optimal policies for intelligent suggestions that could reach the highest gain on D_{train}. Specifically, our RL agents observed the incoming probability estimations and explored different action sequences (i.e., displayed an intelligent suggestion or not) to ultimately find optimal action sequences that would lead to the maximum gain. Additional details about the RL agents are in Section 6.3.

5 STUDY 1 - DATA COLLECTION

The primary goal of the first study was to quantify the cost and benefit of the two optimization objectives. To this end, data was collected from participants while they responded to an intelligent suggestion displayed at different timings. Specifically, this study focused on how much time it took participants to respond to a correct suggestion, the usage percentage of the correct suggestion over time, and the trial completion delay incurred by an incorrect suggestion. Two different task scenarios (manually-intensive vs. mentally-demanding) and two different suggestion types (visual highlighting versus pop-up notification) were used to explore whether these factors would lead to different participant responses. We tested these factors because they could be the main determinants of user behavior towards an intelligent suggestion.

Figure 3: Screenshots of the dense target selection task (left) and the text matching task (right).

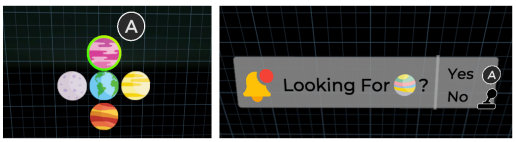


Figure 4: Highlighting notification (left) and pop-up suggestion (right) used in the dense target selection task.

We used a two-session data collection study methodology. In the first session, baseline user performance (e.g., task completion time) was collected while participants performed a dense target selection task and a text matching task. The baseline user performance was used to inform the suggestion timing interval for the second session. In the second session, correct and incorrect suggestions within the earlier determined timing intervals were displayed and the resulting participant behavioral data were recorded. This enabled the measurement of the costs and benefits of the suggestion.

We here prioritize high-level concepts and more relevant contents in our presentation. We refer readers to Appendix A for more detailed descriptions of the task scenarios and suggestion methods and the significance testing results.

5.1 Participants and Apparatus

Sixteen participants (6 women and 10 men) were recruited and provided informed consent on attending the study. Participant ages ranged from 23 to 47 ($mean = 36.6$, $std = 7.7$, one participant did not report their age). All participants had normal or corrected-to-normal vision and all were right-handed. Twelve participants had used VR devices for 0-5 hours per week, three used them for 5-10 hours, and one had never used a VR device before. As participation was remote, participants received equipment to use in the study by mail (i.e., an Oculus Quest 2, two Touch controllers, and a laptop with an GTX 1070 graphics card) and met with the researchers during a video call to complete the study.

5.2 Task Scenarios

Two task scenarios, representative of common interaction tasks that are effortful to perform, were employed (see Figure 3). The dense target selection task represented a manually-intensive task, where participants needed to select a small object located at the center of a cluster [46, 64]. The text matching task served as a mentally-demanding task, where participants needed to find and select an object with text that matched a target text. This task simulated real-world, search-heavy scenarios like searching for ingredients from a receipt, finding street names on a map, or browsing through a menu [52].

5.3 Suggestion Method

Two suggestion methods were used in the study—a highlighting suggestion and a pop-up suggestion. With the highlighting suggestion, a blinking fluorescent outline was displayed around the suggested object (Figure 4 left). The highlighting suggestion was in-situ, so it remained at the object location without following the direction participants were looking. With the pop-up notification suggestion, a suggestion window appeared at the bottom of the participant's current viewing direction (Figure 4 right) [57]. When participants rotated their viewing direction, the pop-up notification followed the viewing direction. For both suggestions, participants could quickly access the suggested object via the Button A or discard the suggestion by tilting the joystick to the right.

5.4 Study Design

The study included two sessions. The first session used a within-subject design with one factor, Task Type (dense target selection and text matching), to collect baseline user performance. Each task had 48 trials, with the first 3 trials being discarded as practice trials. The order of Task Type was counterbalanced. In total, 1440 trials were recorded (= 16 participants × 2 task types × 45 repetitions).

The second session was conducted on a later day with the same pool of participants after they had all finished the first session. It also used a within-subject design but had three factors: Task Type (dense target selection and text matching), Suggestion Method (highlighting and pop-up notification), and Suggestion Mode (correct, incorrect, and no suggestion). A suggestion, if there was one, was generated within a specific timing interval ($[0s, 3.1s]$ for the dense target selection task and $[0s, 7.6s]$ for the text matching task). The suggestion timing was then randomly sampled within this interval in each task to help us better understand how users respond to suggestions over time. The mean task completion time from the first session was used as the maximum suggestion timing for the second session, as users normally finish the task manually before these upper-bound times. The order of Task Type and Suggestion Method were counterbalanced, and the order of Suggestion Mode was randomized within each block. When a participant was working on a certain task type with a suggestion method, a suggestion may or may not appear and could be correct or incorrect. In Session 2, each condition was repeated 32 times (2 practice trials). In total, 5760 trials were recorded (= 16 participants × 2 task types × 2 suggestion methods × 3 suggestion modes × 30 repetitions).

5.5 Study Procedure

The same procedure was used for both sessions of the study. Each session started by introducing the two experimental tasks and suggestion methods (only for session 2). In session 1, participants then practiced the two tasks. In session 2, they practiced the scenarios with and without the two suggestion types in each task. The suggestion timing was shortened to 1/3 of the original intervals during practice to ensure they saw a suggestion. They then started the experiment where they were asked to complete each task as fast and as accurately as possible, and were encouraged to use intelligent

suggestions if they were correct. They were given breaks between blocks. After session 2, they completed a post-study questionnaire.

5.6 Results - Session 1

Before the baseline task completion time was computed, the data was pre-processed to remove outliers that deviated more than three standard deviations from the mean ($mean \pm 3std$). This lead to 9 trials (1.25%) being discarded for the dense target selection task and 19 trials (2.64%) being discarded for the text matching task. A total of 711 trails and 701 trials were left for analysis, respectively.

The completion times for both tasks followed log-normal distributions. Using the maximum-likelihood estimation, the calculated distribution parameters were $\mu = 1.13$, $\sigma = 0.25$ for the dense target selection task and $\mu = 1.88$, $\sigma = 0.60$ for the text matching task. Participants took an average of 3.21 seconds ($std = 0.86$) to complete the dense target selection task and an average of 7.77 seconds ($std = 4.6$) to complete the text matching task. The overall accuracies were 94.09% and 100%, respectively.

5.7 Results - Session 2

Pre-processing the session 2 data involved first discarding trials where participants completed the task before an intelligent suggestion was displayed (i.e., 222 (7.71%) dense target selection trials and 447 (15.52%) text matching tasks). Additionally, trials outside $mean \pm 3std$, were also removed (i.e., 30 (1.04%) dense target selection trials and 40 (1.39%) text matching trials). This left 2628 trials and 2393 trials, respectively, for each task for analysis. The overall accuracy for the dense target selection task was 95.09% and 99.28% for the text matching task.

5.7.1 Response Time. Response time was defined the time elapsed between the appearance of a correct intelligent suggestion and a participant's selection of that suggestion. We used multivariate adaptive regression splines (MARS) to model the relationships between suggestion timing and response time. MARS was used because it tries to find multiple linear regression lines to fit data while balancing goodness-of-fit and simplicity. The linear regression lines were connected through hinge functions ($h(x-c) = max(0, x-c)$ or $h(c-x) = max(0, c-x)$ where c was a constant called knot) to provide non-linear approximations of the data. The maximum number of terms was set to two for the robustness of the model. The resulting equations for the four conditions are summarized in Table 1. Figure 5A shows graphic illustrations of the relationship between suggestion timing and response time of two example conditions .

5.7.2 Response Rate. Response rate was defined as the likelihood that participants accepted a correct suggestion. We applied MARS to model the relationship between the response rates and suggestion timings directly. Specifically, suggestion timing was used as a predictor and the accuracy of the suggestion was as the target variable (0: incorrect, 1: correct). The regression results then approximated the percentage of participants accepting a correct suggestion over time (Figure 5B). Table 1 summarizes the corresponding MARS models.

5.7.3 Delayed Time. Delayed time was the time delay that was incurred due to incorrect suggestions. For a given trial, it was infeasible to record the task completion time both with and without a suggestion (even if we repeated the trial, factors such as learning and familiarity would differ). Therefore, this metric was computed using the task completion time of each trial with an incorrect suggestion minus the average task completion time in the corresponding condition with no suggestion. The calculated distribution then allowed us to determine the average delay an incorrect suggestion would cause across different suggestion timings (Figure 5C). The delayed time data was fit into the MARS model for each condition. The results are summarized in Table 1.

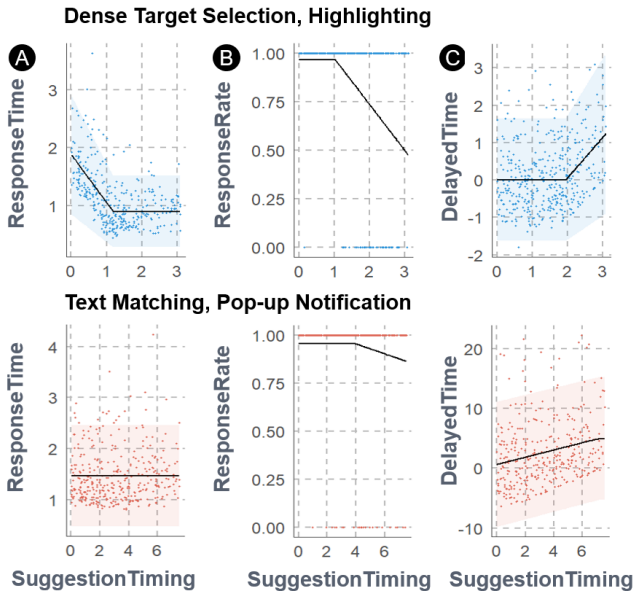


Figure 5: Examples of the modeling results for response time, response rate, and delayed time. The dots represent the data trials, the black lines are the model fitting results provided by MARS, and the ribbons indicate 95% CI.

5.8 Summary

Based on the data collection results, MARS models were able to simulate how participants would respond to an intelligent suggestion at different timings (Table 1). The models resulted in reasonable approximations of cost functions `Cost(t)` and benefit functions `Benefit(t)` for the two objectives. The gain of displaying an intelligent suggestion at timestamp t can thus be calculated via Equation 4 and 5. From the study results, it was also determined that the gain functions for the four conditions (Task Type × Suggestion Method) were quite different. Therefore, the four conditions were handled differently in later evaluations.

6 STUDY 2 - SIMULATION

The primary goal of the second study was to conduct a theoretical evaluation of the two suggestion timing optimization strategies - Optimal Thresholding (OT) and Reinforcement Learning (RL). To achieve this, a mock target prediction model that generated various data trials (D_{train}) during the two task scenarios was built. Simulations were run to estimate the gain of the optimization strategies.

Table 1: Summarization of the modeling results from MARS (multivariate adaptive regression splines).

Task Type	Suggestion Method	Response Time	Response Rate	Delayed Time
Dense Target Selection	Highlighting	$0.90 + 0.83 \cdot h(1.19 - x)$	$0.97 - 0.24 \cdot h(x - 1.02)$	$0.01 + 1.08 \cdot h(x - 1.96)$
Dense Target Selection	Pop-up Notification	$1.13 + 0.13 \cdot h(x - 1.60)$	$1.00 - 0.24 \cdot h(x - 0.98)$	$0.57 + 2.25 \cdot h(x - 2.52)$
Text Matching	Highlighting	2.91	0.90	$0.66 + 0.84 \cdot h(x - 1.29)$
Text Matching	Pop-up Notification	1.47	$0.96 - 0.03 \cdot h(x - 3.90)$	$4.94 - 0.61 \cdot h(7.13 - x)$

To constrain the search space, the study focused on applying highlighting suggestions for the dense target selection task, as it was less intrusive, and using pop-up notifications for the text matching task, as it led to quicker responses.

The following subsections first present the mock target prediction model that was used to generate D_{train} and then introduce the four simulation experiments that were undertaken. In Simulation 1, the performance of OT was bench-marked for the time saved for participants versus the suggestion usage percentage. The performance of the baselines that leveraged the design heuristics were also used to determine thresholds. In Simulation 2, RL was applied for optimization. In Simulation 3, multi-objective optimization (i.e., time saved and usage percentage) was run with OT.

6.1 Target Prediction Model Mock-up

As most models' prediction accuracy values seem to follow sigmoid curves over task progression (e.g., [2, 15, 20, 36]), we simulated a similar model by mimicking the observed sigmoidal relation between accuracy and time to generate D_{train}. Specifically, for each trial, we first sampled trial length T based from the log-normal distribution found in the first session of Study 1 (Figure 6A). Then, a sigmoid function of task progression regarding prediction accuracy was computed (Figure 6B-C) and deviations (i.e., spikes and dips) were added to the sigmoid function (Figure 6D). More details of this mock-up target prediction model can be found in Appendix B.1.

The mock-up target prediction model was limited in that it only mimicked the appearance of the confidence curves, so it did not capture the inherent decision information of a real target prediction model. However, if the optimization strategies worked with a pseudorandom model, then they may also work with an actual target prediction model. Next, we present simulation results based on 30,000 trials generated by the mock-up prediction model for each task scenario. The trials were separated such that 90% were used for training and 10% were used for testing. Among the training data, 10% was used for hold-out validation. We present only testing results in the paper while readers can find the validation results in Appendix B.2.

6.2 Simulation 1: Optimal Thresholding

The Optimal Thresholding (OT) strategy sought to learn an optimized confidence threshold from the dataset that would lead to the best gain. To achieve this, different confidence values were tested ($q_t \in [0, 1]$, 0.01 per step) and the corresponding gain was calculated using Equation 4 and 5 from the first study. Figure 7 presents two examples of how the gain in the time saved condition changed as the confidence threshold q_t varied. The optimized threshold was quite different for the dense target selection task ($thres = 0.47$) compared to the text matching task ($thres = 0.98$).

Table 2: Testing results when using Optimal Thresholding (OT) and Heuristic Thresholding (HT) regarding the dense target selection (DTS) task and the text matching (TM) task.

	Task	Strategy (Th.)	Time Saved/Usage%	% Improved
Time saved	DTS	OT (0.47)	0.4073s (0.3202s)	39.39%
	DTS	HT (0.85)	0.2922s (0.3645s)	-
	TM	OT (0.98)	1.6211s (1.7946s)	260.89%
	TM	HT (0.50)	0.4492s (1.1440s)	-
Usage %	DTS	OT (0.81)	65.69% (18.30%)	0.36%
	DTS	HT (0.85)	65.45% (20.42%)	-
	TM	OT (0.96)	87.17% (18.53%)	51.52%
	TM	HT (0.50)	57.53% (15.63%)	-

To benchmark the performance of OT, we picked a threshold that worked the best on the validation dataset and produced the corresponding results on the testing dataset. The baseline (i.e., Heuristic Thresholding) for the dense target selection task was determined to be $thres = 0.85$, which was directly appropriated from a similar point-and-select task in the literature with sigmoidal prediction curves [20]. The baseline for the text matching task was $thres = 0.50$, which was used to predict participant selections in a search-intensive task like our text matching task (i.e., users' intended ingredients in a sandwich-making task [36]).

From the results, the optimized threshold was found to save 0.1 seconds more than the baseline in the dense target selection task (around 40% of improvement) and 1 second more than the baseline in the text matching task (around 260% of improvement; Table 2). The optimized threshold also led to an 87% suggestion usage percentage in the text matching task (around 50% of improvement). The optimized thresholds were quite different for the dense target selection task for the time saving optimization ($thres = 0.47$) and usage percentage optimization ($thres = 0.81$), while being similar for the text matching task (0.98 *vs.* 0.96).

6.3 Simulation 2: Reinforcement Learning

RL can potentially provide tailored solutions based on the target prediction confidence profile of each trial (e.g., the speed of increase of the model confidence value) by finding an appropriate threshold to display suggestions that works for that specific profile. To achieve this, model-free RL techniques were leveraged because there was a lack of transition dynamics models for our problem. Thus, the model-free RL agents observed the model confidence estimates q_t from a target prediction model trained on D_{train}, which were replayed multiple times to the agent. On each trial, the agent

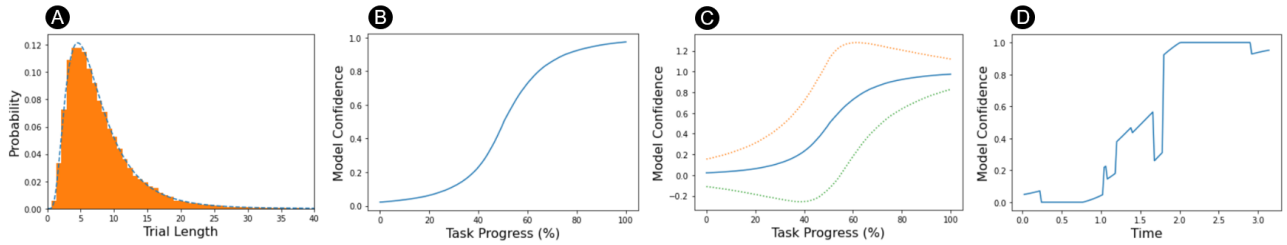


Figure 6: The trial generation process for the mock-up target prediction model. (A) The model first computes the trial length based on the log-normal distribution found in Study 1 for task completion time. (B) The model forms a sigmoid function of task progression with respect to prediction accuracy. (C) The sigmoid function varies within a predefined region (the dashed lines indicate the 95% CI). (D) The model adds deviations (i.e., spikes and dips) to the trial.

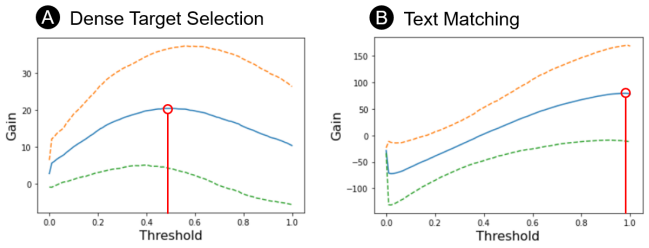


Figure 7: The expected gain for maximizing time savings for participants (y-axis) when using different confidence thresholds (x-axis) based on the validation dataset. The unit of gain is a timestamp, where the time saved in seconds equals 0.02 · timestamps. Dashed lines represent $mean \pm std$**.**

explored different action sequences (i.e., displayed an intelligent suggestion or not) to ultimately find the optimal action sequence for a given q_t trajectory that would lead to the maximum gain.

6.3.1 Problem Formulation. The key elements of the RL agents were:

- *Observation*: For a specific timestamp t, the agent received the following observation $\{p_1, p_2, ..., p_m, d_t\}$. The probability values $\{p_1, p_2, ..., p_m\}$ were the model confidence values produced by the target prediction model over time. The integer m was the memory size of the agent. The list acted like a first-in-first-out queue where p_m represented the most recent confidence value provided by the prediction model and p_1 represented the least recent. The float d_t recorded the last timestamp when a suggestion was displayed.
- *Action*: The agent could take the following two actions based on the observation $\{display, not\ display\}$. The *display* action represented displaying an intelligent suggestion, so d_t was updated to the current timestamp t. The *not display* action hid the suggestion.
- *Reward*: Three reward settings were used to train the RL agents. The first was r_1, where $r_1^t = \texttt{Gain}(t)$ if a suggestion was displayed at t, otherwise $r_1^t = 0$. The second reward setting, r_2, sought to solve the reward sparsity issue in r_1. Specifically, reward shaping was performed when the suggestion wasn't displayed: $r_2^t = \texttt{Gain}(t)$ if a suggestion was displayed at t, otherwise $r_2^t = -k \cdot p_m$. We used a hyper-parameter k to penalize the action of not displaying any suggestion. An agent received more penalties if it did not display a suggestion when the model confidence was high (p_m). The third reward r_3^t also leveraged the benefit of dense rewards, but removed the agents' reliance on the penalty factor k, which may have negative impacts on true reward maximization. Here, $r_3^t = \texttt{Gain}(t) - r_3^{t-1}$ (where $r_3^0 = 0$) at a timestamp t. This essentially rewarded the agent based on how good it performed on a particularly timestamp t, by computing the contribution of agent's action at t towards the gain. More details can be found in Appendix B.1.4.
- *Episode End Criteria*: The current episode ended if t was larger than the maximum length of the trial T, or d_t was larger than 0 (which meant a suggestion was displayed).
- *Initialization*: p_m was initialized to the first confidence value produced by the target prediction model, while all other probability values were set to 0. d_t was initialized to 0.

Table 3: Testing results of RL regarding regarding the dense target selection (DTS) task and the text matching (TM) task.

Task	Strategy	Time Saved	% Improved
DTS	PPO-MLP	0.4087s (0.3285s)	39.87%
DTS	ACER-LSTM	0.4084s (0.3362s)	39.77%
TM	PPO-MLP	1.6050s (1.7877s)	257.30%
TM	ACER-LSTM	1.5671s (1.7328s)	248.86%
Task	**Strategy**	**Usage%**	**% Improved**
TM	PPO-MLP	87.31% (18.05%)	51.76%

6.3.2 Methodology. OpenAI Gym [16] and Stable Baselines [34, 56] were used to build and train the RL agents. Our experimentation demonstrated that PPO2 with MLP policies was a lightweight and effective solution and ACER with LSTM was powerful but may take longer to train. We thus used these two strategies for final benchmarking. Since training these RL agents consumes a lot of resources, for demonstration purposes, we only optimized agents for the time saving objective and one agent for the usage percentage objective. More training details can be found in Appendix B.1.5.

6.3.3 Results. The results showed that RL agents could provide around 40% of improvement in the dense target selection task and

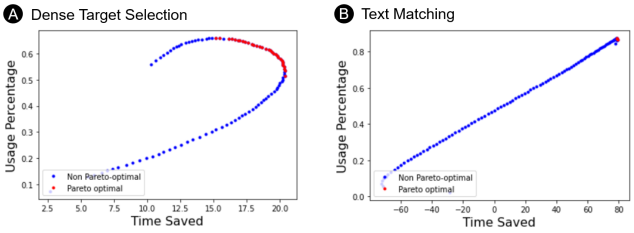


Figure 8: Optimizing both the time saved for participants and the suggestion usage percentage with Pareto Frontier-based multi-objective optimization.

260% of improvement in the text matching task as compared to Heuristic Thresholding (Table 3). Compared to the results from Section 2, OT and RL led to very similar performance improvement in the two task scenarios; while RL did produce dynamic thresholds for each trial. We will return to this in later sections (Section 7.2.3 and 8.2).

6.4 Simulation 3: Multi-Objective Optimization

So far, our approach focused on optimizing a single objective e.g., time saved or usage percentage. However, designers may need to find optimal decisions in the presence of trade-offs between two or more conflicting objectives (e.g., minimizing task completion time and maximizing accuracy) in many applications. Multi-objective optimization is useful in such settings, when more than one objective function need to be optimized simultaneously. Therefore, we explored Pareto Frontier-based multi-objective optimization technique [45, 48], which generates a set of acceptable trade-off optimal solutions, to optimize the two objectives—time saved and suggestion usage percentage simultaneously.

A given condition is called Pareto optimal if one dimension (i.e., objective) could not be improved without worsening other dimensions (i.e., objectives). In our case, we computed the gain of timing saving and usage percentage for each condition q_t and plotted them on a two dimensional xy-plane (Figure 8). A Pareto optimal point was identified if there was no point on the plane that was better in both x and y dimensions. The corresponding threshold q_t of the point was then retraced.

Following the above method, we identified thresholds that could optimize both objectives simultaneously. Thirty-two Pareto optimal values were identified for the dense target selection task ($Thres = 0.47, 0.50 - 0.78, 0.80 - 0.81$) and three Pareto optimal values were identified for the text matching task ($Thres = 0.96 - 0.98$). The results indicated that the time saved and usage percentage objectives were somewhat conflicting in the pointing task but not in the text matching task. Thus COBO can help practitioners who want to trade-off various optimization objectives.

6.5 Summary

These simulation experiments demonstrated different facets of optimization strategies using COBO. The experiments showed how, theoretically, OT and RL were both effective at determining the optimal timings at which to show an intelligent suggestion, while the performance difference between the two strategies was small. We found that for the dense target selection task, an intelligent suggestion should be displayed when the model confidence reached 0.47 for optimizing time saved for users and 0.81 for optimizing suggestion usage percentage. For the text matching task, an intelligent suggestion should be displayed when the model confidence reached around 0.96-0.98 for optimizing both objectives.

It was also found that a non-optimized threshold could lead to much worse performance (e.g., 1 second longer in task completion time and a 30% smaller suggestion usage percentage in the text matching task) compared to an optimized strategy based on COBO. Not all intelligent suggestions were shown to be beneficial, however. Displaying suggestions early in the text matching task lead to a negative gain in terms of task completion time.

7 STUDY 3 - VALIDATION

The third study consisted of two empirical user experiments of COBO because of the high number of conditions. The first one compared the time saved and suggestion usage % for Optimal Thresholding (OT) and Heuristic Thresholding (HT), finding that OT saved participants more time and led to a higher suggestion usage percentage in the text matching task. The second experiment compared OT and RL strategies and found that OT and RL lead to similar performance.

7.1 Validation 1 - Optimal Thresholding vs. Heuristic Thresholding

The goal of validation experiment 1 was to empirically verify the effectiveness of Optimal Thresholding in comparison with Heuristic Thresholding. We also included a No Suggestion condition to help contextualize the impact of suggestion conditions relative to when the interface offers no suggestions.

7.1.1 Participants and Apparatus. Another 26 participants were recruited (i.e., fourteen women, eleven men, and one non-binary). Their ages ranged from 22 to 65 ($mean = 36.1, std = 12.8$). All participants had normal or corrected-to-normal vision and were right-handed. 23 participants had used VR devices 0-5 hours per week, two used 5-10 hours per week, and one had never used any VR device before. The same apparatus was used as in the first study.

7.1.2 Methodology. Participants experienced both task scenarios (i.e., dense target selection and text matching). There were four conditions (Strategy) for the dense target selection task: optimized thresholds for time saved (*TS*, $thres = 0.47$), optimized thresholds for suggestion usage percentage (*UP*, $thres = 0.81$, which was close to HT $thres = 0.85$ from a selection task [20]), balanced optimization for both objectives (*BA*, $thres = 0.64$), and no intelligent suggestions (*NS*). Similar to Study 2, we used highlighting suggestions for the dense target selection task.

There were three conditions (Strategy) for the text matching task: balanced optimization based on OT (*BA*, $thres = 0.97$), HT baseline (*HT*, $thres = 0.50$ from a search-heavy, mentally-demanding task [36]), and no intelligent suggestions (*NS*). The time saved ($thres = 0.98$), suggestion usage percentage ($thres = 0.96$), and balanced ($thres = 0.97$) optimization conditions were combined in this task as the thresholds were very close. We used pop-up notifications for the text matching task. This design enabled us to

investigate multiple factors while keeping the study size reasonable at seven experimental conditions.

48 trials of predictions were generated for each task scenario using the mock-up target prediction model from Study 2. Each trial contained the probability of the model making a correct suggestion (i.e., model confidence) over a fixed period of time. The different thresholding strategies were then applied to each trial to decide the timing of showing a suggestion. The 48 trials were fixed across conditions to minimize the variances caused by the target prediction model. The average global centerline of the 48 trials followed a sigmoid curve. The final correctness of the prediction (i.e., a predicted candidate which participants visually perceived) was determined based on the confidence value when displaying a suggestion. For example, if a strategy decided to display the suggestion when the confidence value was 0.6, the final prediction then had 60% chance to be correct. Among the 48 trials, the first 3 trials were treated as practice trials. In total, 8190 trials were recorded (= 26 participants × 7 conditions × 45 repetitions) during this experiment.

A similar experimental procedure was employed as the first study. However, in this study, after completing each condition, participants were asked to complete a questionnaire that had three 7-point Likert scale questions probing easement, physical workload, and mental workload. The order of the task scenarios was randomized and the conditions within the scenarios were counterbalanced. The order of the formal trials were also randomized, however, the practice trials were always the same.

7.1.3 Analysis and Results. While data was initially collected for 26 participants, P1, P14, P19, and P26 were excluded as they never used intelligent suggestions in one or both of the tasks. The trials where participants had finished before the suggestion appeared (i.e., 169 (3.61%) dense target selection trials and 518 (14.76%) text matching tasks) were removed from the dataset. Because a mock-up target prediction model was used, there could have been trials where participants finished earlier than the pre-determined time period. Thus, only trials where an intelligent suggestion was displayed were considered. We also removed outliers ($mean \pm 3std$) (i.e., 45 (0.96%) dense target selection trials and 42 (1.20%) text matching tasks). These pre-processing steps resulted in 6156 trials remaining for analysis (i.e., 3746 trials for dense target selection and 2410 trials for text matching). The trials were later averaged across participant and condition. The overall accuracy was 93.14% for the dense target selection task and 98.98% for the text matching task.

For the dense target selection task, a linear mixed model with arcsinh transformation (as determined by the `bestNormalize` package) suggested that Strategy had a significant main effect on task completion time ($F = 5.02, p = .003$). A post-hoc analysis with Bonferroni correction showed that the completion time in *NS* was significantly longer than *BA* ($p = .002$), and marginally significant longer than *TS* ($p = .084$) and *UP* ($p = .135$) (all other $p > .887$). Another linear mixed model with exp transformation indicated that Strategy had a significant main effect on suggestion usage percentage ($F = 65.69, p < .001$). Post-hoc analysis suggested that usage percentages of *UP* ($p = 0.056$) and *BA* ($p = 0.140$) were marginally significant higher than *TS*. See Figure 9A-B for an overview.

For the text matching task, a linear mixed model with sqrt transformation suggested that Strategy had a significant main effect on task completion time ($F = 59.79, p < .001$). A post-hoc analysis showed that participants performed significantly faster in *BA* than *HT* ($p < .001$) and *NS* ($p < .001$). *HT* was also found to have a significantly shorter task completion time than *NS* ($p < .001$). Another linear mixed model with exp transformation suggested that Strategy had a significant main effect on suggestion usage percentage ($F = 420.45, p < .001$). A post-hoc analysis indicated that *BA* had a significantly higher suggestion usage percentage than *HT* ($p < .001$). See Figure 9C-D for an overview.

For the subjective questions, pair-wise comparisons (with Bonferroni correction) identified that *BA* led to lower mental workload ($p = .012$), and were possibly easier to use ($p = .053$), than *NS* in the text matching task. This suggests that using an intelligent suggestion could reduce workload and improve user experience.

7.1.4 Discussion. The empirical results demonstrated the effectiveness of the COBO optimization framework for the text matching task. As expected from the theoretical evaluation, the optimized condition (*BA*) led to shorter task completion times and higher suggestion usage % than the baseline conditions (*HT* and *NS*).

The benefits due to COBO were more obvious in the text matching task compared to the dense target selection, mainly because the dense task was very rapid and, as such, it was more difficult to have substantial differences in suggestion timings (thus their effect on time saved for users and suggestion usage percentage). However, the patterns across the two tasks were consistent. The significantly higher suggestion usage in text matching, in particular, could be impactful in lowering user's effort, which is suggested in the lower mental load scores of the balanced optimization.

7.2 Validation 2 - Optimal Thresholding vs. RL

The primary goal of the validation experiment 2 was to compare Optimal Thresholding (OT) vs. RL strategies for time saved and suggestion usage percentage. Based on the findings from validation 1, in this study, only the text matching task was used, as it was more likely to lead to verifiable performance differences in an empirical user study than the dense target selection task.

7.2.1 Participants and Apparatus. 12 participants (6 women, 5 men, and 1 non-binary) who had participated in the first validation study were recruited for the second validation study. Since the time interval between validation experiment 1 and 2 was more than a week and the strategy differences were hard to verify by seeing only the suggestion itself, it was presumed to be reasonable to reuse participants. Participants' age ranged from 22 to 63 ($mean = 35.9, std = 10.9$). The same apparatus were used as in validation study 1.

7.2.2 Methodology. The study employed a 2 × 2 within-subject design: Objective (time saved and suggestion usage percentage) × Strategy (OT and RL). Based on Study 2, $thres = 0.98$ was used for time saved optimization and $thres = 0.96$ was used for suggestion usage percentage optimization. The PPO-MLP agent from Study 2 was used.

The same 48 trials were used to generate the corresponding suggestion timing in each condition, and a similar study protocol was employed as validation study 1. In total, 2160 trials were collected (= 12 participants × 2 objectives × 2 strategies × 45 repetitions).

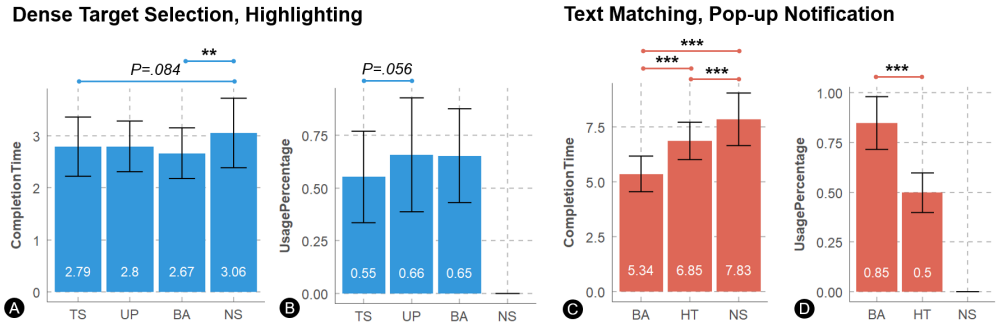


Figure 9: Results of average task completion time and suggestion usage percentage in the first validation experiment of Study 3. The four conditions in the dense target selection task were time saved optimization (TS), usage percentage optimization (UP), balanced optimization (BA), and no suggestion (NS). The three conditions in the text matching task were balanced optimization (BA), Heuristic Thresholding (HT), and no suggestion (NS). The error bars represent $mean \pm std$. ** means $p < .01$ and * means $p < .001$.**

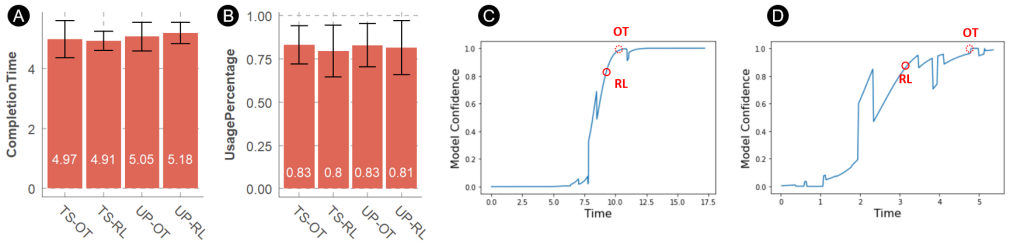


Figure 10: Results of average task completion time (A) and suggestion usage percentage (B) in the second validation experiment of Study 3. The four conditions were usage percentage optimization with RL (UP-RL) and Optimal Thresholding (UP-OT) and time saved optimization with RL (TS-RL) and Optimal Thresholding (TS-OT). The error bars represent $mean \pm std$. (C) and (D) show example trials where RL and Optimal Thresholding (OT) yielded noticeably different suggestion timings. On average, RL saved 0.31s less in (C) and 1.79s more in (D) than OT.

7.2.3 Analysis, Results, and Discussion. After removing outliers ($mean \pm 3std$, 11 trials, 0.51%) and trials where participants finished before the suggestion appeared (709 trials, 32.8%), 1440 trials remained for analysis. The overall accuracy was 99.59%.

A linear mixed model with sqrt transformation was not able to identify that Strategy had a significant main effect on task completion time ($F = 0.18, p = .674$). Another linear mixed model with Yeo-Johnson transformation was not able to identify that strategy had a significant main effect on suggestion usage percentage ($F = 0.74, p = .397$). Strategy was not shown to have significant main effects on any of the subjective scales. In summary, our results did not find any significant differences between OT and RL that lead to identifiable differences in the optimization metrics (Figure 10A-B).

We were further interested to see whether RL proposed different suggestion timings than OT in the 48 trials. For the time saved optimization, RL and OT led to a similar suggestion timing ($\Delta <$ 0.1s) in most cases (72.9%). For 16.8% of the cases, the difference between them was > 0.5s. For usage percentage optimization, there were 68.8% trials where RL and OT led to a similar suggestion timing ($\Delta <$ 0.1s) and 8.3% trials that resulted in difference > 0.5s. In the trials with difference >0.5s, RL *always* attempted to display an earlier suggestion to save more time for users. On average, RL showed the suggestions 0.79s ($std.$ = 0.38s) earlier in these trials as compared to OT.

Figure 10C-D demonstrate two examples wherein RL finds different thresholds than OT. RL strategy seems to be observing the trend of the model confidence curve and displaying a suggestion once the curve is likely to plateau in the near future. Figure 10C shows a trial where RL saved 0.31 less than OT on average, and Figure 10D shows a trial where RL saved 1.79s more. Thus, RL is certainly able to learn a strategy that results in dynamic thresholds that match OT performance on average, but it remains to be seen if/when RL may be able to outperform optimal thresholds.

8 DISCUSSION

We've conducted a series of three studies that demonstrated the theoretical and empirical effectiveness of our COBO (cost-benefit optimization) framework for suggestion timing optimization. In this section, we further reflect on our experiences in terms of the cost and benefit quantification of the two optimization objectives and the strength of RL as an optimization strategy as compared to Optimal Thresholding. We also discuss the generalizability of the framework to other applications and the limitations of our studies.

8.1 Optimization Objectives

Our work demonstrates a successful optimization of two objectives: time saved by users and suggestion usage percentage. The COBO framework is designed to help optimize various objectives, either individually or simultaneously, as long as a cost and benefit quantification method can be determined. We used data collected from participants (Study 1) to construct cost and benefit functions with variables such as response times, response rates, and delayed times. The validation studies indicated that the constructed cost and benefit functions were good approximations of the ground truth.

The time savings in our case, even though significant, are small especially in the dense target selection task. However, existing work has shown that users prefer intelligent suggestions despite negative time costs [54] because they were considered less physically demanding and effortful. The fact motivated us to quantify the benefit of intelligent suggestions beyond performance improvements. While usage percentage is an effective proxy that assumes that higher suggestion usage is always beneficial for a user to lower their interaction friction [38], a highly promising avenue for future work is in optimizing directly for effort, physical and mental-demand especially as we become better at real-time estimations of quantities like arm fatigue [18] and satisfaction [24, 53].

For simplicity, we omitted some rare conditions during cost-benefit quantification. For example, we excluded the trials where users mistakenly triggered the selection of an incorrect suggestion. Such instances were very uncommon (0.4% overall) and did not significantly impact the suggestion usage percentage or the time cost of an incorrect suggestion. However, future endeavors can extend our framework to consider mistaken triggering of an incorrect suggestion especially if those instances are not rare and/or if they require a costly recovery from the mistake [12, 40]. One simple way might be to consider modeling this as a constant time cost (e.g., recovery time).

8.2 RL as an Optimization Strategy

We found that RL was able to learn a successful strategy and produce dynamic thresholds across trials. However, RL's dynamic thresholds weren't able to outperform the single optimal threshold on average in our simulation and validation study.

As we report, there were a small, but significant percentage of trials where RL's suggestion timing differed by >0.5s compared to OT. However, we did not find any big discernible patterns in these trials compared to others. It will be worth investigating task contexts where the percentage of such trials is higher. Another reason for RL's similar performance to OT might be that the room of improvement for RL was small, as Optimal Thresholding (OT) already performed very well. The analysis demonstrated that even the theoretical maximum of a perfect agent (i.e., agent that maximizes the gain by knowing the whole trial profile) can lead to no larger than 0.18s and a 4.3% improvement over OT in task completion time and suggestion usage percentage, respectively, with our dataset. It will be interesting to see if there are contexts where OT does not achieve performance close to the theoretical maximum.

We can propose two variables to explore here that may help diversify our task context. First, is to look at trials with durations that are much more variable. Looking at the validation study data more closely, we found a weak correlation between the time-saving differences ($RL - OT$) and trial length ($R^2 = 0.10$) which indicated that the RL agent saved more time than OT in longer duration trials. Second, is to look at target prediction models that are not sigmoidal in nature (as an example, models that start with a high prior confidence using earlier user activity), and may follow patterns that cannot be easily captured using a single OT.

RL may also prove to be useful in scenarios where an interface wants to show more than one intelligent suggestion and the suggestions get updated based on users' behavior. It might be hard to directly apply OT in these scenarios. Also, in case a designer wants to enable different suggestion types within the same task (example, both highlighting and pop-up notification), an RL agent could choose the most appropriate suggestion type based on the gain of those options at different timings. An interesting area of exploration is the long-term use of such intelligent suggestion interfaces. A user may form an expectation of how well the model performs, which can in turn influence their response behavior, thus changing the cost-benefit quantification over time. An online RL agent may also prove useful in such cases.

8.3 Applications

This research has demonstrated the application of COBO in two task scenarios (dense target selection and text matching) and two objectives (minimizing user task completion time and maximizing intelligent suggestion usage). The two tasks and suggestion types were intentionally chosen to be representative of popular use cases. The dense target selection task aims to simulate physically-demanding tasks where users need to select objects in cluttered environment [46, 64], and the text matching task mimics real-world search-heavy scenarios such as searching for ingredients from a receipt [25, 63]. Object highlighting and pop-up notification are both common visualizations to inform users about system events [57]. Additionally, in Appendix B.3, we also present results on successfully applying COBO on a dataset from the literature which records hand movement trajectories when reaching virtual objects at different locations. We further envision COBO being extensible to other tasks and facilitation.

8.3.1 Extending to other tasks. The framework can be retrained for other applications that want to leverage intelligent predictions using target prediction models that rely on hand, head, gaze, and other contextual information [31, 70] in selection tasks such as pointing, visual search, and text-entry. By following the COBO framework, practitioners may choose different models, objectives, and cost-benefit quantification methods which are tailored for their applications. Overall, based on our user-centric computational framework, designers are more likely to provide intelligent suggestions that support their intended goals, rather than leading to unexpected outcomes [50, 54].

8.3.2 Extending to other facilitation. COBO's framework can also be extended to facilitate techniques other than intelligent suggestions such as expanding [44] or auto-selecting [1, 4] a predicted target, as well as for more than one suggestions simultaneously or sequentially.

8.4 Target Prediction Model

The current research builds on certain assumptions to simplify the complex problem space. One assumption is the use of a mock-up target prediction model, as we wanted to simulate a highly representative prediction model, rather than choosing one at random. Therefore, we carried out a literature survey to extract the commonalities among prediction models and then created a simulation from those commonalities (Section 6.1). However, our inspirations were from human behavior models of target reaching [20] and searching [36] where the model prediction accuracy was typically high during the later stage of the task because the selection indicator (e.g., hand or gaze point) was "approaching" or "almost on" the target and the user was just "fine-tuning" the selection of the target. For example, in the text matching task, we imagined that the gaze direction would reach the targeted object way before the controller-based manual pointing selection (i.e., the model has very high confidence based on gaze features no matter the position of the hand pointer), as Huang et al. [36] could correctly anticipate the intended object through gaze sequences 1.8s before a speech request. We acknowledge that there are other types of models that may not have such rich features. Future work can deploy this framework to any prediction model to test it on new use cases. This, however, did mean that the intelligent suggestions were not delivered dynamically based on a user's behaviour. For experimental control, it was important that this be the case while developing and validating the COBO framework. However, future research should investigate how the framework responds to a real prediction model.

One additional consideration of the current approach is that it requires a dataset of model confidence curves to calculate user-centric costs and benefits over time. In a real scenario where a designer has a target prediction model and its training dataset, the training dataset should contain trials with necessary features (e.g., user behavior data, completion times) so the designer can directly use those for confidence curve generation and cost-benefit computation (see Appendix B.3 for an example). In a condition where the feature dataset is missing, another possible solution is to apply models to simulate user behavior. During the planning phase of this research, our initial idea was to use existing computational models (e.g., minimum jerk model) to generate a large volume of user behavior data. However, we encountered two challenges. First, we did not know how users would behave according to correct/incorrect suggestions that appeared at different timings (so it was hard to incorporate this element into the model). Second, a user behavioral model for the text matching scenario is still largely underexplored (unlike bio-mechanical behavior modeling for pointing and reaching as in Cheema et al. [18] and Fischer et al. [23]). Therefore, we decided to collect new data from real users. However, we do believe using model-generated datasets for user cost-benefit quantification can be helpful in the future with more advances in the field.

9 CONCLUSION

Predictive systems are helpful ways to lower input friction and improve user experiences in current VR/AR systems [38]. Specifically, selection facilitation techniques that leverage target prediction models can alleviate the need for manual pointing and visual search, and can potentially lead to quicker, easier, and more comfortable interaction. While current target prediction models only offer *which* target a user intends to select, we built a framework (COBO) that helps determine *when* an intelligent suggestion should be displayed to maximize its benefits.

COBO is a computational framework that determines the optimal timing of an intelligent suggestion for each interaction based on user-centric costs and benefits. In a set of studies, we demonstrated that COBO is effective at determining the optimal timing of intelligent suggestions. The first study focused on measuring and quantifying the costs and benefits of an intelligent suggestion displayed at different timings when trying to satisfy two objectives (i.e., time saved for users and suggestion usage percentage) during two tasks (i.e., dense target selection and text matching). We then run simulations with two optimization strategies (i.e., Optimal Thresholding and RL) for single- and multi-objective optimizations. We found both Optimal Thresholding and RL led to better performance compared to heuristic-based thresholding approaches. For example, both optimization strategies led to around 40% improvement in terms of task completion time in the dense target selection task and 260% improvement in the text matching task. We also demonstrated the effectiveness of COBO for multi-objective optimization. The third study contained two validation experiments that compared Optimal Thresholding, RL, heuristic-based thresholding, and no suggestion conditions. The experimental results suggested that COBO-based optimization strategies led to shorter task completion times and higher suggestion usage percentages, and were preferred by participants in the text matching task when compared to baselines.

From both theoretical and empirical perspectives, we showed that an optimized strategy based on COBO can perform significantly better than non-optimized heuristic-based approaches in maximizing the time saved by users and increasing suggestion usage percentages. Overall, we envision the introduced framework will unlock effective intelligent suggestions, which will benefit future predictive systems.

ACKNOWLEDGMENTS

We thank Ben Lafreniere, Kashyap Todi, and many others from Meta Reality Labs Research for insightful discussions. We also thank Michael Frederick and other team members for their help with user studies. Icons in figures are from Flaticon.com and 105 Colorful 2D Planet Icons in Unity Asset Store.

REFERENCES

[1] Bashar I. Ahmad, Patrick M. Langdon, Simon J. Godsill, Richard Donkor, Rebecca Wilde, and Lee Skrypchuk. 2016. You Do Not Have to Touch to Select: A Study on Predictive In-Car Touchscreen with Mid-Air Selection. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Ann Arbor, MI, USA) *(Automotive'UI 16)*. Association for Computing Machinery, New York, NY, USA, 113–120. https://doi.org/10.1145/3003715.3005461

[2] Bashar I. Ahmad, Patrick M. Langdon, Simon J. Godsill, Robert Hardy, Eduardo Dias, and Lee Skrypchuk. 2014. Interactive Displays in Vehicles: Improving Usability with a Pointing Gesture Tracker and Bayesian Intent Predictors. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seattle, WA, USA) *(AutomotiveUI '14)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2667317.2667413

[3] Bashar I Ahmad, James Kevin Murphy, Simon Godsill, Patrick M Langdon, and Robery Hardy. 2017. Intelligent interactive displays in vehicles with intent

prediction: A Bayesian framework. *IEEE Signal Processing Magazine* 34, 2 (2017), 82–94. https://doi.org/10.1109/MSP.2016.2638699

[4] Bashar I Ahmad, James K Murphy, Patrick M Langdon, Simon J Godsill, Robert Hardy, and Lee Skrypchuk. 2015. Intent inference for hand pointing gesture-based interactions in vehicles. *IEEE transactions on cybernetics* 46, 4 (2015), 878–889. https://doi.org/10.1109/TCYB.2015.2417053

[5] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2623–2631.

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

[7] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5643–5654. https://doi.org/10.1145/3025453.3025474

[8] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. 2017. A brief survey of deep reinforcement learning. *arXiv preprint arXiv:1708.05866* (2017).

[9] Christian Arzate Cruz and Takeo Igarashi. 2020. A Survey on Interactive Reinforcement Learning: Design Principles and Open Challenges. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1195–1209. https://doi.org/10.1145/3357236.3395525

[10] Takeshi Asano, Ehud Sharlin, Yoshifumi Kitamura, Kazuki Takashima, and Fumio Kishino. 2005. Predictive Interaction Using the Delphian Desktop. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (Seattle, WA, USA) *(UIST '05)*. Association for Computing Machinery, New York, NY, USA, 133–141. https://doi.org/10.1145/1095034.1095058

[11] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. *RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300331

[12] Nikola Banovic, Tovi Grossman, and George Fitzmaurice. 2013. The Effect of Time-Based Cost of Error in Target-Directed Pointing Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1373–1382. https://doi.org/10.1145/2470654.2466181

[13] Xiaojun Bi and Shumin Zhai. 2013. Bayesian Touch: A Statistical Criterion of Target Selection with Finger Touch. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 51–60. https://doi.org/10.1145/2501988.2502058

[14] Pradipta Biswas, Gokcen Aslan Aydemir, Pat Langdon, and Simon Godsill. 2013. Intent recognition using neural networks and Kalman filters. In *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*. Springer, 112–123. https://doi.org/10.1007/978-3-642-39146-0_11

[15] Ali Borji, Andreas Lennartz, and Marc Pomplun. 2015. What do eyes reveal about the mind?: Algorithmic inference of search targets from fixations. *Neurocomputing* 149 (2015), 788–799. https://doi.org/10.1016/j.neucom.2014.07.055

[16] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. Openai gym. *arXiv preprint arXiv:1606.01540* (2016).

[17] Tom Chandler, Maxime Cordeil, Tobias Czauderna, Tim Dwyer, Jaroslaw Glowacki, Cagatay Goncu, Matthias Klapperstueck, Karsten Klein, Kim Marriott, Falk Schreiber, et al. 2015. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*. IEEE, 1–8. https://doi.org/10.1109/BDVA.2015.7314296

[18] Noshaba Cheema, Laura A. Frey-Law, Kourosh Naderi, Jaakko Lehtinen, Philipp Slusallek, and Perttu Hämäläinen. 2020. Predicting Mid-Air Interaction Movements and Fatigue Using Deep Reinforcement Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376701

[19] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D. Wilson. 2017. Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3718–3728. https://doi.org/10.1145/3025453.3025753

[20] Aldrich Clarence, Jarrod Knibbe, Maxime Cordeil, and Michael Wybrow. 2021. Unscripted Retargeting: Reach Prediction for Haptic Retargeting in Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 150–159. https://doi.org/10.1109/VR50410.2021.00036

[21] Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards Gaze-Based Prediction of the Intent to Interact in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 2, 7 pages. https://doi.org/10.1145/3448018.3458008

[22] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtner, and Kaj Grønbæk. 2021. *XRgonomics: Facilitating the Creation of Ergonomic 3D Interfaces*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445349

[23] Florian Fischer, Miroslav Bachinski, Markus Klar, Arthur Fleig, and Jörg Müller. 2021. Reinforcement learning control of a biomechanical model of the upper extremity. *Scientific Reports* 11, 1 (2021), 1–15. https://doi.org/10.1038/s41598-021-93760-1

[24] Erik Frøkjær, Morten Hertzum, and Kasper Hornbæk. 2000. Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 345–352. https://doi.org/10.1145/332040.332455

[25] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. 2019. Learning Cooperative Personalized Policies from Gaze Data. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 197–208. https://doi.org/10.1145/3332165.3347933

[26] Eric J. Gonzalez, Parastoo Abtahi, and Sean Follmer. 2020. REACH+: Extending the Reachability of Encountered-Type Haptics Devices through Dynamic Redirection in VR. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 236–248. https://doi.org/10.1145/3379337.3415870

[27] Joshua Goodman, Gina Venolia, Keith Steury, and Chauncey Parker. 2002. Language Modeling for Soft Keyboards. In *Proceedings of the 7th International Conference on Intelligent User Interfaces* (San Francisco, California, USA) *(IUI '02)*. Association for Computing Machinery, New York, NY, USA, 194–195. https://doi.org/10.1145/502716.502753

[28] Tovi Grossman and Ravin Balakrishnan. 2005. 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* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 281–290. https://doi.org/10.1145/1054972.1055012

[29] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. 2017. On calibration of modern neural networks. In *International Conference on Machine Learning*. PMLR, 1321–1330.

[30] Rorik Henrikson, Daniel Clarke, Thomas White, Frances Lai, Michael Glueck, Stephanie Santosa, Daniel Wigdor, Tovi Grossman, Sean Trowbridge, and Hrvoje Benko. 2020. Head-Coupled Kinematic Template Matching for Target Selection in Hangry Piggos. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3334480.3383176

[31] Rorik Henrikson, Tovi Grossman, Sean Trowbridge, Daniel Wigdor, and Hrvoje Benko. 2020. *Head-Coupled Kinematic Template Matching: A Prediction Model for Ray Pointing in VR*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376489

[32] Niels Henze, Markus Funk, and Alireza Sahami Shirazi. 2016. Software-Reduced Touchscreen Latency. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 434–441. https://doi.org/10.1145/2935334.2935381

[33] Lorenz Hetzel, John Dudley, Anna Maria Feit, and Per Ola Kristensson. 2021. Complex Interaction as Emergent Behaviour: Simulating Mid-Air Virtual Keyboard Typing using Reinforcement Learning. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (2021), 4140–4149. https://doi.org/10.1109/TVCG.2021.3106494

[34] Ashley Hill, Antonin Raffin, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, Rene Traore, Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, and Yuhuai Wu. 2018. Stable Baselines. https://github.com/hill-a/stable-baselines.

[35] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. 2021. Fixationnet: Forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2681–2690. https://doi.org/10.1109/TVCG.2021.3067779

[36] Chien-Ming Huang, Sean Andrist, Allison Sauppé, and Bilge Mutlu. 2015. Using gaze patterns to predict task intent in collaboration. *Frontiers in psychology* 6 (2015), 1049. https://doi.org/10.3389/fpsyg.2015.01049

[37] Chien-Ming Huang and Bilge Mutlu. 2016. Anticipatory robot control for efficient human-robot collaboration. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 83–90. https://doi.org/10.1109/HRI.2016.7451737

[38] Tanya R Jonker, Ruta Desai, Kevin Carlberg, James Hillis, Sean Keller, and Hrvoje Benko. 2020. The Role of AI in Mixed and Augmented Reality Interactions. In

CHI2020 ai4hci Workshop Proceedings. ACM.

[39] Fatemeh Koochaki and Laleh Najafizadeh. 2018. Predicting intention through eye gaze patterns. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 1–4. https://doi.org/10.1109/BIOCAS.2018.8584665

[40] Ben Lafreniere, Tanya R. Jonker, Stephanie Santosa, Mark Parent, Michael Glueck, Tovi Grossman, Hrvoje Benko, and Daniel Wigdor. 2021. False Positives vs. False Negatives: The Effects of Recovery Time and Cognitive Costs on Input Error Preference.. In *Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3472749.3474735

[41] Edward Lank, Yi-Chun Nikko Cheng, and Jaime Ruiz. 2007. Endpoint Prediction Using Motion Kinematics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 637–646. https://doi.org/10.1145/1240624.1240724

[42] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.

[43] Huy Viet Le, Valentin Schwind, Philipp Göttlich, and Niels Henze. 2017. PredicTouch: A System to Reduce Touchscreen Latency Using Neural Networks and Inertial Measurement Units. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (Brighton, United Kingdom) *(ISS '17)*. Association for Computing Machinery, New York, NY, USA, 230–239. https://doi.org/10.1145/3132272.3134138

[44] Michael McGuffin and Ravin Balakrishnan. 2002. Acquisition of Expanding Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 57–64. https://doi.org/10.1145/503376.503388

[45] Kaisa Miettinen. 2012. *Nonlinear multiobjective optimization*. Vol. 12. Springer Science & Business Media.

[46] Martez E. Mott and Jacob O. Wobbrock. 2014. Beating the Bubble: Using Kinematic Triggering in the Bubble Lens for Acquiring Small, Dense Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 733–742. https://doi.org/10.1145/2556288.2557410

[47] Atsuo Murata. 1998. Improvement of pointing time by predicting targets in pointing with a PC mouse. *International Journal of Human-Computer Interaction* 10, 1 (1998), 23–32. https://doi.org/10.1207/s15327590ijhc1001_2

[48] Patrick Ngatchou, Anahita Zarei, and A El-Sharkawi. 2005. Pareto multi objective optimization. In *Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems*. IEEE, 84–91. https://doi.org/10.1109/ISAP.2005.1599245

[49] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua V Dillon, Balaji Lakshminarayanan, and Jasper Snoek. 2019. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. *arXiv preprint arXiv:1906.02530* (2019).

[50] Kseniia Palin, Anna Maria Feit, Sunjun Kim, Per Ola Kristensson, and Antti Oulasvirta. 2019. How Do People Type on Mobile Devices? Observations from a Study with 37,000 Volunteers. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 9, 12 pages. https://doi.org/10.1145/3338286.3340120

[51] Phillip T. Pasqual and Jacob O. Wobbrock. 2014. Mouse Pointing Endpoint Prediction Using Kinematic Template Matching. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 743–752. https://doi.org/10.1145/2556288.2557406

[52] Philip Quinn and Andy Cockburn. 2008. The effects of menu parallelism on visual search and selection. In *Proceedings of the ninth conference on Australasian user interface-Volume 76*. 79–84.

[53] Philip Quinn and Andy Cockburn. 2020. Loss Aversion and Preferences in Interaction. *Human–Computer Interaction* 35, 2 (2020), 143–190. https://doi.org/10.1080/07370024.2018.1433040

[54] Philip Quinn and Shumin Zhai. 2016. A Cost-Benefit Study of Text Entry Suggestion Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 83–88. https://doi.org/10.1145/2858036.2858305

[55] Antonin Raffin. 2018. RL Baselines Zoo. https://github.com/araffin/rl-baselines-zoo.

[56] Antonin Raffin, Ashley Hill, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, and Noah Dormann. 2019. Stable Baselines3. https://github.com/DLR-RM/stable-baselines3.

[57] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 199–211. https://doi.org/10.1145/3311350.3347190

[58] Hosnieh Sattar, Mario Fritz, and Andreas Bulling. 2020. Deep gaze pooling: Inferring and visually decoding search intents from human gaze fixations. *Neurocomputing* 387 (2020), 369–382. https://doi.org/10.1016/j.neucom.2020.01.028

[59] Hosnieh Sattar, Sabine Muller, Mario Fritz, and Andreas Bulling. 2015. Prediction of search targets from fixations in open-world settings. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 981–990. https://doi.org/10.1109/CVPR.2015.7298700

[60] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2021. *Ninja Hands: Using Many Hands to Improve Target Selection in VR*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445759

[61] Ronal Singh, Tim Miller, Joshua Newn, Liz Sonenberg, Eduardo Velloso, and Frank Vetere. 2018. Combining Planning with Gaze for Online Human Intention Recognition. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (Stockholm, Sweden) *(AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 488–496.

[62] Kashyap Todi, Gilles Bailly, Luis Leiva, and Antti Oulasvirta. 2021. *Adapting User Interfaces with Model-Based Reinforcement Learning*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445497

[63] Christina Trepkowski, David Eibich, Jens Maiero, Alexander Marquardt, Ernst Kruijff, and Steven Feiner. 2019. The effect of narrow field of view and information density on visual search performance in augmented reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 575–584. https://doi.org/10.1109/VR.2019.8798312

[64] Lode Vanacken, Tovi Grossman, and Karin Coninx. 2007. 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. https://doi.org/10.1109/3DUI.2007.340783

[65] Datong Wei, Chaofan Yang, Xiaolong (Luke) Zhang, and Xiaoru Yuan. 2021. *Predicting Mouse Click Position Using Long Short-Term Memory Model Trained by Joint Loss Function*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3451651

[66] Ryen W. White, Peter Bailey, and Liwei Chen. 2009. Predicting User Interests from Contextual Information. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Boston, MA, USA) *(SIGIR '09)*. Association for Computing Machinery, New York, NY, USA, 363–370. https://doi.org/10.1145/1571941.1572005

[67] Ryen W. White, Paul N. Bennett, and Susan T. Dumais. 2010. Predicting Short-Term Interests Using Activity-Based Search Context. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management* (Toronto, ON, Canada) *(CIKM '10)*. Association for Computing Machinery, New York, NY, USA, 1009–1018. https://doi.org/10.1145/1871437.1871565

[68] Haijun Xia, Ricardo Jota, Benjamin McCanny, Zhe Yu, Clifton Forlines, Karan Singh, and Daniel Wigdor. 2014. Zero-Latency Tapping: Using Hover Information to Predict Touch Locations and Eliminate Touchdown Latency. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 205–214. https://doi.org/10.1145/2642918.2647348

[69] Difeng Yu, Hai-Ning Liang, Kaixuan Fan, Heng Zhang, Charles Fleming, and Konstantinos Papangelis. 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 (2019), 2762–2774. https://doi.org/10.1109/TVCG.2019.2905580

[70] Difeng Yu, Hai-Ning Liang, Xueshi Lu, Kaixuan Fan, and Barrett Ens. 2019. Modeling Endpoint Distribution of Pointing Selection Tasks in Virtual Reality Environments. *ACM Trans. Graph.* 38, 6, Article 218 (Nov. 2019), 13 pages. https://doi.org/10.1145/3355089.3356544

[71] Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Fully-occluded target selection in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3402–3413. https://doi.org/10.1109/TVCG.2020.3023606

[72] Gregory Zelinsky, Zhibo Yang, Lihan Huang, Yupei Chen, Seoyoung Ahn, Zijun Wei, Hossein Adeli, Dimitris Samaras, and Minh Hoai. 2019. Benchmarking gaze prediction for categorical visual search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 0–0. https://doi.org/10.1109/CVPRW.2019.00111

[73] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

[74] Brian Ziebart, Anind Dey, and J. Andrew Bagnell. 2012. Probabilistic Pointing Target Prediction via Inverse Optimal Control. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces* (Lisbon, Portugal) *(IUI '12)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2166966.2166968

A STUDY 1 - DATA COLLECTION

A.1 Task Scenarios

Two task scenarios, representative of common interaction tasks that are effortful to perform, were employed. The dense target selection task represented a manually-intensive task, where participants needed to select a small object located at the center of a cluster [46, 64]. The text matching task served as a mentally-demanding task, where participants needed to find and select an object with text that matched a target text. This task simulated real-world, search-heavy scenarios like searching for ingredients from a receipt, finding street names on a map, or browsing through a menu [52].

A.1.1 Dense Target Selection Task. This task was inspired by existing literature on small and dense target selection [46, 64]. The goal was to select the earth icon at the center of a planet cluster (Figure 3, left). The cluster was surrounded by other planet icons, which were randomly sized and distributed to add noise to the task environment. This setting required participants to aim precisely [64] and simulated scenarios where participants need to select objects in a cluttered virtual scene (e.g., select a keychain in a messy room).

The angular size of the target was set to 1°, which was determined by previous research to be sufficiently challenging [70]. The angular distance, or required movement amplitude, was fixed to 90°, and the target was generated in a predefined list of locations that were no more than 30° away from the horizontal plane. This target placement required participants to rotate their heads to find the out-of-view object, which added physical workload, without requiring that they overextend their neck. The distractors that were located directly adjacent to the target were the same size as the target, while others were randomly sized between 0.6° and 2°.

Participants started the task by pointing at a button at a fixed center position. A blue 3D arrow then appeared to indicate the location of the target. The arrow was designed to minimize search time in this task [69]. Participants then followed the direction of the arrow to point at the target through the right-hand controller and pressed the trigger to confirm their selection.

A.1.2 Text Matching Task. This task was designed to require participants to perform a difficult visual search (mentally-demanding) [25, 63]. Participants were required to find a target text string that matched a prompt (Figure 3, right) in a 6×7 grid of texts strings.

The angular distance between the candidates was 10° horizontally and 2.8° vertically to make sure all objects were located within field of view of participants to minimize their physical workload (e.g., turning their bodies to search for the target). The object radius was set to 1.5° and all objects were placed on a spherical plane.

Participants started the task by memorizing the target string and selecting a button at a fixed center position. All candidate strings then appeared with the goal text reminder at the top of the grid. To complete a task trial, participants pointed at the target icon using the controller and pressed the trigger to select it.

A.2 Suggestion Method

Two suggestion methods were used in the study—a highlighting suggestion and a pop-up suggestion. With the highlighting suggestion, a blinking fluorescent outline was displayed around the suggested object (Figure 4 left). A symbol of Button A also appeared at a pre-determined, unoccluded position close to the indicated object to depict that the object could be selected by pressing the Button A on the Touch controller. Participants could also cancel the suggestion by tilting the joystick to the right. Note that the highlighting suggestion was in-situ, so it remained at the object location without following the direction participants were looking.

With the pop-up notification suggestion, a suggestion window appeared at the bottom of the participant's current viewing direction (Figure 4 right) [57]. The suggestion presented either a predicted icon in the dense target selection task or a text string in the text matching task. When participants rotated their viewing direction, the pop-up notification followed the viewing direction using horizontal linear interpolation. Linear interpolation was not applied in the vertical dimension to avoid the suggestion being "stuck" on the head-mounted display, which may have caused visual discomfort. Like the highlighting suggestion, participants could quickly access the suggested object via the Button A or discard the suggestion by tilting the joystick to the right.

A.3 Example Data Trials

We show example data trials collected in session 2 in Figure 11.

A.4 Results - Session 2

Figure 12 shows the average response times and delayed times for the suggestion methods and task types. We performed significance tests with linear mixed models on response time and delayed time.

A.4.1 Response Time. Response time was defined the time elapsed between the appearance of a correct intelligent suggestion and a participant's selection of that suggestion. First, the Yeo-Johnson transformation, as chosen by the `bestNormalize` package in `R`, was applied to normalize the data. A linear mixed model was then used to identify whether different task types and suggestion methods lead to different response times across various suggestion timings. We set Task Type, Suggestion Method, and Suggestion Timing as fixed factors and Participant as a random factor. The linear mixed model indicated that there were interaction effects between Suggestion Method × Suggestion Timing ($F = 125.18, p < .001$) and Task Type × Suggestion Timing ($F = 49.47, p < .001$). As Task Type and Suggestion Method led to different response times across Suggestion Timing, we used multivariate adaptive regression splines (MARS) to model the relationships between suggestion timing and response time.

A.4.2 Response Rate. Response rate was defined as the likelihood that participants accepted a correct suggestion. Significance testing was not applied because the "rate" variable was only meaningful if we considered multiple data points.

A.4.3 Delayed Time. Delayed time was the time delay that was incurred due to incorrect suggestions. Similar to response time, an arcsinh transformation as suggested by the `bestNormalize` package, was applied and a linear mixed model was used to identify significant interaction effects between Task Type and Suggestion Method with regard to Suggestion Timing. The results indicated a significant effect of Task Type × Suggestion Timing ($F = 5.30, p = .021$), but not Suggestion Method × Suggestion

PID	Trial	FormalTrial	TaskType	SuggMethod	SuggTiming	SuggCorrectness	Distance	Angle	CompletionTime	Correctness	UseSugg	CancelSugg
0	0	FALSE	0	0	0.4754066	TRUE	90	84.00597	2.39489	TRUE	FALSE	FALSE
0	1	FALSE	0	0	0.5343446	FALSE	90	112.9455	2.979116	TRUE	FALSE	FALSE

Figure 11: Example data trials from session 2.

Figure 12: Average response times and delayed times for the suggestion methods (highlighting and pop-up notification) and task types (dense target selection and text matching). The error bars represent *mean* ± *std*.

Timing ($F = 1.24, p = 0.267$) nor Suggestion Method × Task Type × Suggestion Timing ($F = 0.01, p = .928$).

B STUDY 2 - SIMULATION

B.1 Target Prediction Model Mock-up

B.1.1 Target Prediction Model Observations. A selection prediction model based on the available data [20] was replicated and we observed how the predicted probability of the most likely object changed as the task progressed. Further, we drew inspiration from existing research on gaze-based target prediction [15, 36]. From these explorations, we made the following observations:

- The *global centerline* of model confidence over time (i.e., the average trend across all trials) seems to be a sigmoid-like curve [14, 15, 20, 74]. Intuitively, model confidence accelerates from a low point and becomes steady as it approaches an asymptote.
- By replicating [20] and observing results in [36], we found that while the *local centerline* of the model confidence value (i.e., the general trend of each trial) seems to roughly follow a sigmoid-like curve, it can deviate from the global centerline. While the local centerline can still be approximated by a sigmoid curve, the speed of increase can differ on each trial.
- The final confidence curve of each trial, rather than the general trend, contains seemingly randomly-distributed deviations (i.e., spikes and dips) from the local centerline. The evidence was found by replicating [20] and observing results in [36].

B.1.2 Mock-up Prediction Model Generation. Based on these observations, the following trial generation process was formulated for our mock-up prediction model. Our goal was to produce reasonable model confidence curves that mimic an actual prediction model.

- When starting to generate a data trial, the model first samples a trial length t_{max} based on the log-normal distribution regarding user task completion time found in Study 1 (Figure 6A). This sampling approach allows the final dataset to approximate the distribution of user task completion time.
- The model then generates a global centerline based on a sigmoid function $y_1 = sigmoid(x, k, x_0, u, l)$ where k is the logistic growth rate, x_0 is the sigmoid's midpoint, u is the upper bound, and l is the lower bound (Equation 7). This simulates the observation that the global centerline follows a sigmoid curve in an actual prediction model (Figure 6B).

$$y_1 = \frac{u - l}{1 + e^{-k(x - x_0)}} + l \tag{7}$$

- To simulate the variances in a local centerline, the model generates a Bell curve $y_2 = bell(x, \mu, \sigma)$ (Equation 8) to define the area of deviation (see Figure 6C). The distance between the local centerline y_3 and the global centerline is probabilistically sampled from a Gaussian distribution following Equation 9, where μ_r and σ_r are the predefined mean and standard deviation of a Gaussian distribution. By generating random numbers from a Gaussian distribution (with random.gauss), it is more likely that a local centerline is close to the global centerline than further away.

$$y_2 = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{8}$$

$$y_3 = y_1 + y_2 \cdot \text{random.gauss}(\mu_r, \sigma_r) \tag{9}$$

- The final step of the mock-up model is to generate spikes and dips based on the local centerline. To achieve this, the model uses a pre-determined probability j_p to represent the likelihood of jumping to another randomly generated local centerline (new y_3) at a particular timestamp t. The model goes through all timestamps in the trial and modifies the curve depending on it a jump will occur. The resulting curve preserves the property of previous steps: by averaging all generated trials, the centerline still follows a sigmoid function and the local centerline deviates within a predefined region. The model further corrects all probabilities larger than 1 to 1 and smaller than 0 to 0. A sample of a generated trial can be found in Figure 6D.

B.1.3 Dataset Generation. We pre-defined the parameters for the trial generation in later analyses. For the global centerline-related parameters, we set logistic growth rate $k = 2$, sigmoid's midpoint $x_0 = t_{max}/2$, upper bound $u = 1$, lower bound $l = 0$. This simulated a model that knew little information when users started a trial and increased its confidence over time until it reached an almost perfect understanding when users finished the trial, similar to the prediction models in [20] and [36]. Regarding the local centerline-related parameters, we set the bell curve mean $\mu = t_{max}/1.9$ and standard deviation $\sigma = 1$. We also set the Gaussian distribution mean $\mu_r = 0$ and standard deviation $\sigma_r = 1$. The random jump rate

j_p was fixed at 0.05. The final results yielded visually similar curves as in the literature [20, 36]. The frame rate was determined to be 50 (0.02 seconds per frame).

B.1.4 RL Reward Settings. Three reward settings were used to train the RL agents. The first reward setting was r_1, where $r_1^t = \texttt{Gain}(t)$ if a suggestion was displayed at t, otherwise $r_1^t = 0$. However, the sparsity in r_1 (i.e., the agent only receives a single reward per trial) prevented many of the agents from learning to display a suggestion at all.

The second reward setting, r_2, sought to solve the reward sparsity issue. Specifically, reward shaping was performed when the suggestion wasn't displayed: $r_2^t = \texttt{Gain}(t)$ if a suggestion was displayed at t, otherwise $r_2^t = -k \cdot p_m$. We used k to penalize the action of not displaying any suggestion. Furthermore, an agent received more of a penalty if it did not display a suggestion when the model confidence value was high (p_m). The penalty factor k was treated as a hyper-parameter during training. While r_2 worked well and enabled the agents to learn to display suggestions, a static value of k might have been limiting. In particular, the penalty of not displaying a suggestion should have changed as training progresses for true reward (i.e., gain function) maximization. In other words, the agent reliance on k should be reduced over the training process. Thus, k was decreased as the training progressed.

The third reward setting also leveraged the benefit of dense rewards, but removed the agents' reliance on the penalty factor k, which may have negative impacts on true reward maximization. In this setting, $r_3^t = \texttt{Gain}(t) - r_3^{t-1}$ (where $r_3^0 = 0$) at a timestamp t. This setting essentially rewarded the agent based on how good it performed on a particularly timestamp t, by computing the contribution of agent's action at t towards the gain. This reward setting thus allowed agents to learn directly from gain functions with dense feedback.

B.1.5 RL Training Methodology. OpenAI Gym [16] with Stable Baselines [34] (for recurrent policies) and Stable Baselines3 [56] (for MLP policies) were used to build and train the RL agents. A preliminary analysis was first run on the toy dataset to determine the appropriate model-free RL training algorithms (PPO2, DQN, A2C, and ACER), reward settings (r_1, r_2, and r_3), policy architectures (MLP and LSTM), policy network size, and training epochs for both task scenarios using the default hyper-parameter settings from the Stable Baselines. This experimentation demonstrated that the PPO2 training with MLP policies was a lightweight and effective solution. ACER with LSTM was the other powerful solution that worked well, but may take longer to train. r_3 was also found to be more suitable for the dense target selection task, while r_2 was better for the text matching task. The training with $4e6$ steps was sufficient for MLP policies and $2e6$ steps was adequate for LSTM policies, based on the convergence of gain in the validation dataset.

After the preliminary exploration, full-range hyper-parameter searches were performed with Optuna [5] using the training dataset for memory size m, penalty k, network size, activation function, learning rate, batch size, discount factor γ, and other algorithm-related parameters following the guidance of RL Baselines Zoo [55]. The model was then fine-tuned by focusing on several key parameters related to training. The training was stopped when the gain in the validation dataset converged. After training all the agents, their performance on the validation and testing dataset were benchmarked.

B.2 Validation and testing results

Detailed validation and testing results of Optimal Thresholding, Heuristic Thresholding, and RL can be found in Table 4 and Table 5.

B.3 Simulation 4: Revisiting a Prior Study

To determine the optimal timing of highlighting suggestions if we were to use an existing model for intelligent suggestion, we ran another simulation using an open-sourced dataset from a prior work [20]. The dataset contained 809 trials with four prediction features over time (i.e., position x, y, z, and rotation yaw every 10 milliseconds) and a final selected target. The original work was replicated with respect to data augmentation, LSTM structure, and training protocol, resulting in a model with 95.06% testing accuracy. For COBO, the features were refit to the trained model to obtain model confidence values over time for the 807 trials.

While it could be challenging to replicate the original study and acquire empirical data on participant response behavior towards intelligent suggestions, the following assumptions were made for the cost and benefit functions: (1) It would take participants 0.5 seconds (i.e., 0.25 seconds reaction time and 0.25 seconds trigger pressing time) to respond to a correct suggestion; (2) An incorrect suggestion would cause 0.25 seconds (i.e., reaction time) of delay; (3) participants would act rationally [62] and would not use a suggestion if the estimated response time (current time + 0.5 seconds) was larger than task completion time of that trial without any suggestion.

Under these assumptions, the optimized threshold for the two objectives were calculated using the COBO framework. The results show that the optimized threshold for completion time ($thres = 0.90$) was able to save 0.0801 seconds ($std. = 0.1540$ seconds) and the optimized threshold for the usage percentage ($thres = 0.82$) led to 52.27% ($std. = 42.20\%$) of clicks. Nine Pareto optimal values were also found($thres = 0.82 - 0.90$). The performance improvement in terms of time savings was small for this selection task, although a higher suggestion usage percentage could lead to better user experiences. The original authors' estimate based on the prediction accuracy alone ($thres = 0.85$) was close to our simulation results.

Table 4: Validation and testing results when using Optimal Thresholding and Heuristic Thresholding on the time saved for users and on suggestion usage percentages.

	Task Type	Strategy (Threshold)	Time Saved/Usage% (Std.) *Validation*	% Improved	Time Saved/Usage% *Test*	% Improved
Time saved	Dense Target Selection	Optimal Thresholding (0.47)	0.4073s (0.3169s)	44.07%	0.4073s (0.3202s)	39.39%
	Dense Target Selection	Heuristic Thresholding (0.85)	0.2827s (0.3597s)	-	0.2922s (0.3645s)	-
	Text Matching	Optimal Thresholding (0.98)	1.5822s (1.7991s)	268.38%	1.6211s (1.7946s)	260.89%
	Text Matching	Heuristic Thresholding (0.50)	0.4295s (1.1225s)	-	0.4492s (1.1440s)	-
Usage %	Dense Target Selection	Optimal Thresholding (0.81)	65.85% (17.70%)	0.64%	65.69% (18.30%)	0.36%
	Dense Target Selection	Heuristic Thresholding (0.85)	65.43% (20.24%)	-	65.45% (20.42%)	-
	Text Matching	Optimal Thresholding (0.96)	87.33% (18.44%)	50.72%	87.17% (18.53%)	51.52%
	Text Matching	Heuristic Thresholding (0.50)	57.94% (15.85%)	-	57.53% (15.63%)	-

Table 5: Validation and testing results of RL regarding time saved for users and suggestion usage percentages.

Task Type	Strategy	Time Saved (Std.) *Validation*	% Improved	Time Saved *Test*	% Improved	Usage% (Std.) *Validation*	% Improved	Usage% *Test*	% Improved
Pointing	PPO-MLP	0.4078s (0.3253s)	44.25%	0.4087s (0.3285s)	39.87%	-	-	-	-
Pointing	ACER-LSTM	0.4079s (0.3354s)	44.29%	0.4084s (0.3362s)	39.77%	-	-	-	-
Text Matching	PPO-MLP	1.5673s (1.7878s)	265.91%	1.6050s (1.7877s)	257.30%	87.33% (18.18%)	50.72%	87.31% (18.05%)	51.76%
Text Matching	ACER-LSTM	1.5275s (1.7418s)	240.05%	1.5671s (1.7328s)	248.86%	-	-	-	-

Chapter 8

DISCUSSION

In this chapter, we reflect on our solutions of occlusion visualizations, complementary modalities, and predictive models to address the research challenge of enhancing Virtual Hand and Raycasting. Based on our research findings, we illustrate how to apply the proposed techniques to improve selection and manipulation in complex VR interaction scenarios with different environmental settings and task requirements and discuss the caveats of using them. We also elaborate on the internal and external validity of our studies and point out potential limitations.

Going beyond the research work presented in this thesis, we envision what future VR selection and manipulation should look like. To move steadily towards more usable and useful 3D user interactions, we also present a framework for developing and determining appropriate solutions for different application scenarios. We outline future research directions regarding, for example, explainable theories and technique accessibility.

8.1 Selection and Manipulation for Complex VR Interaction

The most prevalent mid-air interaction techniques for object selection and manipulation (i.e., Virtual Hand and Raycasting) have limited capability in dealing with more complex application scenarios that contain small, distant, and occluded targets and require efficient, precise, versatile, and prolonged operations. In this thesis, we present new solutions that can enhance the selection and manipulation of complex VR interactions (**RQ**).

> **RQ**. How to enhance Virtual Hand and Raycasting for target selection and manipulation in complex VR interaction scenarios?

In Chapters 4, 5, 6, and 7, we presented occlusion visualization techniques, complementary modalities of gaze and on-body surface, and optimized predictive models to improve VR selection and manipulation. These solutions enable new interactions with small, distant, and occluded objects, and were demonstrated, through a set of user studies, to be effective, efficient, comfortable, and satisfying. The solutions were also shown to be usable and useful for various application scenarios. We have summarized the contributions of the solutions in Table 1 (we pasted the table in the following for a better reading experience). Next, we reflect upon the solutions based on the study results.

		Article I Occluded Selection	**Article II** Gaze Support	**Article III** On-Body Support	**Article IV** Intelligent Suggestion
Env.	Small	✓	✓	✓	✓
	Distant	✓	✓	✓	✓
	Occluded	✓		✓	
Task	Effectiveness	✓		✓	✓
	Efficiency	✓	✓	✓	✓
	Ergonomics		✓		✓
	Experience	✓	✓	✓	✓
	Expressivity	✓	✓	✓	✓

8.1.1 Occlusion Visualizations

In Chapter 4, we introduced various occlusion visualizations, including, for example, techniques that reorganize potential objects onto a grid for selection (i.e., *GridWall*, *LassoGrid+*, and *FlowerCone*), techniques that leverage a depth cursor to control the appearance of the objects (i.e., *AlphaCursor* and *GravityZone+*), and techniques that create a tiny replica of the virtual world (*MagicBall+*). The techniques differed regarding their visualization types, disambiguation mechanisms, and selection methods. The choices of techniques should depend on their intended usage in the application and their usefulness under different interaction scenarios.

Based on our studies, we found that all methods were effective in interacting with small, distant, and occluded targets. *LassoGrid+*, where users can select a group of candidates with a lasso and narrow down to the target with a 2D grid, was shown to be the most efficient and robust across multiple experimental conditions [25, 198]. These grid-based techniques were seen as the easiest and less intrusive. One note is that these techniques completely shifted the position of the objects, so they might not be suitable when maintaining the object location information is essential [40].

GravityZone+, where all objects in the environment could move towards the user by controls, and *MagicBall+*, where the remote selection was performed on a virtual world replica, could preserve the relative and exact location information. *GravityZone+* should be preferred if better efficiency is needed, and *MagicBall+* should be favored if the application focuses on providing a pleasant user experience (since *MagicBall+* was shown to have the highest hedonic quality with UEQ-S [150]). One consideration when applying these techniques is that they may be more sensitive to changes in environmental factors (e.g., depth of the target and density of the target area). For example, the efficiency of *GravityZone+* might decrease significantly if the target was located at the far end of the depth dimension. Our design recommendations presented in Chapter 4 are a useful guide for future designs of VR occluded target acquisition.

8.1.2 Complementary Modalities

In Chapter 5 and 6, we investigated the incorporation of gaze and on-body surfaces as complementary modalities to support the existing interaction workflow based on Virtual Hand and Raycasting. In Chapter 5, we presented techniques that snap a remote object onto a user's hand for manipulation and techniques that leverage collaborative movements of gaze and hand to translate and rotate a target. In general, we found that gaze could not offer significant benefits in terms of efficiency in manipulating objects in front of the user and within arm-reach distance but was helpful to handle distant objects in a larger environment [195]. When integrating gaze into real applications, it is essential to decide on the coordination and transition strategy between gaze and hands. For example, an explicit transition that requires a specific confirmation (e.g., a button click) to switch between the two modalities may be more robust but demand extra workload in performing the switching command. On the other hand, an implicit transition can enable smooth and concurrent transformations but will induce false triggering of the functions because of classification errors.

In Chapter 6, we proposed six design patterns that treat on-body and mid-air surfaces as input or output modalities for interaction. The design patterns inspired techniques that are capable of handling a variety of selection and manipulation tasks, such as occluded target selection, one degrees-of-freedom translation [108], and manipulations with adjustable CD ratio [48] (i.e., tuning the hand movement speed to be more rapid or more precise). With the high-level design concepts in our design patterns, on-body interfaces can assist mid-air interaction by providing quick access to different tools through subtle thumb-on-finger gestures [59] and achieving additional helpful, comfortable, and effective functions with finger-on-arm and on-body displays [163]. Our study results also uncovered where to provide on-body input and output to ensure better user experiences. For example, we found it beneficial to restrict the thumb-on-finger touching areas to the first and second segments of the index and middle fingers to satisfy general users while allowing customization to comply with individual needs.

One additional caveat when applying multiple modalities for input is that it increases the complexity of the interaction and may induce a higher cognitive load on the users [74, 102]. Evidence from our studies indicates that users may have limited cognitive bandwidth in performing simultaneous input in both modalities at the same time. For example, users were found to perform on-body input after they finished the mid-air input (i.e., sequentially), even though the interfaces allowed users to trigger the input altogether.

8.1.3 Predictive Models

In Chapter 7, we introduced a framework to optimize the support of target prediction models by offering timely, intelligent suggestions. The target prediction models actively and implicitly determine the likelihood of users interacting with an object over time. With the probability

distribution, our framework presents the most likely target through highlighting or notifications at the optimal timing. The framework was demonstrated to speed up the interaction and lower the interaction friction (i.e., reduce workload and enable seamless experience) [73]. Importantly, the framework can be extended for different optimization objectives and task scenarios as long as a quantification method can be determined.

However, there are still a few challenges in applying intelligent suggestion techniques in everyday scenarios. The most important challenge is that the target prediction models are not powerful enough to provide accurate predictions, given the complexities and noises involved in a daily interaction scenario. While we have seen target prediction models based on features of hand reaching [28] and gaze searching [69] to produce accurate results, they are only limited to a controlled experimental setting. With our optimization framework, we offer a proof-of-concept that these models can assist users in an interaction task, which can inspire predictive systems that are helpful to lower input friction and improve user experiences in current VR systems.

8.1.4 Generalizability of the Findings

We illustrate how we ensured internal and external validity through rigorous study and analysis protocols. We also discuss the limitations of our studies.

§1 Internal Validity. We employed carefully-designed study procedures, as described in Chapter 3 and detailed in the publications, to ensure the internal validity of our user studies. For example, we mitigated the ordering effects through randomized or counterbalanced designs. We also tried to minimize the variances in the data through multiple repetitive trials and recruiting sufficient participants [22]. Further, we conducted rigorous statistical tests to analyze our data. For instance, we did not fully rely on results from the significance tests to derive a conclusion but also considered other measures such as effect size.

§2 External Validity. In our research, we struck a balance between concreteness and abstraction when constructing an evaluation methodology to ensure the external validity of our work (i.e., the study results can be used for similar applications). First, when designing the controlled experiments, we closely approximated the features in concrete 3D interaction tasks. For example, in the intelligent suggestion study in Chapter 7, we used two types of tasks (dense target selection and text matching) that closely mimic the settings of intended applications which are manually intensive and mentally demanding [50, 167, 170].

Second, we assessed our techniques not only in the controlled experiments but also evaluated them in real-world settings to ensure ecological validity. For instance, in the gaze-supported manipulation study in Chapter 5, we proposed an interaction scenario that allowed users to reconstruct an empty virtual room as in similar applications such as Mozilla Hubs and

Minecraft VR. Users could move around the space and place their desired objects in an intended location. The techniques were embedded into users' own workflow and experiences.

Third, we validated our results in wide parameter ranges that may be contained in a real-world application. For example, in the fully-occluded target selection study in Chapter 4, we varied environmental factors that can influence user performance, including occurrence area, area density, occlusion layer, and target depth to provide an in-depth comparison of the techniques under different application scenarios.

§3 Limitations. While we attempted to ensure the validity of our results through rigorous studies and analyses, we acknowledge several limitations of our research. First, results from a limited number of studies may not provide a complete picture of the user experience, as they only captured a snapshot in time (i.e., single-point measurements). Although we attempted to vary different experiment settings, more replication or iterative studies are needed to further determine the significance of the study results and make meaningful conclusions. Second, while we tried to mimic the application scenarios, the studies were not conducted in the wild (i.e., in a real-world environment). Additionally, we did not test the use of the techniques longitudinally to explore the potential learning and adaptations in a longer-term scenario. To further ensure the validity of the results, we should perform meta-analyses with the increase of similar studies and conduct longitudinal, outside-of-lab studies in the future.

8.2 Advancing the Field with Multi-Objective Optimization

While numerous interaction techniques (including ours) have been developed, we now take a step back by envisioning the possible future of VR selection and manipulation techniques based on our current practices. We then present a framework for developing and determining appropriate selection and manipulation techniques for different application scenarios so that we can advance steadily towards our goals and apply the solutions more confidently.

8.2.1 Envisioning Future VR Selection and Manipulation

We believe that success measures are the main determinants of shaping future selection and manipulation methods—they are treated as optimization objectives of our endeavors. Based on our literature review and studies, our prediction is that the future solution should perform reasonably well in the *5Es* (effectiveness, efficiency, ergonomics, experience, and expressivity) and other success measurements (robustness, realism, behavior, and consistency). However, we should also note that the successful measurements may correlate or conflict with each other. For example, our analysis of the literature has shown that performance measures (effectiveness and efficiency) correlated with experience measures quite well; 76.9% of the proposed artifacts outperformed the baselines in experience measures when they achieved better performance. In other cases, researchers and designers might need to decide the tradeoff between the measures

like speed vs. accuracy tradeoff [133] and flexibility (expressivity) vs. efficiency tradeoff [92]. Eventually, the future selection and manipulation method will have to pick a subset of success measurements to optimize while trading off other objectives.

Another related inference about the future VR selection and manipulation methods is that there will be a clear separation between generalized solutions, which aim to handle numerous interaction scenarios, and specialized tools, which are dedicated to specific use cases [95, 191] (i.e., the breadth/depth dichotomy [65]). For example, the primary interaction metaphors based on Raycasting (or pointing in general) and Virtual Hand are unlikely to change significantly because of the significant commercialization and their flexibility to be used in various interaction scenarios of selecting and manipulating properly-sized, unoccluded menus, buttons, and objects. An implicit assistance from target prediction models can be applied to enhance their usability when appropriate [61, 194]. More explicit enhancements like visualizations and modalities [102, 154, 192], because of the added functionalities (and complexities), may continue proliferation to provide solutions for more specific scenarios such as for dense [9, 193], occluded [174, 198, 200], group-based [114, 162], and hands-free [94, 153] target selection and manipulation.

8.2.2 An Optimization Framework

As discussed in the previous section, VR object selection and manipulation solutions may only optimize a subset of the success measurements that are most useful for the intended application because trade-offs can exist between different objectives. In this case, it is essential to determine which techniques are the most ideal and how to develop the most appropriate solutions for a given application to guide future research and designs. To achieve that, we build a framework based on Pareto Frontier [118] for deciding the most suitable technique(s) given multiple success measurements. Pareto Frontier contains a set of solutions that cannot be better off in any targeted objective without making it worse off in another objective. The main idea of our proposed framework is to first determine the desired success measurements for a given application and then choose the Pareto optimal solutions (i.e., Pareto Frontier) for the application. We illustrate the detailed process through an example.

Suppose we are looking for the best techniques for mid-range (around 1-5 meters) target selection. We want to maximize its performance measures (i.e., efficiency and effectiveness). We pick two papers [9, 96] as our knowledge base for choosing the desired technique(s). In Lu et al.'s work [96], six techniques can be used for our purpose: *Go-Go*, *Raycasting* (i.e., *Naive Ray* in the paper), *Heuristic Ray*, *Quad Cone*, *BubbleRay-E*, and *BubbleRay-A*. From their comparison study, we can infer that for efficiency: *BubbleRay-A*>*BubbleRay-E*, *Heuristic Ray*, *Quad Cone*>*Go-Go*, *Raycasting*. For effectiveness: *BubbleRay-A*, *BubbleRay-E*, *Quad Cone*>*Go-Go*, *Heuristic Ray*, *Raycasting*. Similarly, we derive the following relationships based on the study

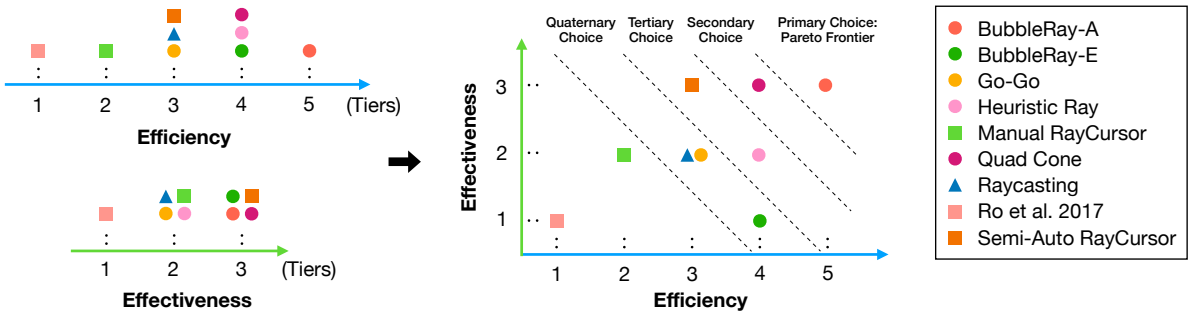


Fig. 9. To maximize performance measures (i.e., efficiency and effectiveness), we placed all the candidate techniques onto a 2D plot with efficiency as the x-axis and effectiveness as the y-axis. We can compute Pareto Frontier to determine the most optimal solution (i.e., *BubbleRay-A* in this case). The secondary, tertiary, and quaternary choices can also be concluded.

in Baloup et al. [9]. For efficiency: *Raycasting*, *Semi-Auto RayCursor*>*Manual RayCursor*>*Ro et al. 2017*. For effectiveness: *Semi-Auto RayCursor*>*Manual RayCursor*, *Raycasting*>*Ro et al. 2017*.

Based on the above relationships, we can place the techniques from the two papers onto an efficiency and effectiveness scale by using the common technique of *Raycasting* as the anchor (see Figure 9 left). In both scales, the higher the tiers, the better the techniques perform on that scale. Note that we should be cautious about combining results from different papers, as the experimental environments differed. To demonstrate the optimization framework, we assume these conclusions are generalizable.

Since we want to optimize two objectives, we can place all the candidate techniques onto a 2D plot with efficiency as the x-axis and effectiveness as the y-axis (see Figure 9 middle). Following the definition of Pareto Frontier, where we identify the solution(s) that either one of the dimensions could not be improved without worsening the other dimension, we conclude that *BubbleRay-A* should be our primary choice. A similar procedure can be followed to determine the secondary options (i.e., *Quad Cone*) by excluding the primary choices, also for tertiary and quaternary options.

The demonstrated framework can be easily extended for future use cases. For example, we can consider higher dimensions with more optimization objectives (e.g., four objectives including efficiency, effectiveness, experience, and robustness) if we have enough data support from previous empirical studies. Eyeballing the solutions might be difficult in higher dimensions, but the solutions can be computed programmatically. Furthermore, the optimization objectives do not need to be restricted to the nine success measurements categorized in this research, but with more detailed separations (e.g., fun vs. perceived ease-of-use).

The framework can not only help practitioners decide which techniques to choose given a set of optimization objectives but also guide the future development of VR selection and manipulation solutions. The framework indicates that we should aim to develop solutions locate at the Pareto Frontier of different combinations of objectives. It also suggests that we should conduct studies to verify the "tiers" of solutions, maybe with multiple studies to evaluate the generalizability of the conclusions in a given interaction scenario. In addition, future research should try to compare a newly proposed solution to common baselines (e.g., Virtual Hand or Raycasting) or the state-of-the-art to position it in the landscape of the techniques in the literature.

8.3 Future Research Directions

We have mentioned a few research directions in the thesis. In our literature review (Chapter 2), we identified small but emerging topics in the field, such as coping with the limitations in a user's physical space, integrating the selection and manipulation tasks into broad contexts and workflows, and enabling collaborative manipulation. Furthermore, through our research studies (Chapter 4, 5, 6, and 7), we identified that occlusion visualizations, multi-modality interaction, and predictive model integration were promising ways to enhance Virtual Hand and Raycasting and may require future research to mature the interactions. In this section, we emphasize several other research areas that we see as essential.

8.3.1 Proposal of Theories that can Explain

Recent discussions in HCI, in general, have been putting substantial attention on theory building (e.g., [67, 93, 122, 123]). A recent survey on selection and manipulation by Bergström et al. also highlighted the importance of evidence accumulation for theory building [13]. While we have seen a few papers on empirical models that can predict user selection behavior [194] or intended target of interest [61], these models are mainly descriptive—they do not provide a sense of understanding about the causes of the predicted event. Since being able to explain the causal mechanism is indispensable for a scientific theory [141], we should aim to build theories that can provide understanding regarding the underlying cognitive and motor mechanisms of VR object selection and manipulation.

8.3.2 User Behavior in Moving Target Acquisition

Selecting moving targets is commonplace in VR games and social platforms. In popular VR games such as Beat Saber, Fruit Ninja, and Robot Recall, players must often aim, catch, or grab flying targets (like fruit or bullets). One interesting direction is to model user behavior in such moving target acquisition tasks. With a generalizable user model for moving target acquisition, a game designer can better predict how users will behave when adjusting the parameters to unseen conditions without costly user tests (i.e., enabling automated playtesting [196]). A user may also better handle challenging game scenes or complex interaction scenarios that contain moving target selection (e.g., selecting a datapoint in VR traffic flow visualizations [26, 181]).

Additionally, understanding user general selection behavior regardless of static or moving targets may provoke relevant theories on target selection.

8.3.3 Study Generalizability and User Simulation

Good generalizability allows research findings to remain valid across relevant application scenarios. One issue that we have constantly been reflecting upon during the development of this thesis is to improve the study methodology to ensure the generalizability of the study findings to be reusable for other applications. The challenge here is that user studies are normally costly, and only a limited number and levels of variables can be included in one study. User fatigue and disengagement may also affect the validity of user data. One interesting future direction is to simulate users with an AI agent. For example, Ikkala et al. [70] applied biomechanical and perception models with reinforcement learning to reproduce user behaviors in simple tasks like pointing and object tracking. We deem this a promising future research direction as various experimental conditions and application scenarios may be easily simulated in computers to evaluate a newly proposed interaction technique.

8.3.4 Measuring Accessibility

In addition to the success measurements discussed in this thesis, it would be helpful to add the measure of accessibility to expand access to VR selection and manipulation techniques. There is a significant number of people who suffer from disabilities worldwide [116, 120]. Furthermore, every user could experience situational impairments depending on their situations, contexts, and environments (e.g., a user may not be able to use their arms for interaction while lifting heavy goods) [146, 151, 186]. It is thus essential to consider how to adjust the proposed methods to support people with (situational) disabilities such as visual [201] or motor [117] impairments to accomplish selection and manipulation tasks in VR.

Chapter 9

CONCLUSION

Object selection and manipulation are the foundation of VR interactions—users perform selections to identify target(s) of interest and execute manipulations to transform the target into a desired configuration (i.e., location, rotation, and scale). Existing VR systems primarily rely on Virtual Hand and Raycasting which are imprecise, inefficient, and cumbersome, especially in complex scenarios that contain small, distant, and occluded targets.

In this thesis, we have presented a set of solutions to enhance Virtual Hand and Raycasting for object selection and manipulation in complex VR interaction scenarios. Occlusion visualization techniques help reveal fully-occluded targets and empower efficient disambiguation for target acquisition. Complementary modalities, including gaze and on-body surfaces, can enable access to various helpful functionalities and shortcuts that are indispensable for complex VR interactions. Predictive models, which actively infer a user's intention in the background, can provide prompt intelligent suggestions to improve user performance and experience.

The results from a series of user studies have demonstrated that our solutions can help handle small, distant, and occluded targets and are effective, efficient, comfortable, and satisfying in different application scenarios. Moreover, our design space, framework, and design recommendations can guide researchers and designers to effortlessly adapt the solutions to a multitude of VR interaction experiences.

In this thesis, we have also anticipated future VR selection and manipulation techniques and proposed a framework for developing and determining appropriate selection and manipulation techniques for different applications with multiple design objectives. Additionally, we have outlined future directions regarding explainable theories, moving target acquisition, user simulation, and technique accessibility. Finally, we envision the technical solutions and findings presented in this thesis inspire more usable and useful 3D user interfaces in VR systems.

References

[1] Parastoo Abtahi, Sidney Q. Hough, James A. Landay, and Sean Follmer. 2022. Beyond Being Real: A Sensorimotor Control Perspective on Interactions in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 358, 17 pages. https://doi.org/10.1145/3491102.3517706

[2] Anne Adams, Peter Lunt, and Paul Cairns. 2008. A qualititative approach to HCI research. (2008).

[3] Remi Alkemade, Fons J Verbeek, and Stephan G Lukosch. 2017. On the efficiency of a VR hand gesture-based interface for 3D object manipulations in conceptual design. *International Journal of Human–Computer Interaction* 33, 11 (2017), 882–901. https://doi.org/10.1080/10447318.2017.1296074

[4] Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. 2015. Interaction with Large Displays: A Survey. *ACM Comput. Surv.* 47, 3, Article 46 (feb 2015), 38 pages. https://doi.org/10.1145/2682623

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

[6] Oscar Ariza, Gerd Bruder, Nicholas Katzakis, and Frank Steinicke. 2018. 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)*. IEEE, 327–334. https://doi.org/10.1109/VR.2018.8446317

[7] Jatin Arora, Aryan Saini, Nirmita Mehra, Varnit Jain, Shwetank Shrey, and Aman Parnami. 2019. VirtualBricks: Exploring a Scalable, Modular Toolkit for Enabling Physical Manipulation in VR. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300286

[8] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1968–1979. https://doi.org/10.1145/2858036.2858226

[9] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300331

[10] Anil Ufuk Batmaz, Mayra Donaji Barrera Machuca, Junwei Sun, and Wolfgang Stuerzlinger. 2022. The Effect of the Vergence-Accommodation Conflict on Virtual Hand Pointing in Immersive Displays. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 633, 15 pages. https://doi.org/10.1145/3491102.3502067

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

[12] Anil Ufuk Batmaz and Wolfgang Stuerzlinger. 2019. Effects of 3D rotational jitter and selection methods on 3D pointing tasks. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1687–1692. https://doi.org/10.1109/VR.2019.8798038

[13] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. 2021. How to evaluate object selection and manipulation in VR? Guidelines from 20 years of studies. In *Proceedings of the 2021*

CHI Conference on Human Factors in Computing Systems. 1–20. https://doi.org/10.1145/3411764.3445193

[14] Mark Billinghurst, Maxime Cordeil, Anastasia Bezerianos, and Todd Margolis. 2018. Collaborative immersive analytics. In *Immersive Analytics*. Springer, 221–257. https://doi.org/10.1007/978-3-030-01388-2_8

[15] Gunnar AV Borg. 1982. Psychophysical bases of perceived exertion. *Medicine & science in sports & exercise* (1982). https://doi.org/10.1249/00005768-198205000-00012

[16] Sebastian Boring, Marko Jurmu, and Andreas Butz. 2009. Scroll, Tilt or Move It: Using Mobile Phones to Continuously Control Pointers on Large Public Displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7* (Melbourne, Australia) *(OZCHI '09)*. Association for Computing Machinery, New York, NY, USA, 161–168. https://doi.org/10.1145/1738826.1738853

[17] Doug A Bowman, Brian Badillo, and Dhruv Manek. 2007. Evaluating the need for display-specific and device-specific 3D interaction techniques. In *International Conference on Virtual Reality*. Springer, 195–204. https://doi.org/10.1007/978-3-540-73335-5_22

[18] Doug A. Bowman and Larry F. Hodges. 1997. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 35–ff. https://doi.org/10.1145/253284.253301

[19] Doug A Bowman and Larry F Hodges. 1999. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing* 10, 1 (1999), 37–53. https://doi.org/10.1006/jvlc.1998.0111

[20] Doug A Bowman, Donald B Johnson, and Larry F Hodges. 1999. Testbed evaluation of virtual environment interaction techniques. In *Proceedings of the ACM symposium on Virtual reality software and technology*. 26–33. https://doi.org/10.1145/323663.323667

[21] Doug A Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan Poupyrev. 2005. *3D user interfaces: theory and practice*. Addison-Wesley.

[22] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 981–992. https://doi.org/10.1145/2858036.2858498

[23] Fabio M Caputo, Marco Emporio, and Andrea Giachetti. 2018. The Smart Pin: An effective tool for object manipulation in immersive virtual reality environments. *Computers & Graphics* 74 (2018), 225–233. https://doi.org/10.1016/j.cag.2018.05.019

[24] Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. 1990. The Design Space of Input Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 117–124. https://doi.org/10.1145/97243.97263

[25] Jeffrey Cashion, Chadwick Wingrave, and Joseph J LaViola Jr. 2012. Dense and dynamic 3d selection for game-based virtual environments. *IEEE transactions on visualization and computer graphics* 18, 4 (2012), 634–642. https://doi.org/10.1109/TVCG.2012.40

[26] Qianwen Chao, Huikun Bi, Weizi Li, Tianlu Mao, Zhaoqi Wang, Ming C Lin, and Zhigang Deng. 2020. A survey on visual traffic simulation: Models, evaluations, and applications in autonomous driving. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 287–308. https://doi.org/10.1111/cgf.13803

[27] Di Laura Chen, Ravin Balakrishnan, and Tovi Grossman. 2020. Disambiguation techniques for freehand object manipulations in virtual reality. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 285–292. https://doi.org/10.1109/VR46266.2020.00048

[28] Aldrich Clarence, Jarrod Knibbe, Maxime Cordeil, and Michael Wybrow. 2021. Unscripted retargeting: Reach prediction for haptic retargeting in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 150–159. https://doi.org/10.1109/VR50410.2021.00036

[29] Nathan Cournia, John D. Smith, and Andrew T. Duchowski. 2003. Gaze- vs. Hand-Based Pointing in Virtual Environments. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI EA '03)*. Association for Computing Machinery, New York, NY, USA, 772–773. https://doi.org/10.1145/765891.765982

[30] Tor-Salve Dalsgaard, Jarrod Knibbe, and Joanna Bergström. 2021. Modeling Pointing for 3D Target Selection in VR. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 42, 10 pages. https://doi.org/10.1145/3489849.3489853

[31] Nguyen-Thong Dang. 2007. A survey and classification of 3D pointing techniques. In *2007 IEEE international conference on research, innovation and vision for the future*. IEEE, 71–80. https://doi.org/10.1109/RIVF.2007.369138

[32] Henrique G Debarba, Jad-Nicolas Khoury, Sami Perrin, Bruno Herbelin, and Ronan Boulic. 2018. Perception of redirected pointing precision in immersive virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 341–346. https://doi.org/10.1109/VR.2018.8448285

[33] Henrique G. Debarba, Sami Perrin, Bruno Herbelin, and Ronan Boulic. 2015. Embodied Interaction Using Non-Planar Projections in Immersive Virtual Reality. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology* (Beijing, China) *(VRST '15)*. Association for Computing Machinery, New York, NY, USA, 125–128. https://doi.org/10.1145/2821592.2821603

[34] Thibauld Delrieu, Vincent Weistroffer, and Jean Pierre Gazeau. 2020. Precise and realistic grasping and manipulation in Virtual Reality without force feedback. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 266–274. https://doi.org/10.1109/VR46266.2020.00046

[35] Diane Dewez, Ludovic Hoyet, Anatole Lecuyer, and Ferran Argelaguet. 2022. Do You Need Another Hand? Investigating Dual Body Representations During Anisomorphic 3D Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2047–2057. https://doi.org/10.1109/TVCG.2022.3150501

[36] Diane Dewez, Ludovic Hoyet, Anatole Lécuyer, and Ferran Argelaguet Sanz. 2021. Towards "Avatar-Friendly" 3D Manipulation Techniques: Bridging the Gap Between Sense of Embodiment and Interaction in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 264, 14 pages. https://doi.org/10.1145/3411764.3445379

[37] Ian Dey. 2003. *Qualitative data analysis: A user friendly guide for social scientists*. Routledge.

[38] Ciro Donalek, S George Djorgovski, Alex Cioc, Anwell Wang, Jerry Zhang, Elizabeth Lawler, Stacy Yeh, Ashish Mahabal, Matthew Graham, Andrew Drake, et al. 2014. Immersive and collaborative data visualization using virtual reality platforms. In *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 609–614. https://doi.org/10.1109/BigData.2014.7004282

[39] Gregory W Edwards, Woodrow Barfield, and Maury A Nussbaum. 2004. The use of force feedback and auditory cues for performance of an assembly task in an immersive virtual environment. *Virtual reality* 7, 2 (2004), 112–119. https://doi.org/10.1007/s10055-004-0120-6

[40] Niklas Elmqvist and Philippas Tsigas. 2007. A taxonomy of 3D occlusion management techniques. In *2007 IEEE Virtual Reality Conference*. IEEE, 51–58. https://doi.org/10.1109/VR.2007.352463

[41] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (2019), 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011

[42] Cathy Mengying Fang and Chris Harrison. 2021. Retargeted Self-Haptics for Increased Immersion in VR without Instrumentation. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 1109–1121. https://doi.org/10.1145/3472749.3474810

[43] Martin Feick, Scott Bateman, Anthony Tang, André Miede, and Nicolai Marquardt. 2020. Tangi: Tangible proxies for embodied object exploration and manipulation in virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 195–206. https://doi.org/10.1109/ISMAR50242.2020.00042

[44] Alex Olwal Steven Feiner. 2003. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proc. UIST*, Vol. 3. 81–82.

[45] James D Foley, Foley Dan Van, Andries Van Dam, Steven K Feiner, and John F Hughes. 1996. *Computer graphics: principles and practice*. Vol. 12110. Addison-Wesley Professional.

[46] Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. 1996. Aperture Based Selection for Immersive Virtual Environments. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 95–96. https://doi.org/10.1145/237091.237105

[47] Scott Frees and G Drew Kessler. 2005. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* IEEE, 99–106. https://doi.org/10.1109/VR.2005.1492759

[48] Scott Frees, G. Drew Kessler, and Edwin Kay. 2007. PRISM Interaction for Enhancing Control in Immersive Virtual Environments. *ACM Trans. Comput.-Hum. Interact.* 14, 1 (may 2007), 2–es. https://doi.org/10.1145/1229855.1229857

[49] Zihan Gao, Huiqiang Wang, Hongwu Lv, Moshu Wang, and Yifan Qi. 2020. Evaluating the Effects of Non-isomorphic Rotation on 3D Manipulation Tasks in Mixed Reality Simulation. *IEEE Transactions on Visualization and Computer Graphics* 28, 2 (2020), 1261–1273. https://doi.org/10.1109/TVCG.2020.3010247

[50] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. 2019. Learning Cooperative Personalized Policies from Gaze Data. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 197–208. https://doi.org/10.1145/3332165.3347933

[51] P Christopher Gloumeau, Wolfgang Stuerzlinger, and JungHyun Han. 2020. Pinnpivot: Object manipulation using pins in immersive virtual environments. *IEEE transactions on visualization and computer graphics* 27, 4 (2020), 2488–2494. https://doi.org/10.1109/TVCG.2020.2987834

[52] Mar Gonzalez-Franco and Tabitha C Peck. 2018. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74. https://doi.org/10.3389/frobt.2018.00074

[53] Jerônimo Gustavo Grandi, Henrique Galvan Debarba, and Anderson Maciel. 2019. Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 127–135. https://doi.org/10.1109/VR.2019.8798080

[54] Jerônimo Gustavo Grandi, Henrique Galvan Debarba, Luciana Nedel, and Anderson Maciel. 2017. Design and Evaluation of a Handheld-Based 3D User Interface for Collaborative Object Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5881–5891. https://doi.org/10.1145/3025453.3025935

[55] Jan Gugenheimer, David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2016. FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality. In *Proceedings of the 29th Annual*

Symposium on User Interface Software and Technology (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 49–60. https://doi.org/10.1145/2984511.2984576

[56] Kim Halskov and Caroline Lundqvist. 2021. Filtering and Informing the Design Space: Towards Design-Space Thinking. *ACM Trans. Comput.-Hum. Interact.* 28, 1, Article 8 (jan 2021), 28 pages. https://doi.org/10.1145/3434462

[57] Chris Hand. 1997. A survey of 3D interaction techniques. In *Computer graphics forum*, Vol. 16. Wiley Online Library, 269–281. https://doi.org/10.1111/1467-8659.00194

[58] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908. https://doi.org/10.1177/15419312060500090

[59] Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. Plane, Ray, and Point: Enabling Precise Spatial Manipulations with Shape Constraints. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1185–1195. https://doi.org/10.1145/3332165.3347916

[60] Chris Heape. 2007. The Design Space: the design process as the construction, exploration and expansion of a conceptual space. (2007).

[61] Rorik Henrikson, Tovi Grossman, Sean Trowbridge, Daniel Wigdor, and Hrvoje Benko. 2020. Head-Coupled Kinematic Template Matching: A Prediction Model for Ray Pointing in VR. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376489

[62] Kenneth P. Herndon, Andries van Dam, and Michael Gleicher. 1994. The Challenges of 3D Interaction: A CHI '94 Workshop. *SIGCHI Bull.* 26, 4 (oct 1994), 36–43. https://doi.org/10.1145/191642.191652

[63] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[64] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A Survey of Design Issues in Spatial Input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology* (Marina del Rey, California, USA) *(UIST '94)*. Association for Computing Machinery, New York, NY, USA, 213–222. https://doi.org/10.1145/192426.192501

[65] Ken Hinckley and Daniel Wigdor. 2002. Input technologies and techniques. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications* (2002), 151–168.

[66] Kasper Hornbæk. 2006. Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies* 64, 2 (2006), 79–102. https://doi.org/10.1016/j.ijhcs.2005.06.002

[67] Kasper Hornbæk. 2022. *Implications for Theory — Keynote at NordiCHI 2022.* Retrieved January 17, 2023 from https://www.kasperhornbaek.dk/presentations/Hornb%C3%A6k-ImplicationsForTheory-Nordichi2022Keynote.pdf

[68] Wen-jun Hou and Xiao-lin Chen. 2021. Comparison of Eye-Based and Controller-Based Selection in Virtual Reality. *International Journal of Human–Computer Interaction* 37, 5 (2021), 484–495. https://doi.org/10.1080/10447318.2020.1826190

[69] Chien-Ming Huang, Sean Andrist, Allison Sauppé, and Bilge Mutlu. 2015. Using gaze patterns to predict task intent in collaboration. *Frontiers in psychology* 6 (2015), 1049. https://doi.org/10.3389/fpsyg.2015.01049

[70] Aleksi Ikkala, Florian Fischer, Markus Klar, Miroslav Bachinski, Arthur Fleig, Andrew Howes, Perttu Hämäläinen, Jörg Müller, Roderick Murray-Smith, and Antti Oulasvirta. 2022. Breathing Life Into Biomechanical User Models. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 90, 14 pages. https://doi.org/10.1145/3526113.3545689

[71] ISO. 2018. 9241-11:2018(en). Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. *The international organization for standardization* (2018).

[72] Bret Jackson, Brighten Jelke, and Gabriel Brown. 2018. Yea big, yea high: A 3D user interface for surface selection by progressive refinement in virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 320–326. https://doi.org/10.1109/VR.2018.8447559

[73] Tanya R Jonker, Ruta Desai, Kevin Carlberg, James Hillis, Sean Keller, and Hrvoje Benko. 2020. The Role of AI in Mixed and Augmented Reality Interactions. In *CHI2020 ai4hci Workshop Proceedings. ACM.*

[74] Sungchul Jung, Andrew L Wood, Simon Hoermann, Pramuditha L Abhayawardhana, and Robert W Lindeman. 2020. The impact of multi-sensory stimuli on confidence levels for perceptual-cognitive tasks in vr. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 463–472. https://doi.org/10.1109/VR46266.2020.00067

[75] Saleh Kalantari and Jun Rong Jeffrey Neo. 2020. Virtual environments for design research: Lessons learned from use of fully immersive virtual reality in interior design research. *Journal of Interior Design* 45, 3 (2020), 27–42. https://doi.org/10.1111/joid.12171

[76] Nikolaos Katzakis, Lihan Chen, Oscar Ariza, Robert J Teather, and Frank Steinicke. 2019. Evaluation of 3d pointing accuracy in the fovea and periphery in immersive head-mounted display environments. *IEEE transactions on visualization and computer graphics* 27, 3 (2019), 1929–1936. https://doi.org/10.1109/TVCG.2019.2947504

[77] Bohyun Kim. 2019. Virtual Reality for 3D Modeling. *Beyond Reality: Augmented, Virtual, and Mixed Reality in the Library* (2019), 31–46.

[78] MyoungGon Kim and JungHyun Han. 2019. Effects of switchable dof for mid-air manipulation in immersive virtual environments. *International Journal of Human–Computer Interaction* 35, 13 (2019), 1147–1159. https://doi.org/10.1080/10447318.2018.1514163

[79] Panayiotis Koutsabasis and Panagiotis Vogiatzidakis. 2019. Empirical research in mid-air interaction: A systematic review. *International Journal of Human–Computer Interaction* 35, 18 (2019), 1747–1768.

[80] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. 2017. Augmented reality versus virtual reality for 3d object manipulation. *IEEE transactions on visualization and computer graphics* 24, 2 (2017), 1038–1048. https://doi.org/10.1109/TVCG.2017.2658570

[81] Stanislav Kyian and Robert Teather. 2021. Selection Performance Using a Smartphone in VR with Redirected Input. In *Symposium on Spatial User Interaction* (Virtual Event, USA) *(SUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 6, 12 pages. https://doi.org/10.1145/3485279.3485292

[82] Markus Lange, Frank Steinicke, Gerd Bruder, et al. 2017. Vibrotactile assistance for user guidance towards selection targets in VR and the cognitive resources involved. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 95–98. https://doi.org/10.1109/3DUI.2017.7893323

[83] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice.* Addison-Wesley Professional.

[84] Morgan Le Chénéchal, Jérémy Lacoche, Jérôme Royan, Thierry Duval, Valérie Gouranton, and Bruno Arnaldi. 2016. When the giant meets the ant an asymmetric approach for collaborative and concurrent object manipulation in a multi-scale environment. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*. IEEE, 18–22. https://doi.org/10.1109/3DCVE.2016.7563562

[85] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation Strategies for HCI Toolkit Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3173574.3173610

[86] Chia-Yang Lee, Wei-An Hsieh, David Brickler, Sabarish V. Babu, and Jung-Hong Chuang. 2021. Design and Empirical Evaluation of a Novel Near-Field Interaction Metaphor on Distant Object Manipulation in VR. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction* (Virtual Event, USA) *(SUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 13, 11 pages. https://doi.org/10.1145/3485279.3485296

[87] Jaeyeon Lee, Mike Sinclair, Mar Gonzalez-Franco, Eyal Ofek, and Christian Holz. 2019. TORC: A Virtual Reality Controller for In-Hand High-Dexterity Finger Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300301

[88] Jan Leusmann. 2021. A literature review on distant object selection methods. (2021).

[89] Jialei Li, Isaac Cho, and Zachary Wartell. 2018. Evaluation of Cursor Offset on 3D Selection in VR. In *Proceedings of the Symposium on Spatial User Interaction* (Berlin, Germany) *(SUI '18)*. Association for Computing Machinery, New York, NY, USA, 120–129. https://doi.org/10.1145/3267782.3267797

[90] Nianlong Li, Teng Han, Feng Tian, Jin Huang, Minghui Sun, Pourang Irani, and Jason Alexander. 2020. Get a Grip: Evaluating Grip Gestures for VR Input Using a Lightweight Pen. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376698

[91] Zhen Li, Joannes Chan, Joshua Walton, Hrvoje Benko, Daniel Wigdor, and Michael Glueck. 2021. Armstrong: An Empirical Examination of Pointing at Non-Dominant Arm-Anchored UIs in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 123, 14 pages. https://doi.org/10.1145/3411764.3445064

[92] William Lidwell, Kritina Holden, and Jill Butler. 2010. *Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* Rockport Pub.

[93] Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. CHI 1994-2013: Mapping Two Decades of Intellectual Progress through Co-Word Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3553–3562. https://doi.org/10.1145/2556288.2556969

[94] Xueshi Lu, Difeng Yu, Hai-Ning Liang, and Jorge Goncalves. 2021. IText: Hands-Free Text Entry on an Imaginary Keyboard for Augmented Reality Systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 815–825. https://doi.org/10.1145/3472749.3474788

[95] Yujun Lu, BoYu Gao, Huawei Tu, Huiyue Wu, Weiqiang Xin, Hui Cui, Weiqi Luo, and Henry Been-Lirn Duh. 2022. Effects of physical walking on eyes-engaged target selection with ray-casting pointing in virtual reality. *Virtual Reality* (2022), 1–23. https://doi.org/10.1007/s10055-022-00677-9

[96] Yiqin Lu, Chun Yu, and Yuanchun Shi. 2020. Investigating bubble mechanism for ray-casting to improve 3D target acquisition in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 35–43. https://doi.org/10.1109/VR46266.2020.00021

[97] Jaime Maldonado and Christoph Zetzsche. 2021. Object Manipulations in VR Show Task- and Object-Dependent Modulation of Motor Patterns. In *Proceedings of the 27th ACM Symposium on Virtual Reality*

Software and Technology (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 41, 9 pages. https://doi.org/10.1145/3489849.3489858

[98] Dhruv B Manek. 2004. *Effects of Visual Displays on 3D Iteraction in Virtual Environments*. Ph.D. Dissertation. Virginia Tech.

[99] Gary Marchionini and John Sibert. 1991. An Agenda for Human-Computer Interaction: Science and Engineering Serving Human Needs. *SIGCHI Bull.* 23, 4 (oct 1991), 17–32. https://doi.org/10.1145/126729.126741

[100] Diako Mardanbegi, Benedikt Mayer, Ken Pfeuffer, Shahram Jalaliniya, Hans Gellersen, and Alexander Perzl. 2019. Eyeseethrough: Unifying tool selection and application in virtual environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 474–483. https://doi.org/10.1109/VR.2019.8797988

[101] Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, and Bruce H Thomas. 2018. *Immersive analytics*. Vol. 11190. Springer.

[102] Daniel Martin, Sandra Malpica, Diego Gutierrez, Belen Masia, and Ana Serrano. 2022. Multimodality in VR: A Survey. *ACM Comput. Surv.* 54, 10s, Article 216 (sep 2022), 36 pages. https://doi.org/10.1145/3508361

[103] Alejandro Martin-Gomez, Ulrich Eck, and Nassir Navab. 2019. Visualization techniques for precise alignment in VR: A comparative study. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, 735–741. https://doi.org/10.1109/VR.2019.8798135

[104] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174227

[105] Daniel Mendes, Fabio Marco Caputo, Andrea Giachetti, Alfredo Ferreira, and Joaquim Jorge. 2019. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, Vol. 38. Wiley Online Library, 21–45. https://doi.org/10.1111/cgf.13390

[106] Daniel Mendes, Daniel Medeiros, Maurício Sousa, Eduardo Cordeiro, Alfredo Ferreira, and Joaquim A Jorge. 2017. Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement. *Computers & Graphics* 67 (2017), 95–102. https://doi.org/10.1016/j.cag.2017.06.003

[107] Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. 2016. The Benefits of DOF Separation in Mid-Air 3D Object Manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/2993369.2993396

[108] Daniel Mendes, Maurício Sousa, Rodrigo Lorena, Alfredo Ferreira, and Joaquim Jorge. 2017. Using Custom Transformation Axes for Mid-Air Manipulation of 3D Virtual Objects. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) *(VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 27, 8 pages. https://doi.org/10.1145/3139131.3139157

[109] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[110] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, Vol. 2351. Spie, 282–292. https://doi.org/10.1117/12.197321

[111] Mark R. Mine, Frederick P. Brooks, and Carlo H. Sequin. 1997. Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., USA, 19–26. https://doi.org/10.1145/258734.258747

[112] Mathias Moehring and Bernd Froehlich. 2011. Effective manipulation of virtual objects within arm's reach. In *2011 IEEE Virtual Reality Conference*. IEEE, 131–138. https://doi.org/10.1109/VR.2011.5759451

[113] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4 (2009), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

[114] Roberto A Montano-Murillo, Cuong Nguyen, Rubaiat Habib Kazi, Sriram Subramanian, Stephen DiVerdi, and Diego Martinez-Plasencia. 2020. Slicing-volume: Hybrid 3d/2d multi-target selection technique for dense virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 53–62. https://doi.org/10.1109/VR46266.2020.00023

[115] Roberto A. Montano Murillo, Sriram Subramanian, and Diego Martinez Plasencia. 2017. Erg-O: Ergonomic Optimization of Immersive Virtual Environments. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 759–771. https://doi.org/10.1145/3126594.3126605

[116] Martez Mott, Edward Cutrell, Mar Gonzalez Franco, Christian Holz, Eyal Ofek, Richard Stoakley, and Meredith Ringel Morris. 2019. Accessible by design: An opportunity for virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 451–454. https://doi.org/10.1109/ISMAR-Adjunct.2019.00122

[117] Martez Mott, John Tang, Shaun Kane, Edward Cutrell, and Meredith Ringel Morris. 2020. "I Just Went into It Assuming That I Wouldn't Be Able to Have the Full Experience": Understanding the Accessibility of Virtual Reality for People with Limited Mobility. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 43, 13 pages. https://doi.org/10.1145/3373625.3416998

[118] Patrick Ngatchou, Anahita Zarei, and A El-Sharkawi. 2005. Pareto multi objective optimization. In *Proceedings of the 13th international conference on, intelligent systems application to power systems*. IEEE, 84–91. https://doi.org/10.1109/ISAP.2005.1599245

[119] Sergiu Oprea, Pablo Martinez-Gonzalez, Alberto Garcia-Garcia, John A Castro-Vargas, Sergio Orts-Escolano, and Jose Garcia-Rodriguez. 2019. A visually realistic grasping system for object manipulation and interaction in virtual reality environments. *Computers & Graphics* 83 (2019), 77–86. https://doi.org/10.1016/j.cag.2019.07.003

[120] World Health Organization. 2022. *Disability - Key Facts*. Retrieved January 20, 2023 from https://www.who.int/news-room/fact-sheets/detail/disability-and-health

[121] Francisco R Ortega, Katherine Tarre, Mathew Kress, Adam S Williams, Armando B Barreto, and Naphtali D Rishe. 2019. Selection and manipulation whole-body gesture elicitation study in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1723–1728. https://doi.org/10.1109/VR.2019.8798105

[122] Antti Oulasvirta and Kasper Hornbæk. 2022. Counterfactual thinking: What theories do in design. *International Journal of Human–Computer Interaction* 38, 1 (2022), 78–92. https://doi.org/10.1080/10447318.2021.1925436

[123] Antti Oulasvirta, Jussi P. P. Jokinen, and Andrew Howes. 2022. Computational Rationality as a Theory of Interaction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 359, 14 pages. https://doi.org/10.1145/3491102.3517739

[124] Yun Suen Pai, Tilman Dingler, and Kai Kunze. 2019. Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality* 23, 2 (2019), 119–131. https://doi.org/10.1007/s10055-018-0371-2

[125] Kseniia Palin, Anna Maria Feit, Sunjun Kim, Per Ola Kristensson, and Antti Oulasvirta. 2019. How Do People Type on Mobile Devices? Observations from a Study with 37,000 Volunteers. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 9, 12 pages. https://doi.org/10.1145/3338286.3340120

[126] Frol Periverzov and Horea Ilieş. 2015. IDS: The intent driven selection method for natural user interfaces. In *2015 IEEE symposium on 3D user interfaces (3DUI)*. IEEE, 121–128. https://doi.org/10.1109/3DUI.2015.7131736

[127] Duc-Minh Pham and Wolfgang Stuerzlinger. 2019. Is the Pen Mightier than the Controller? A Comparison of Input Devices for Selection in Virtual and Augmented Reality. In *25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) *(VRST '19)*. Association for Computing Machinery, New York, NY, USA, Article 35, 11 pages. https://doi.org/10.1145/3359996.3364264

[128] Jeffrey S. Pierce, Brian C. Stearns, and Randy Pausch. 1999. Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments. In *Proceedings of the 1999 Symposium on Interactive 3D Graphics* (Atlanta, Georgia, USA) *(I3D '99)*. Association for Computing Machinery, New York, NY, USA, 141–145. https://doi.org/10.1145/300523.300540

[129] Márcio S. Pinho, Doug A. Bowman, and Carla M.D.S. Freitas. 2002. Cooperative Object Manipulation in Immersive Virtual Environments: Framework and Techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (Hong Kong, China) *(VRST '02)*. Association for Computing Machinery, New York, NY, USA, 171–178. https://doi.org/10.1145/585740.585769

[130] Marcio S Pinho, Doug A Bowman, and Carla M Freitas. 2008. Cooperative object manipulation in collaborative virtual environments. *Journal of the Brazilian Computer Society* 14, 2 (2008), 53–67. https://doi.org/10.1007/BF03192559

[131] Thammathip Piumsomboon, David Altimira, Hyungon Kim, Adrian Clark, Gun Lee, and Mark Billinghurst. 2014. Grasp-Shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 73–82. https://doi.org/10.1109/ISMAR.2014.6948411

[132] Thammathip Piumsomboon, Gun Lee, Robert W Lindeman, and Mark Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE symposium on 3D user interfaces (3DUI)*. IEEE, 36–39. https://doi.org/10.1109/3DUI.2017.7893315

[133] Réjean Plamondon and Adel M Alimi. 1997. Speed/accuracy trade-offs in target-directed movements. *Behavioral and brain sciences* 20, 2 (1997), 279–303. https://doi.org/10.1017/S0140525X97001441

[134] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 79–80. https://doi.org/10.1145/237091.237102

[135] Ivan Poupyrev and Tadao Ichikawa. 1999. Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing* 10, 1 (1999), 19–35. https://doi.org/10.1006/jvlc.1998.0112

[136] Ivan Poupyrev, Tadao Ichikawa, Suzanne Weghorst, and Mark Billinghurst. 1998. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer graphics forum*, Vol. 17. Wiley Online Library, 41–52. https://doi.org/10.1111/1467-8659.00252

[137] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa. 1997. A framework and testbed for studying manipulation techniques for immersive VR. In *Proceedings of the ACM symposium on Virtual reality software and technology*. 21–28. https://doi.org/10.1145/261135.261141

[138] Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don't Have It: An Empirical Comparison of Head-Based and Eye-Based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 91–98. https://doi.org/10.1145/3131277.3132182

[139] Philip Quinn and Andy Cockburn. 2016. When Bad Feels Good: Assistance Failures and Interface Preferences. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4005–4010. https://doi.org/10.1145/2858036.2858074

[140] Iulian Radu, Tugce Joy, Yiran Bowman, Ian Bott, and Bertrand Schneider. 2021. A Survey of Needs and Features for Augmented Reality Collaborations in Collocated Spaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 169 (apr 2021), 21 pages. https://doi.org/10.1145/3449243

[141] Paul Davidson Reynolds. 2015. *Primer in theory construction: An A&B classics edition.* Routledge.

[142] Warren Robinett and Richard Holloway. 1992. Implementation of Flying, Scaling and Grabbing in Virtual Worlds. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics* (Cambridge, Massachusetts, USA) *(I3D '92)*. Association for Computing Machinery, New York, NY, USA, 189–192. https://doi.org/10.1145/147156.147201

[143] Katja Rogers, Jana Funke, Julian Frommel, Sven Stamm, and Michael Weber. 2019. Exploring Interaction Fidelity in Virtual Reality: Object Manipulation and Whole-Body Movements. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300644

[144] Majed Samad, Elia Gatti, Anne Hermes, Hrvoje Benko, and Cesare Parise. 2019. Pseudo-Haptic Weight: Changing the Perceived Weight of Virtual Objects By Manipulating Control-Display Ratio. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300550

[145] Mahnaz Samadbeik, Donya Yaaghobi, Peivand Bastani, Shahabeddin Abhari, Rita Rezaee, and Ali Garavand. 2018. The applications of virtual reality technology in medical groups teaching. *Journal of advances in medical education & professionalism* 6, 3 (2018), 123.

[146] Zhanna Sarsenbayeva, Niels Van Berkel, Eduardo Velloso, Jorge Goncalves, and Vassilis Kostakos. 2022. Methodological Standards in Accessibility Research-PLXBCCR- on Motor Impairments: A Survey. *ACM Comput. Surv.* 55, 7, Article 143 (dec 2022), 35 pages. https://doi.org/10.1145/3543509

[147] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of Three One-Question, Post-Task Usability Questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1599–1608. https://doi.org/10.1145/1518701.1518946

[148] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2021. Ninja Hands: Using Many Hands to Improve Target Selection in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 130, 14 pages. https://doi.org/10.1145/3411764.3445759

[149] Samuel B. Schorr and Allison M. Okamura. 2017. Fingertip Tactile Devices for Virtual Object Manipulation and Exploration. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3115–3119. https://doi.org/10.1145/3025453.3025744

[150] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108.* (2017). https://doi.org/10.9781/ijimai.2017.09.001

[151] Andrew Sears, Min Lin, Julie Jacko, and Yan Xiao. 2003. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. In *HCI international*, Vol. 2. 1298–1302.

[152] Ludwig Sidenmark, Christopher Clarke, Xuesong Zhang, Jenny Phu, and Hans Gellersen. 2020. Outline Pursuits: Gaze-Assisted Selection of Occluded Objects in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376438

[153] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

[154] Ludwig Sidenmark, Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2022. Weighted Pointer: Error-aware Gaze-based Interaction through Fallback Modalities. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3585–3595. https://doi.org/10.1109/TVCG.2022.3203096

[155] Mel Slater, Bernhard Spanlang, Maria V Sanchez-Vives, and Olaf Blanke. 2010. First person experience of body transfer in virtual reality. *PloS one* 5, 5 (2010), e10564. https://doi.org/10.1371/journal.pone.0010564

[156] Leonardo Pavanatto Soares, Regis Kopper, and Márcio Sarroglia Pinho. 2018. Ego-exo: A cooperative manipulation technique with automatic viewpoint control. In *2018 20th Symposium on Virtual and Augmented Reality (SVR)*. IEEE, 82–88. https://doi.org/10.1109/SVR.2018.00023

[157] Chang Geun Song, No Jun Kwak, and Dong Hyun Jeong. 2000. Developing an Efficient Technique of Selection and Manipulation in Immersive V.E.. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (Seoul, Korea) *(VRST '00)*. Association for Computing Machinery, New York, NY, USA, 142–146. https://doi.org/10.1145/502390.502417

[158] Suzanne Sorli, Dan Casas, Mickeal Verschoor, Ana Tajadura-Jiménez, and Miguel A Otaduy. 2021. Fine Virtual Manipulation with Hands of Different Sizes. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 304–309. https://doi.org/10.1109/ISMAR52148.2021.00046

[159] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300767

[160] Anthony Steed. 2006. Towards a general model for selection in virtual environments. In *3D User Interfaces (3DUI'06)*. IEEE, 103–110. https://doi.org/10.1109/VR.2006.134

[161] Frank Steinicke, Timo Ropinski, and Klaus Hinrichs. 2006. Object selection in virtual environments using an improved virtual pointer metaphor. In *Computer vision and graphics*. Springer, 320–326. https://doi.org/10.1007/1-4020-4179-9_46

[162] Rasmus Stenholt. 2012. Efficient Selection of Multiple Objects on a Large Scale. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology* (Toronto, Ontario, Canada) *(VRST '12)*. Association for Computing Machinery, New York, NY, USA, 105–112. https://doi.org/10.1145/2407336.2407357

[163] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 265–272. https://doi.org/10.1145/223904.223938

[164] Gail M Sullivan and Richard Feinn. 2012. Using effect size—or why the P value is not enough. *Journal of graduate medical education* 4, 3 (2012), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

[165] Ivan E. Sutherland. 1968. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I* (San Francisco, California) *(AFIPS '68 (Fall, part I))*. Association for Computing Machinery, New York, NY, USA, 757–764. https://doi.org/10.1145/1476589.1476686

[166] Ryo Suzuki, Eyal Ofek, Mike Sinclair, Daniel Leithinger, and Mar Gonzalez-Franco. 2021. HapticBots: Distributed Encountered-Type Haptics for VR with Multiple Shape-Changing Mobile Robots. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 1269–1281. https://doi.org/10.1145/3472749.3474821

[167] Christina Trepkowski, David Eibich, Jens Maiero, Alexander Marquardt, Ernst Kruijff, and Steven Feiner. 2019. The effect of narrow field of view and information density on visual search performance in augmented reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 575–584. https://doi.org/10.1109/VR.2019.8798312

[168] Huawei Tu, Susu Huang, Jiabin Yuan, Xiangshi Ren, and Feng Tian. 2019. Crossing-Based Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300848

[169] Allen B Tucker. 2004. *Computer science handbook*. Chapman and Hall/CRC.

[170] Lode Vanacken, Tovi Grossman, and Karin Coninx. 2007. 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. https://doi.org/10.1109/3DUI.2007.340783

[171] András Vargha and Harold D Delaney. 2000. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics* 25, 2 (2000), 101–132. https://doi.org/10.3102/107699860250021

[172] Jorge Wagner, Wolfgang Stuerzlinger, and Luciana Nedel. 2021. Comparing and combining virtual hand and virtual ray pointer interactions for data manipulation in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2513–2523. https://doi.org/10.1109/TVCG.2021.3067759

[173] Jia Wang and Robert W Lindeman. 2015. Object impersonation: Towards effective interaction in tablet- and HMD-based hybrid virtual environments. In *2015 IEEE virtual reality (VR)*. IEEE, 111–118. https://doi.org/10.1109/VR.2015.7223332

[174] Lili Wang, Jianjun Chen, Qixiang Ma, and Voicu Popescu. 2021. Disocclusion headlight for selection assistance in vr. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 216–225. https://doi.org/10.1109/VR50410.2021.00043

[175] Lili Wang, Xiaolong Liu, and Xiangyu Li. 2021. VR Collaborative Object Manipulation Based on Viewpoint Quality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 60–68. https://doi.org/10.1109/ISMAR52148.2021.00020

[176] Miao Wang, Zi-Ming Ye, Jin-Chuan Shi, and Yang-Liang Yang. 2021. Scene-context-aware indoor object selection and movement in vr. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 235–244. https://doi.org/10.1109/VR50410.2021.00045

[177] Yanbin Wang, Yizhou Hu, and Yu Chen. 2021. An experimental investigation of menu selection for immersive virtual environments: fixed versus handheld menus. *Virtual Reality* 25, 2 (2021), 409–419. https://doi.org/10.1007/s10055-020-00464-4

[178] Matthias Weise, Raphael Zender, and Ulrike Lucke. 2019. A Comprehensive Classification of 3D Selection and Manipulation Techniques. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 321–332. https://doi.org/10.1145/3340764.3340777

[179] Johann Wentzel, Greg d'Eon, and Daniel Vogel. 2020. Improving Virtual Reality Ergonomics Through Reach-Bounded Non-Linear Input Amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376687

[180] Jonathan Wieland, Johannes Zagermann, Jens Müller, and Harald Reiterer. 2021. Separation, Composition, or Hybrid?–Comparing Collaborative 3D Object Manipulation Techniques for Handheld Augmented Reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 403–412. https://doi.org/10.1109/ISMAR52148.2021.00057

[181] David Wilkie, Jason Sewall, Weizi Li, and Ming C Lin. 2015. Virtualized traffic at metropolitan scales. *Frontiers in Robotics and AI* 2 (2015), 11. https://doi.org/10.3389/frobt.2015.00011

[182] Graham Wilson, Mark McGill, Matthew Jamieson, Julie R. Williamson, and Stephen A. Brewster. 2018. Object Manipulation in Virtual Reality Under Increasing Levels of Translational Gain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173673

[183] C Wingrave and D Bowman. 2005. Baseline factors for raycasting selection. In *Proceedings of HCI International*. 61–68.

[184] Chadwick A Wingrave, Doug A Bowman, and Naren Ramakrishnan. 2002. Towards preferences in virtual environment interfaces. In *EGVE*, Vol. 2. 63–72. https://doi.org/10.5555/509709.509720

[185] Chadwick A Wingrave, Ryan Tintner, Bruce N Walker, Doug A Bowman, and Larry F Hodges. 2005. Exploring individual differences in raybased selection: strategies and traits. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* IEEE, 163–170. https://doi.org/10.1109/VR.2005.1492770

[186] Jacob O Wobbrock. 2019. Situationally-induced impairments and disabilities. In *Web accessibility*. Springer, 59–92. https://doi.org/10.1007/978-1-4471-7440-0_5

[187] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[188] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (apr 2016), 38–44. https://doi.org/10.1145/2907069

[189] Dennis Wolf, Jan Gugenheimer, Marco Combosch, and Enrico Rukzio. 2020. Understanding the Heisenberg Effect of Spatial Interaction: A Selection Induced Error for Spatially Tracked Input Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376876

[190] Huiyue Wu, Kaini Huang, Yanyi Deng, and Huawei Tu. 2022. Exploring the design space of eyes-free target acquisition in virtual environments. *Virtual Reality* 26, 2 (2022), 513–524. https://doi.org/10.1007/s10055-021-00591-6

[191] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173616

[192] Xin Yi, Leping Qiu, Wenjing Tang, Yehan Fan, Hewu Li, and Yuanchun Shi. 2022. DEEP: 3D Gaze Pointing in Virtual Reality Leveraging Eyelid Movement. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 3, 14 pages. https://doi.org/10.1145/3526113.3545673

[193] Difeng Yu, Ruta Desai, Ting Zhang, Hrvoje Benko, Tanya R. Jonker, and Aakar Gupta. 2022. Optimizing the Timing of Intelligent Suggestion in Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 20 pages. https://doi.org/10.1145/3526113.3545632

[194] Difeng Yu, Hai-Ning Liang, Xueshi Lu, Kaixuan Fan, and Barrett Ens. 2019. Modeling Endpoint Distribution of Pointing Selection Tasks in Virtual Reality Environments. *ACM Trans. Graph.* 38, 6, Article 218 (nov 2019), 13 pages. https://doi.org/10.1145/3355089.3356544

[195] Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-Supported 3D Object Manipulation in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 734, 13 pages. https://doi.org/10.1145/3411764.3445343

[196] Difeng Yu, Brandon Victor Syiem, Andrew Irlitti, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2023. Modeling Temporal Target Selection: A Perspective from Its Spatial Correspondence. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3544548.3581011

[197] Difeng Yu, Qiushi Zhou, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2022. Blending On-Body and Mid-Air Interaction in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 637–646. https://doi.org/10.1109/ISMAR55827.2022.00081

[198] Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Fully-occluded target selection in virtual reality. *IEEE transactions on visualization and computer graphics* 26, 12 (2020), 3402–3413. https://doi.org/10.1109/TVCG.2020.3023606

[199] Difeng Yu, Qiushi Zhou, Benjamin Tag, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Engaging participants during selection studies in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 500–509. https://doi.org/10.1109/VR46266.2020.00071

[200] Futian Zhang, Keiko Katsuragawa, and Edward Lank. 2022. Conductor: Intersection-Based Bimanual Pointing in Augmented and Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 560 (nov 2022), 15 pages. https://doi.org/10.1145/3567713

[201] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. 2019. SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300341

[202] Qiushi Zhou, Difeng Yu, Martin N Reinoso, Joshua Newn, Jorge Goncalves, and Eduardo Velloso. 2020. Eyes-free target acquisition during walking in immersive mixed reality. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3423–3433. https://doi.org/10.1109/TVCG.2020.3023570