Motion-based Interaction for Head-Mounted Displays

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Wenge Xu

May 2021

PGR Declaration of Academic Honesty

NAME (Print)	**Wenge Xu**
STUDENT NUMBER	**201324739**
SCHOOL/INSTITUTE	**University of Liverpool**
TITLE OF WORK	**Motion-based Interaction for Head-Mounted Displays**

This form should be completed by the student and appended to any piece of work that is submitted for examination. Submission by the student of the form by electronic means constitutes their confirmation of the terms of the declaration.

Students should familiarise themselves with Appendix 4 of the PGR Code of Practice: PGR Policy on Plagiarism and Dishonest Use of Data, which provides the definitions of academic malpractice and the policies and procedures that apply to the investigation of alleged incidents.

Students found to have committed academic malpractice will receive penalties in accordance with the Policy, which in the most severe cases might include termination of studies.

<u>STUDENT DECLARATION</u>

I confirm that:

- I have read and understood the University's PGR Policy on Plagiarism and Dishonest Use of Data.
- I have acted honestly, ethically and professionally in conduct leading to assessment for the programme of study.
- I have not copied material from another source nor committed plagiarism nor fabricated, falsified or embellished data when completing the attached material.
- I have not copied material from another source, nor colluded with any other student in the preparation and production of this material.
- If an allegation of suspected academic malpractice is made, I give permission to the University to use source-matching software to ensure that the submitted material is all my own work.

SIGNATURE……Wenge Xu……

DATE……7th May 2021……

Abstract

Recent advances in affordable sensing technologies have enabled motion-based interaction (MbI) for head-mounted displays (HMDs). Unlike traditional input devices like the mouse and keyboard, which often offer comparatively limited interaction possibilities (e.g., single-touch interaction), MbI does not have these constraints and is more natural because they reflect more closely people do things in real life. However, several issues exist in MbI for HMDs due to the technical limitations of the sensing and tracking devices, higher degrees of freedom afforded to users, and limited research in the area due to the rapid advancement of HMDs and tracking technologies.

This thesis first outlines four core challenges in the design space of MbI for HMDs: (1) boundary awareness for hand-based interaction, (2) efficient hands-free head-based interface for HMDs, (3) efficient and feasible full-body interaction for general tasks with HMDs, and (4) accessible full-body interaction for applications in HMDs. Then, this thesis presents an investigation into the contributions of these challenges in MbI for HMDs. The first challenge is addressed by providing visual feedback during interaction tailored for such technologies. The second challenge is addressed by using a circular layout with a go-and-hit selection style for head-based interaction using text entry as the scenario. In addition, this thesis explores additional interaction mechanisms that leverage the affordances of these techniques, and in doing so, we propose directional full-body motions as an interaction approach to perform general tasks with HDMs as an example to address the third challenge. The last challenge is addressed by (1) exploring the differences between performing full-body interaction for HMDs and common displays (i.e., TV) and (2) providing a set of design guidelines that are specific to current and future HMDs.

The results of this thesis show that: (1) visual methods for boundary awareness can help with mid-air hand-based interaction in HMDs; (2) head-based interaction and interfaces that take advantages of MbI, such as a circular interface, can be very efficient and low error hands-free input method for HMDs; (3) directional full-body interaction can be a feasible and efficient interaction approach for general tasks involving HMDs; (4) full-body interaction for applications in HMDs should be

designed differently than for traditional displays. In addition to these results, this thesis provides a set of design recommendations and takeaway messages for MbI for HMDs.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Hai-Ning Liang for his support, encouragement, and remarkable patience. I never expected to learn so much as a PhD student, and I could not have asked for a better person to work with. Without his guidance and constant feedback, this thesis would not have been possible.

Secondly, I sincerely extend my gratitude to my supervisory team (i.e., Prof. Yong Yue, Dr. Bing Wu Berberich, and Mr. Phil Jimmieson) and advisory members (i.e., Dr. Terry Pane, Dr. Xing Huang, and Dr. Jie Zhang) for their feedback and insightful comments on my work.

Thirdly, I would like to thank everybody at the X-CHI lab for being great friends, colleagues, and collaborators. Lei, Rongkai, Yiming, and Mengqi, I wish you guys an enjoyable PhD journey. Special thanks to my PhD colleague Diego Monteiro: thank you for everything in the past three years, I cannot wait to see you again in the UK.

Fourthly, I am very grateful and owe a lot to my family. I would like to thank my parents, Yunming Xu and Haiyan Tang, for their supports, care, and love. My brother Junkai Xu for having fun with me in my leisure time and holidays. Finally, special thanks to my fiancée Xiaolin Zhao for understanding my choice of doing a PhD and her encouragement and support throughout my PhD journey.

Lastly, thank you very much to everyone who helped me during my PhD studies.

Publications and Author's Contributions Statements

Materials from this dissertation have been previously published in the journals and conferences listed below. The publication from which each chapter is derived is noted in the parentheses.

1. Wenge Xu et al., "Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays," in Proceedings of 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China, 2019, pp. 279-288, doi: 10.1109/ISMAR.2019.00026. (Chapter 3)
2. Wenge Xu et al., "Exploring Visual Techniques for Boundary Awareness During Interaction in Augmented Reality Head-Mounted Displays," in Proceedings of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 2020, pp. 204-211, doi: 10.1109/VR46266.2020.00039. (Chapter 5)
3. Wenge Xu et al., "RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 5, pp. 1991-2001, May 2019, doi: 10.1109/TVCG.2019.2898736. (Chapter 6)
4. Wenge Xu et al., "DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 2019, paper 444, pp. 1–14, doi: 10.1145/3290605.3300674. (Chapter 7)
5. Wenge Xu et al., "Assessing the Effects of a Full-body Motion-based Exergame in Virtual Reality," in Proceedings of the Seventh International Symposium of Chinese CHI, Xiamen, China, 2019, pp. 1–6, doi: 10.1145/3332169.3333574. (Chapter 8)
6. Wenge Xu et al., "Studying the Effect of Display Type and Viewing Perspective on User Experience in Virtual Reality Exergames," in Games for Health Journal, 2020, ahead of print, doi: 10.1089/g4h.2019.0102. (Chapter 9)

Materials that contribute to this dissertation but not used as a chapter:

1. Wenge Xu et al., "Directional Motion-based Interfaces for Virtual and Augmented Reality Head-mounted Displays," in proceedings of 2018 International Computers, Signals and Systems Conference (ICOMSSC), Dalian, China, 2018, pp. 1-5, doi: 10.1109/ICOMSSC45026.2018.8942021.
2. Difeng Yu et al., "PizzaText: Text Entry for Virtual Reality Systems Using Dual Thumbsticks," in IEEE Transactions on Visualization and Computer Graphics, 2018, vol. 24, no. 11, pp. 2927-2935, Nov 2018, doi: 10.1109/TVCG.2018.2868581.
3. Difeng Yu et al., "DepthMove: Leveraging Head Motions in the Depth Dimension to Interact with Virtual Reality Head-Worn Displays," in Proceedings of 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China, 2019, pp. 103-114, doi: 10.1109/ISMAR.2019.00-20.
4. Wenge Xu et al., "Health Benefits of Digital Videogames for the Aging Population: A Systematic Review," in Games for Health Journal, 2020, ahead of print, doi: 10.1089/g4h.2019.0130.
5. Wenge Xu et al., "Results and Guidelines From a Repeated-Measures Design Experiment Comparing Standing and Seated Full-Body Gesture-Based Immersive Virtual Reality Exergames: Within-Subjects Evaluation," in JMIR Serious Games vol. 8, no. 3, pp. e17972, 2020, doi: 10.2196/17972.
6. Wenge Xu et al., "VirusBoxing: A HIIT-based VR Boxing Game," in Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY'20 EA), Virtual Event, Canada, 2020, pp 1-5, doi: 10.1145/3383668.3419958.
7. Xueshi Lu et al., "Exploration of Hands-free Text Entry Techniques For Virtual Reality," in Proceedings of 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Virtual, Brazil, 2020, pp 1-6, doi: 10.1109/ISMAR50242.2020.00061.
8. Wenge Xu et al., "Effect of Gameplay Uncertainty, Display Type, and Age on Virtual Reality Exergames," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 2021, pp 1-14, doi: 10.1145/3411764.3445801.

Table of Contents

List of Abbreviations

ANOVA	Analysis of Variance
AR	Augmented Reality
AvgHR%	Average Heart Rate Percentage
CC	Core Challenge
CER	Corrected Error Rate
CV	Commercial Version
CVSQ	Computer Vision Syndrome Questionnaire
DC	Dwell Circular
DCL	Dynamic Coordinate Line
DFC	Dwell-free Circular
DoF	Degrees of Freedom
DQ	Dwell QWERTY
DS	Dynamic Surface
DT	Display Type
EEG	Electroencephalography
ERP	Event-Related Potential
FEMD	Finger-Earth Mover’s Distance
FoV	Field-of-View
GEQ	Game Experience Questionnaire
HMD	Head-Mounted display
HR	Heart Rate
HWD	Head-Worn Display
IMI	Intrinsic Motivation Inventory
LD	Large Display
LPR	Letters per Region
MbI	Motion-based Interaction
MR	Mixed Reality
MSAQ	Motion Sickness Assessment Questionnaire
MTR	Missing Target Rate
NASA-TLX	National Aeronautics and Space Administration-Task Load Index
NCER	Not Corrected Error Rate

NE	North-East
NW	North-West
PAR-Q	Physical Activity Readiness Questionnaire
PID	Proportional–Integral–Derivative
RPE	Rating of Perceived Exertion
RQ	Research Question
SC	Swype Circular
SCL	Static Coordinate Line
SE	South-East
SPSS	Statistical Package for the Social Sciences
SQ	Swype QWERTY
SS	Static Surface
SSQ	Simulator Sickness Questionnaire
SUS	Slater-Usoh-Steed
SW	South-West
TER	Total Error Rate
UEQ	User Experience Questionnaire
VP	Viewing Perspective
VR	Virtual Reality
WPM	Words Per Minute
1pp	First-person Viewing Perspective
3pp	Third-person Viewing Perspective
%HRmax	Percentage of a Participant’s Estimated Maximum Heart Rate

List of Equations

List of Figures

List of Tables

Chapter 1 Introduction

Since Oculus raised a $2.5 million campaign for its Rift head-mounted displays (HMDs) series, there has been increasing popularity in the use of such devices. According to a report published in early 2020, sales of virtual reality (VR) HMDs will reach around seven million units, while augmented reality (AR) HMDs will climb to about 600 thousand units [254]. Forecasts project massive growth in HMD sales in the coming years, with HMDs expected to sell over 30 million units per year by 2023 [254]. As the popularity of HMDs rapidly increases, improving interaction performance and experience for HMDs is of great value.

There are two types of methods used by commercial HMDs: physical input devices (e.g., handheld controllers by Oculus Quest, Magic Leap 1) or motion-based input (e.g., bare hand interaction by Meta 2, head-based interaction by HoloLens). Handheld controllers, also called 3D mice [32], could provide position, orientation, and motion data of the hand in 3D space with the ability to complete a variety of tasks. However, controller-based interaction can be troublesome in many situations and may not be suitable for many users. Their limitations include, but not limited to, (1) batteries issues—(i) running out of batteries during the interaction could cause errors and hence lead to poor interaction performance and experience; (ii) going out for battery supplies during times that are risky for users to go out like COVID-19 pandemic period; (2) not suitable for people with special needs, such as users with hand disability due to hand tremors, could not manipulate a controller at all or with the precision required for certain tasks (e.g., text entry); (3) not suitable when the hands are occupied with other activities (e.g., cooking) and the controllers are not around (e.g., outdoor).

This thesis mainly focuses on motion-based input, not only because they could avoid the limitations of the controller-based interaction but also because they are natural, practical, and more suitable for HMDs [275]. The rest of the chapter introduces motion-based interaction (its definition, types, and what this thesis intended to investigate for these interactions), followed by research questions, thesis statement, contributions, and thesis organization.

Section 1.1 Motion-based Interaction

Aligned with a larger trend in human-computer interaction (HCI) around embodied interactions [59], motion-based interaction puts the body in the center of the interactive experience. In this thesis, the term motion-based interaction is defined as “interaction that relies on the changes in acceleration, orientation, the velocity of the user’s body part(s), where there is no need for direct contact with a pre-defined button or interactive surface.”

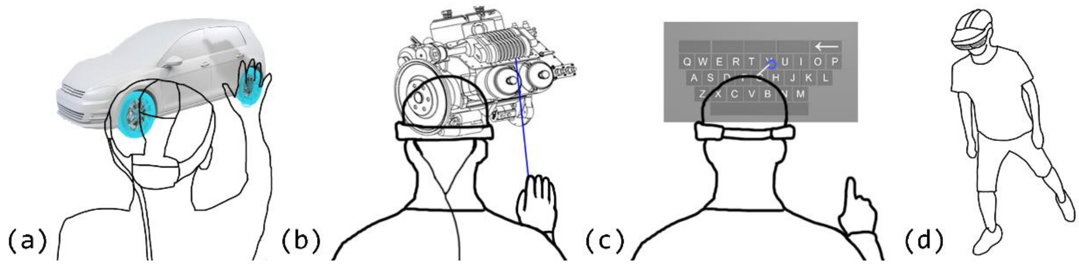


Figure 1-1. Commercially available motion-based interaction: Hand-based interaction (a) for selecting a nearby object, (b) for selecting a distant object. (c) Head-based (Head+Hand) interaction. Possible addition: (d) Full-body interaction.

Section 1.1.1 Hand-based Interaction

Hand-based interaction has gained rapid attention since projects like *Videoplace* [143] and the movie *Minority Report*. The proliferation of reasonably-priced motion-tracking cameras and sensors has warranted the possibility of hand-based interaction in many systems. For instance, it has been used for controlling (1) arbitrary medical computerized systems [24], (2) robotic hands [214], (3) unmanned aerial vehicles [118]. This interaction allows users to control a system without holding a physical input device, avoiding the issues brought by handheld controllers.

Several HMDs (e.g., Oculus Quest, HoloLens, Meta 2) have enabled hand-based interaction for interacting with 3D applications. There are two types of interaction depending on the location of the object: (1) for an object that is reachable by hand, users need to move their hand to the item that they want to select and perform a hand gesture (e.g., grab) to confirm the selection (see Figure 1-1a); (2) for a distant object, users use hand ray point to the object and perform finger gestures to confirm the selection (see Figure 1-1b).

Although hand-based interaction is often thought to allow natural user interfaces, designing a good hand interaction experience for HMDs is challenging. A typical reason is the lack of boundary information of the limited tracked area for the hand-based interactions. Because of this, users may easily move their hand(s) outside the tracked area during the interaction, especially in dynamic tasks (e.g., when translating an object). Boundary awareness issues have been observed in early works with other displays [48,52,183] and are also an issue for HMDs due to technical limitations of the motion sensors. Therefore, the research presented in this thesis firstly aims to investigate how boundary awareness can be provided in HMDs during hand-based interaction.

Section 1.1.2 Head-based Interaction

Even though it is rarely studied in other displays, head-based pointing, controlled by head motions, has become one of the primary metaphors for acquiring targets in current HMDs [9]. A ray is cast from the virtual camera to the virtual environment to serve as a pointing mechanism, where the end of the ray is akin to a cursor. This head-based pointing method is often used together with hand gestures, where users need to use their head to move the cursor to point to the target and perform hand gestures for confirming selection (see Figure 1-1c). This type of hybrid interaction suffers hand-related issues since it involves using the hand for indicating the selection.

Dwell technique has been used for head-based interaction to enable hands-free interaction [294]. Instead of using a hand gesture for indicating a selection, Head+Dwell selects the target by dwelling over it for a period of time. However, this technique also has limitations. For instance, a long dwell time could decrease the performance, while a short dwell time could cause errors [127]. In addition, the pre-set dwell time always “pushed” users to make quick decisions, which could be stressful [142]. Thus, the research presented in this thesis secondly aims to explore how to design an efficient hands-free and dwell-free head-based interaction for HMDs.

Section 1.1.3 Full-body Interaction

Instead of just using the hand or head gestures (motions), full-body motion-based interaction uses the human body as a whole unit [20] (see Figure 1-1d). This type of interaction has been initially widely studied in video games [85,176,196] and has been now used in a broader context such as museums [211], motor rehabilitation [233], learning environments [172]. Full-body interaction could avoid the pitfalls of hand-based interaction (i.e., arm/hand fatigue). Besides, it encourages physical activity in offices and homes and, as such, can bring health benefits to users who are living a sedentary lifestyle—e.g., just ten minutes of physical activity can help users gain cognitive and physical benefits [137].

Although full-body interaction could provide various benefits to the HMD users, the feasibility of this type of interaction for HMDs remains unknown (i.e., studies on full-body motion-based interaction were with computer monitors or televisions, where motion sickness is not an issue). Thus, the research presented in this thesis thirdly aims to investigate the feasibility and efficiency of full-body interaction for general tasks with HMDs and finally aims to explore full-body interaction for applications in HMDs.

Section 1.2 Research Questions

This thesis aims to answer the following research questions (RQ):

RQ1 – Chapter 5 How can visual boundary awareness techniques support mid-air hand-based interaction?

RQ2 – Chapter 6 Can other types of non-standard interfaces, such a circular layout, achieve an efficient hands-free head-based interaction?

RQ3 – Chapter 7 Are directional full-body interaction feasible and efficient for general tasks with HMDs?

RQ4 – Chapter 8-9 Will HMDs affect users experiencing full-body interaction?

RQ5 – Chapter 8-9 Will sickness mitigation factors in other contexts work for full-body motion-based interaction?

Section 1.3 Thesis Statement

The goal of this dissertation is to design motion-based interaction techniques and interfaces for HMDs with consideration of user performance and user experience. In specifics, this research focuses on addressing the following Core Challenges (CC) in motion-based interaction for HMDs: (CC1) boundary awareness for hand-based interaction, (CC2) efficient hands-free head-based interface for HMDs, (CC3) efficient and feasible full-body interaction for general tasks with HMDs, and (CC4) accessible full-body interaction for applications in HMDs.

Section 1.4 Contributions

This dissertation addresses the Core Challenges in motion-based interaction and interfaces for HMDs. In this context, it makes the following main contributions:

- Visual methods for boundary awareness can help with hand-based interaction in HMDs, but their effectiveness and application are user-dependent (CC1).
- Head-based interaction with other types of interfaces, such as a circular layout for a typical keyboard, can be valuable additions to dwell-, device-, and hands-free interaction for HMDs. It is an efficient and low error input technique for HMDs (CC2).
- Directional motion-based interaction can be an efficient and feasible input technique for general tasks with HMDs. It could outperform (1) hand-based interaction regarding task performance and user experience and (2) hybrid-based (head+hand) interaction in user experience (CC3).
- Providing a list of full-body gestures and design guidelines for full-body exergame in HMDs (CC4).

Section 1.5 Dissertation Organization

As shown in Figure 1-2, in the following three chapters, we focus on identifying Core Challenges of motion-based Interaction for HMDs. Specifically, Chapter 2 and Chapter 3 identify challenges of motion-based Interaction for HMDs and Chapter 4 summarizes four core challenges that we selected to address in this thesis. From Chapter 5 to 9, we focus on addressing these four Core Challenges. Finally, in Chapter 10, we discuss the findings of this thesis, conclude the work and list future work of motion-based interaction for HMDs. Details of the dissertation organization shown below.

Chapter 2: Literature Review – This chapter provides a detailed description of HMDs (i.e., history, types of HMDs in the consumer market, user experience-related issues with current HMDs), and summarizes current and potential motion-based interactions that can be used for HMDs.

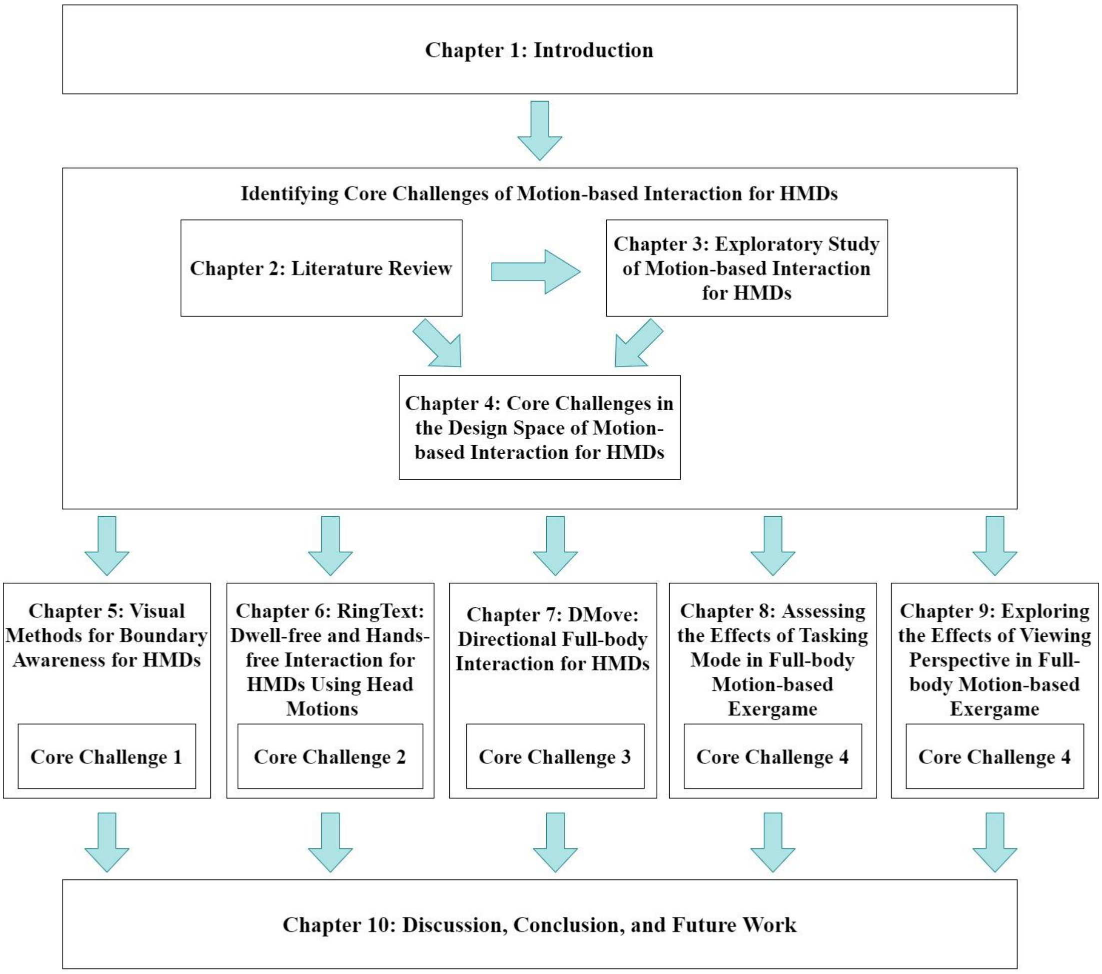


Figure 1-2. Structure of the thesis.

Chapter 3: Exploratory Study of Motion-based Interaction for HMDs – This part of the dissertation aims to confirm the validation of issues pointed by Chapter 2 (i.e., boundary awareness for hand-based interaction and efficiency issue of Head+Dwell technique).

Chapter 4: Core Challenges and Research Questions in the Design Space of Motion-based Interaction for HMDs – This chapter first lists four Core Challenges needed to be addressed in this thesis based on findings from the literature review (i.e., Chapter 2) and exploratory study (i.e., Chapter 3). Then it explains why other challenges (e.g., sweating in HMDs, gesture vocabulary, the ideal gesture set) are not covered in this thesis. The four Core Challenges selected are: (i) boundary awareness for hand-based interaction, (ii) efficient hands-free head-based interaction for HMDs, (iii) efficient and feasible full-body interaction for general tasks with HMDs, and (iv) accessible full-body interaction for applications in HMDs.

Chapter 5: Visual Methods for Boundary Awareness for HMDs – In this chapter, we explore visual techniques for boundary awareness in HMDs, focusing on object translation tasks. Through a systematic formative study, we first identify the challenges that users might face when interacting with HMDs without any boundary awareness information (i.e., how current systems work). Based on the findings, we then propose four methods (i.e., static surfaces, dynamic surface(s), static coordinated lines, and dynamic coordinate line(s)) and evaluate them against the benchmark (i.e., baseline condition without boundary awareness) to make users aware of the tracked interaction area. Our results show that visual methods for boundary awareness can help with dynamic mid-air hand interactions in HMDs, but their effectiveness and application are user-dependent.

Chapter 6: RingText: Dwell-free and Hands-free Interaction for HMDs Using Head Motions – In this chapter, we present a case for text entry using a circular keyboard layout for HMDs that is hands-free for letter selection. The design of RingText follows an iterative process, where we initially conduct one first study to optimize its design. Our second study compares the text entry performance of RingText with four other hands-free techniques and the results show that RingText outperforms them. Finally, we run a third study lasting four consecutive days with ten participants (five novice users and five expert users) doing two daily sessions and the results show that RingText is quite efficient and yields a low error rate. At the end of the eighth session, the novice users can achieve a text entry speed of 11.30 words per minute (WPM) after 60 minutes of training while the expert (more experienced) users can reach an average text entry speed of 13.24 WPM after 90 minutes of training.

Chapter 7: DMove: Directional Full-body Interaction for HMDs – This chapter presents DMove, directional full-body interaction for HMDs that is hands-free and device-free. It uses directional walking as a way to interact with virtual objects. To use DMove, a user needs to perform directional motions such as moving one foot forward or backward. We compare DMove with two approaches—hand-based interaction and hybrid-based (head+hand) interaction for menu selection tasks. Results show DMove causes fewer errors than hand-based interaction, leads to a lower overall workload than hand-based interaction, and brings a better user experience than hand-based interaction and hybrid-based interaction.

Chapter 8: Assessing the Effects of Tasking Mode in Full-body Motion-based Exergame – This chapter investigates the effect of the (1) task mode (single- and multi-tasking) on exergame and (2) explore the differences of user performance and experience in between playing exergame between HMDs and a 50-inch 4K TV. Findings show that (1) participants have the same level of game experience and simulator sickness when playing the exergame in VR and Large Display; (2) VR has increased participants' Theta wave; (3) participants believe multi-tasking is more challenging and show a higher level of simulator sickness than single-tasking; (4) participants have a worse game performance in multi-tasking than single-tasking.

Chapter 9: Exploring the Effects of Viewing Perspective in Full-body Motion-based Exergame – This chapter investigates the effect of the (1) viewing perspective (first-person and third-person perspective) on exergame and (2) explore the differences of user performance and experience in between playing exergame between HMDs and a 50-inch 4K TV. This study demonstrates that youth who played gesture-based exergame in HMD had a higher level of exertion (%HRmax, calories consumption, and Borg RPE), although the number of performed gestures were not significantly different. They also felt that HMD was much more challenging, immersive (flow, sensory and imaginative immersion), and had a lower negative effect than the TV; however, HMD was more likely to make youth have higher cybersickness.

Chapter 10: Discussion, Conclusion, and Future Work – This chapter first presents how we addressed the Core Challenges and Research Questions; meanwhile, it lists contributions that have been made for each Core Challenge. Then, this chapter proposes a set of design recommendations and takeaway messages of motion-based interactions for HMDs. Finally, it concludes the dissertation and provides future work on motion-based interaction for HMDs.

Chapter 2 Literature Review

This chapter reviews the existing literature, that is essential to the investigations within this thesis, on head-mounted display (HMD) technologies and their limitations, as well as motion-based interactions for HMDs.

Section 2.1 Head-mounted Displays

In 1960, Morton Heilig invented the first HMD, Telesphere Mask, which can provide stereoscopic 3D and comprehensive vision with stereo sound and play non-interactive films. However, unlike modern HMDs, Telesphere Mask is not capable of motion tracking. The first motion tracking-enabled HMD—Headsight was created in 1961. It was developed for immersive remote viewing of dangerous situations by the military. The user's head movements would move a remote camera, allowing the user to look around the environment naturally, allowing the user to look around the environment naturally. Headsight lacked the integration of computer and image generation.

The first computer-supported HMD was developed in 1968 by Sutherland [256]. However, due to its weight, it had to be suspended from the ceiling. The user also had to be strapped into the device. The graphics generated by the computer were wireframe rooms and objects. After that, HMDs had been used for many different applications and received significant research and development in the 1990s. However, the HMDs did not reach consumers. Possible reasons are poor technical quality (e.g., resolution, field-of-view, comfort), graphical quality, and prohibitive cost [121]. The precise dividing line between the commercial failures of consumer HMD in the past and the modern HMD revolution happened in 2012, where Oculus raised a $2.5 million campaign for its product Rift. With its two development kits released in 2013 and 2014, Oculus released its first commercial version (CV)—Oculus CV 1 ($699) in 2016. Since then, affordable HMDs such as HTC VIVE ($799), Microsoft HoloLens 1 ($3000), Magic Leap 1 ($2300) have become more readily available to the public. The International Data Corporation report[1] indicates the worldwide shipments of HMDs are expected to reach 7.1 million units in 2020 to 76.7 million units in 2024.

[1]https://www.idc.com/getdoc.jsp?containerId=prUS46143720

Section 2.1.1 Tethered HMD

Tethered HMDs are HMDs connected to powerful and expensive gaming desktops. A typical device of this type is Oculus Rift Virtual Reality (VR) HMD, Meta 2 Augmented Reality (AR) HMD, and HP Windows Mixed Reality (MR) HMD, with others, like HTC VIVE, Valve Index, Lenovo Explorer. All tethered HMDs are capable of six degrees of freedom (DoF) tracking, which means that they could check the user's position and orientation. In addition, since a dedicated gaming desktop does the computing power, the tethered HMDs are generally capable of rendering vibrant graphical scenes at high frame rates and visual quality. However, this type of HMD has limited mobility and is high cost.

Section 2.1.2 Standalone HMD

Standalone HMDs are a more portable and accessible way to interact with the extended reality (AR/VR/XR) applications since these contents are installed and run on the device itself or the smartphone connected to the HMD. Standalone HMDs are widely available due to their low cost, simple setup process, light, and accessibility, but suffer from low graphical quality, poor battery life, and thermal radiation [27].

Standalone HMDs can be classified depending on the source of computing power and their DoF tracking.

1. *Smartphone-empowered HMDs*: this type of HMDs (e.g., Google Cardboard or Gear VR) requires an additional smartphone to be connected to the HMDs (typically via USB-C or micro-USB) and acts as the headset's display, processor, and rotational tracker (i.e., smartphone's built-in inertial measurement unit). Due to the absence of RGB and depth cameras, this type of HMDs can only provide 3-DoF tracking (i.e., rotational tracking), limiting its experience.
2. *All-in-one HMDs*: This type of HMDs contains all necessary components such as processor, operating system (built-in Android or Windows system), the audio system to provide an extended reality experience. Depending on their DoF tracking capabilities, they can be further classified into all-in-one (i) 3-DoF HMDs and (ii) 6-DoF HMDs. All-in-one 3DoF HMDs (e.g., Oculus Go, Xiaomi Mi) rely on its built-in inertial measurement unit to

provide rotational tracking. On the other hand, all-in-one 6-DoF HMDs (e.g., Oculus Quest, Microsoft HoloLens, Magic Leap) provide both rotational and positional tracking by on-device RGB and depth cameras. This type of HMDs has gained rapid attention due to its low cost when compared with tethered HMDs.

Section 2.1.3 User Experience-related Issues with Current HMDs

One major drawback that frequently happened on HMDs, especially for VR HMDs, is sickness. It has been found that many people report experiencing simulator sickness symptoms (e.g., headaches, stomach awareness, nausea, disorientation [150]) from HMD use [64,84,102,218]. Here we classified these factors into hardware, human, and content.

Hardware field-of-view (FoV) is one of the hardware-related features that could lead to VR sickness. Literature shows that reducing the hardware FoV could alleviate users' discomfort. Several methods have been developed to reduce sickness: (1) changing the size of display or distance between the user and the screen [243], (2) using a dynamic FoV system based on the electrophysiological signals of the participant [135], (3) applying content-type based directional FoV restriction method [134]. Latency could also lead to sickness, especially when the latency is inconsistent during the experience [61]. In addition, DiZio and Lackner [57] suggest that the severity of sickness increased as the latency increased. However, this is not supported by [61]. Overall, to minimize sickness caused by latency, designers should minimize the latency and keep it consistent throughout the VR experience.

Age, gender, motion sickness susceptibility are common human factors that are discussed in VR sickness. Regarding age, a recent meta-analysis by Saredakis et al. [231] suggests that younger adults (<35) often reported a higher simulator sickness compared with the older age group. The effect of gender on sickness also remains mixed, with a few studies suggesting females are more susceptible to VR sickness than males [103]. However, a meta-analysis by Saredakis et al. [231] did not find a significant correlation between gender and sickness. Motion sickness susceptibility could also act as an essential index for predicting the degree of VR sickness. Several studies indicate that users vulnerable to motion sickness are likely to report higher VR discomfort [157,253].

Content-related factors can be associated with VR sickness. A typical content-related factor is the reference frame. Studies suggest that VR sickness decreases when fixed visual stimuli are presented (e.g., a virtual nose [274]). Secondly, the duration of VR experience could also lead to sickness; literature suggests that users can experience sickness even for a short period (<10 min) of VR play [55]. Furthermore, body motion also plays a crucial role in inducing simulator sickness. Rotational movements could lead to more significant discomfort when compared to translational movements [26,132]. Furthermore, the discomfort could be worsened when the dual-axis are involved [132].

Overall, this thesis has considered sickness an essential factor that has to be measured during HMD usage, especially interaction involving rapid head and full-body motions. This thesis also aims to address sickness by identifying possible factors that could lead to higher sickness levels (see Chapter 8 and Chapter 9).

Section 2.2 Motion-based Interaction for HMDs

As mentioned in the introduction, the term of motion-based interaction in this thesis is defined as "interaction that relies on the changes in acceleration, orientation, the velocity of the user's body part(s), where there is no need for direct contact with a pre-defined button or interactive surface." Here, we summarize four methods that can be used for HMDs.

Section 2.2.1 Hand-based Interaction

Hand-based interaction is one of the most commonly used selection methods for HMDs [170] because it is assumed to be natural and practical. For selecting a near object [182], users first need to choose the target object by hovering the hand over it and then selecting it by performing a gesture—e.g., in Meta 2, users select the item by making a grab gesture. To select an item that is placed further away from the user, Mine [182] suggests that users can utilize their finger to point at the object, followed by a selection gesture. Studies have looked at finger-pointing [22,154], but these techniques require an additional external sensor like Kinect placed at a distance to detect and classify the gestures. Recently, Oculus Quest and HoloLens 2 have proposed palm pointing, where a ray is extended from users' palms towards the virtual objects; however, these interactions suffer from limitations such as sensitivity to lighting conditions and line-of-sight conditions of the motion camera.

One limitation of mid-air hand interaction is boundary awareness (or lack of it), which is an issue that can occur in motion tracking applications that rely on any sensor. For instance, for mid-air interaction, in particular, the user's hand can easily go out of or leave the tracking volume (or area) that the devices' sensor(s) can capture, but the user may not have a conscious awareness that their hands are no longer tracked [183]. This has been observed in early works with motion tracking devices such as Leap Motion [52,183] and Kinect [48] that unavoidably had a restricted tracked area due to technical limitations.

In HMDs, motion sensors are embedded in the front of the HMD. Hence, users are required to keep their hands to chest level, which is uncomfortable and can quickly lead to fatigue [114,251]. Because of the small tracked area by the motion sensors and the fatigue during the interaction, there are chances that users could move their hands off the motion camera's tracked area (i.e., boundary awareness), which often leads to issues such as misrecognition of gestures, registration errors [144]. Therefore, there is a need first to confirm whether this is an issue for HMDs and address it if it is a validated issue.

Section 2.2.2 Head-based Interaction

Head-based interaction has been actively studied for HMDs [32,46,147]. Relying on the HMDs' built-in IMU sensors, head-based pointing has been widely adopted as a standard way of interacting with virtual objects in HMDs [147]. However, head-based pointing alone can only identify the target object to be selected. It lacks an integrated method to indicate selections [182].

Several methods have been proposed to fill this gap for head-only input. One possible solution is the crossing-based technique. It has been successfully used with many input methods (e.g., gaze input [145], paper-based input [60], and direct touch input [160]) and applied on many display systems (e.g., desktop [8], touchscreen [160], remote screen [194]). Yan et al. [289] proposed *HeadCross*, which allows users to select an object by moving the pointer across the target boundary and then turn it back immediately. This goal crossing selection design is more expressive than the point and click interfaces [2,8,160] and allows users to issue several actions in one single stroke [8,56]. However, crossing requires users to make additional turning for selection quickly, which could increase the risk of simulator sickness [295].

The most used method is dwell, where users need to turn the head and move the cursor over the target object in a fixed time [126,202,245,263]. It has been used for menu selection and text entry [294]. Research [45] suggests that the dwell-based technique could lead to fewer errors and is perceived as more usable, more comfortable, and less fatiguing than the touchpad-based technique. However, the dwell technique is slower than the touchpad-based technique.

Since head-only input lacks efficiency, head-based interaction has been coupled with other input modalities. One typical motion-based input that is often used with head-based interaction is hand-based gestures. For instance, HoloLens require users to move the cursor to the target item and select it by finger “air tap” gesture. Although this hybrid (Head+Hand) interaction style might improve the efficiency and avoid extra sickness that causes by additional rapid head motions [295], it unavoidably suffers issues that are related to hand-based interaction (i.e., see Hand-based Interaction).

Section 2.2.3 Foot-based Interaction

Foot-based interaction techniques [265] have been widely explored for many scenarios (e.g., interactive animation system [292], 3D interaction tasks [249], and navigating spatial data [234]) in different using poses (e.g., seated [264], standing [232], and walking [288]). It can be grouped into two categories based on how foot actions are mapped to system commands [5]: (1) *Discrete* foot gestures are mapped to specific tasks. For instance, it has been widely researched for operating an in pocket mobile phone (e.g., locking and unlocking a mobile phone, making a phone call, performing foot-step to operate a menu selection system while jogging [49,237,288]); (2)

Continuous gestures are those that are mapped to tasks with a spatial component (e.g., moving the foot in one direction in space). It has been widely used in many areas, includes but not limited to target selection in a desktop computer [116], making a menu selection with mobile applications [219], navigating spatial data with a large display [234].

Pure foot-based interactions have been proposed to increase the input space for desktop [248] and mobile [18,70]. It has also been used in conjunction with many other input methods. For instance, foot interaction has been used with (1) multi-touch hand gestures for navigating spatial data with a large display [234] and playing games on mobile phone [162], (2) mid-air hand gestures for interacting with a handheld device [161,163], (3) head motions to navigate for the game World of Warcraft [248], (4) gaze input for interacting with the desktop environment [89,216], and (5) mouse and keyboard for target selection [232].

Since the emergence of HMDs (e.g., VR and AR) in 2012, the applicability of foot-based interaction for HMDs has been studied. Early work by Matthies et al. [177] presented a proof of concept wearable foot interface prototype to provide hands-free interaction for virtual and real environments. Later, Fukahori et al. [76] used sock-placed pressure sensors to detect the shifting of the user’s weight on their foot for subtle gestures to control HMDs interfaces. Recently, Muller et al. [192] proposed foot tap-based interaction for HMDs using an optical tracking system. Furthermore, foot-based interactions have been used (1) as locomotion technique for HMDs [277], (2) for controlling an AR game [70], and (3) for exploring a VR representation of a planet [69].

Section 2.2.4 Full-body Interaction

Instead of just relying on the hand or head gestures (motions), full-body motion-based interaction uses the human body as a whole unit [20]. Full-body interaction has been widely used/studied in video games because (1) the consumer level motion-tracking devices (i.e., Kinect) were initially published with video game consoles, and (2) video game is a good platform for exploring gestures and testing gesture recognition performance [32]. Nowadays, full-body interaction has been now used in a broader context such as museum [211], motor rehabilitation [233], learning environments [172].

Full-body interaction could avoid the pitfalls of hand-based interaction (i.e., arm/hand fatigue—holding the hand in the mid-air for long periods). In addition, it can encourage physical activity in offices and homes and, as such, can bring health benefits to their users—e.g., just ten minutes of physical activity can help users gain cognitive and physical benefits [137].

Despite the potentials and benefits of full-body interaction are promising, it only receives limited attention for HMDs. One reason is that the feasibility of this interaction for HMDs remains unknown because HMDs could bring motion sickness and related issues to users. Hence, designing feasible full-body interaction is a key research challenge. Addressing this research challenge and providing design guidelines for full-body interaction for HMDs could benefit many applications (e.g., exergame [85], rehabilitation [233], learning [172]).

Section 2.3 Summary

This chapter presents an overview of the field of motion-based interaction for HMDs. We provided a detailed description of HMDs regarding their history, type of HMDs available in the current consumer market, and user experience issues while using HMDs. Then, we define motion-based interaction used in this thesis and present four types of motion-based interaction that can be used for HMDs.

The literature review shows that (1) there is a lack of comparison between motion-based interaction and controller-based interaction, (2) boundary awareness issues might affect mid-air hand-based interaction in HMDs, (3) there might be a need to propose a hands-free efficient head-based input because Head+Dwell can be inefficient, (4) there is a need to design feasible full-body interactions for HMDs.

The following chapter outlines an exploratory study to compare commercially used motion-based interaction with controller-based interaction. Most importantly, it is to confirm whether (1) boundary awareness is an issue for hand-based interaction for HMDs and (2) current hands-free head-based interaction is inefficient.

We selected text entry as the interaction task in the exploratory study (i.e., next chapter) because it is an essential activity in all interactive systems, including HMDs [32], and it is a relatively new research area with modern HMDs. In addition, it is also a major interface for many content production applications, including but not limited to document editing, programming, web browsing [98]. Further, text entry activities like instant messaging and email communication are common platforms for communicating with family, friends, and colleagues.

In addition, text entry is chosen because it could expose issues that we explored from the literature review (see Table 2-1):

1. Hand-based: to select the letter from the virtual keyboard, users have to hold and move their hand in the mid-air, which could cause arm and hand fatigue [114,251]. This tiredness would likely cause their hands to gradually move outside the interaction area (lack of boundary awareness [183]).
2. Head+Hand: to indicate a selection, users are required to keep their hands to chest level, which is uncomfortable and can quickly lead to fatigue [114,251], leading to move their hands outside the interaction area (lack of boundary awareness [183]).
3. Head+Dwell: to indicate a selection, users must keep the cursor staying on the target for a period of time. As mentioned in [127], a long dwell time could decrease the performance, while a short dwell time could cause errors. Besides, a pre-set dwell time always “pushed” users to make quick decisions, which could be stressful for users [142].

Table 2-1. Interaction techniques tested in the next chapter and the related issues that would occur based on the literature.

Technique	Issue
Hand	Arm and hand fatigue [114,251]; Boundary awareness [183]
Head+Hand	Arm and hand fatigue [114,251]; Boundary awareness [183]
Head+Dwell	Dwell-related performance issue [127]; Stress [142]

Chapter 3 Exploratory Study of Motion-based Interaction for HMDs

Section 3.1 Introduction

Text entry is an essential activity in all interactive systems, including virtual reality (VR) and augmented reality (AR) head-mounted displays (HMDs). There have been some advances in this area for VR [98,138,294,295], but it is still quite underexplored for AR. Unlike VR, AR users can see through the transparent HMD and it is possible to access a physical keyboard. For example, the HoloLens can connect to a wireless physical keyboard. However, traditional input devices such as mice and keyboards are not suitable for outdoor environments, as they require a type of flat surface to operate on [260]. Moreover, AR HMDs are meant to be mobile devices that enable users to move within both indoor and outdoor environments [66,156]. Therefore, using a physical keyboard can be useful for text entry in VR settings [98] as the VR HMDs are commonly used in indoor scenarios, but it is unlikely the most suitable way for AR HMDs.

Text entry in AR differs from VR in many aspects. The hand representation can be hidden or virtually presented in VR [99] but not for AR HMDs. There are some known issues that only exist in AR, including layer interference, color blending problem, and layout foreground-background. These issues affect the text readability, visibility, depth ordering, object segmentation, and scene distortion [144] and make it difficult for users to acclimate to the content viewed through see-through displays [198]. Since the text and the virtual keyboard are typically viewed in a fixed location within an HMD screen, other people and objects in the background can become noise and hinder accomplishing various tasks, including entering text.

Early work has investigated using a glove for AR HMDs to interact with the system to support direct manipulation of virtual objects, interaction with symbolic data (e.g., text entry), and doing military logistics tasks in both indoor and outdoor settings [260]. However, current AR HMDs do not come with an expensive glove specially designed to support such interactions. On the other hand, pointing methods are not only low-cost but can also be used in both indoor and outdoor scenarios. In addition to head-based pointing, other methods rely on the user's hand or involve a handheld device for cursor positioning. Pointing methods are widely used in both VR and AR HMDs and as such it is worth exploring their suitability and relative performance with virtual keyboards. In this research, our primary goal is to explore pointing methods in AR that can work with a virtual keyboard and does not rely on specialized peripheral devices (i.e., Chord [164]) that typically do not come with the AR HMDs.

Our exploration considers three user case scenarios.

1. *When users have access to a ray-casting handheld device.* The assumption is that the users have access to a controller that can interact with an AR HMDs using ray-casting, a technique commonly used in VR HMDs and is also available in AR HMDs [9]. For example, the Magic Leap 1 provides a handheld controller that uses this technique.
2. *Hand-based but device-free*. There are two scenarios in this condition. (i) Hybrid interaction (head+hand), which relies on the use of the head to position the cursor pointer on the letters of the keyboard and the hand to trigger their selection. This approach has been used partially in some AR HMDs like HoloLens. (ii) Hand-based interaction, which only relies on mid-air hand motions to move a pointer over the letters and a hand gesture to indicate their selection. This approach has been used partially in the Meta 2 and it is thought to be one of the most natural selection methods used to interact with an AR environment [170].
3. *Both device-free and hands-free.* This represents the cases where no device is available, and it is based on head motions only for positioning the cursor and making letter selections. It is suitable for cases where users cannot use their hand or lift it comfortably (e.g., a user using AR HMD seating on a chair inside a bus that has limited space or with their hands encumbered because they are holding other objects). This is suitable also for environments that are too noisy for hand tracking (e.g., a user using the AR HMD while walking within a shopping mall because there are likely other moving objects in the background).

In short, we are comparing four standard, common HMDs Pointing Methods: *Head*, *Hand*, *Hybrid* (i.e., Head plus Hand like what HoloLens uses), and *Controller*. We also want to test two of the most common Input Mechanisms for making selections: *Tap* and *Swype*. Both Pointing Methods or Input Mechanisms have been partially studied for VR HMDs (e.g., [145,251,294]) but, to our knowledge, not for AR HMDs. Therefore, we want to compare eight text entry combinations of Pointing Methods and Input Mechanisms for text entry with respect to their performance, error rates, and user preferences. The results of our experiment with 24 participants (12 using Swype and 12 Tap) show that text entry performance of the Controller is comparable to other studies in VR [251,294] and non-VR [88,175]. When compared with all the three device-free pointing techniques, the Controller approach outperforms them in text entry performance and leads to better overall user experience. Our results also show that Swype is as fast as Tap and could cause lower uncorrected errors even for users who are new to Swype. On the other hand, these two input mechanisms do not show any significant difference in terms of a user's text entry experience, feeling of immersion, motion sickness, and most NASA TLX workload subscales. Finally, Swype is found to cause a heavier temporal workload and frustration than Tap.

Table 3-1 reviews examples of text entry techniques from other domains and devices that could be tailored for AR HMDs. To our knowledge, there has been no study that has explored text entry performance and user experience for AR HMDs. Our study represents the first systematic study of the eight possible combinations of Pointing Methods and Input Mechanisms. As such, the main contributions of this work include: (1) a first evaluation of four Pointing Methods × two Input Mechanisms (that is, eight possible combinations) for text input in AR HMDs regarding performance and user preference; (2) a set of design recommendations that are derived from our experimental results and observations during the experiment.

Section 3.2 Evaluated Text Entry Techniques

In this section, we describe how each combination of four Pointing Methods (Controller, Head, Hand, and Hybrid) and two Input Mechanisms (Tap and Swype) are operationalized in our experiment.

Table 3-1. Overview of text entry methods that have already been evaluated in VR that can potentially be used in AR (adapted from [251]): (1) *hands-only*, (2) *head-only*, (3) *hybrid*, (4) *controller*.

Pointing Method	Input Method	Qwerty	Eyes-free	Hands	Haptic feedback	Potential device-free for current AR HMDs?	WPM in VR	WPM other
(1)	Soft button selection	✓	×	1-2	×	×	4-7 [95]	33-36 [11]
(2)/(3)/(4)	Mid-air pointing	✓	×	1-2	×	(✓)	15.4 [251]	13-19 [175, 244]
(2)/(3)	Head pointing	✓	×	0-1	×	✓	10-15 [294]	4.5 [88]
(1)	Gamepad	×	(✓)	2	(✓)	×	8-15 [295]	6-7 [278]
(1)/(3)	Physical keyboard	✓	(✓)	1-2	✓	×	24-67 [138, 155]	45-67 [138]
(1)	Finger gestures	×	✓	1-2	×	(✓)	6 [95]	22-29 [252]
(1)	Chording	×	✓	1	✓	×	3 [95]	47 [164]
(1)	Multi-tap	×	✓	1	✓	×	12 [95]	20 [164]

Section 3.2.1 Controller

One of the most common ways of interacting with virtual environments and their objects is via a handheld controller [222]. The device uses a ray cast from it to the virtual environment to serve as a pointing mechanism. The end of the ray is akin to a cursor. To implement it, we have adapted the HTC VIVE controller (ray-casting enabled with at least one active button) and used the SteamVR Unity plugin to enable it to work with an AR HMD. The users would type on a virtual keyboard by merely moving the controller to point to the desired letters (see Figure 3-1a). Selection is done by either Tap or Swype.

Controller+Tap. To select a letter, the user needs to move the cursor to the letter on the virtual keyboard and press the trigger button for selection (see Figure 3-1a). A Tap action is also required to select a recommended word and special characters (e.g., space/backspace).

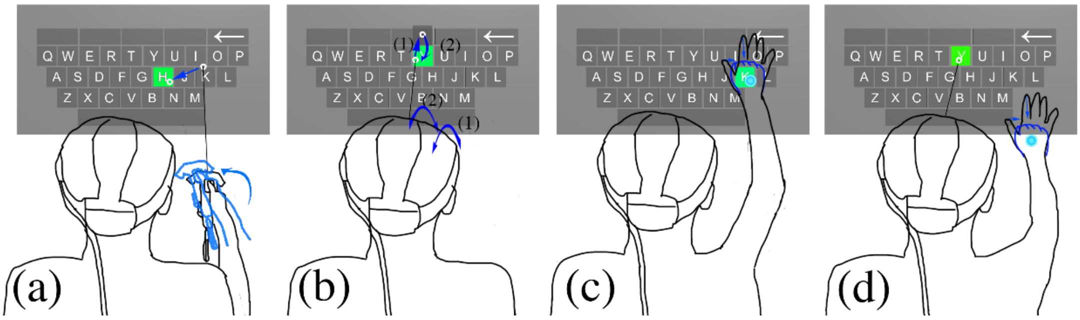


Figure 3-1. This figure shows how to Tap a letter by using the four pointing methods for AR HMDs. (a) *Controller*—the user uses a controller to move the cursor on the letter 'H', and then presses the trigger button to confirm the selection; (b) *Head*—the user positions the cursor on the letter 'Y', waits for 500 ms for the popup button to appear, then (1) moves the cursor to the popup button, and (2) returns to the letter 'Y' to select it; (c) *Hand*—the user moves the hand to the letter 'K' and makes a close palm gesture to selects it; (d) *Hybrid* (Head+Hand)—the user uses the head to move the cursor to the letter 'Y' and makes a close palm gesture to make the selection.

Controller+Swype. To type a word, the user needs to move the cursor to the first letter of the intended word and then click the trigger button on the controller to start the Swype action. When the user finishes Swyping, clicking the trigger button again ends the typing process. For special characters, the user needs to move the cursor to the corresponding block and then clicks the trigger button for selection. Figure 3-2 shows an example of a Swype action.

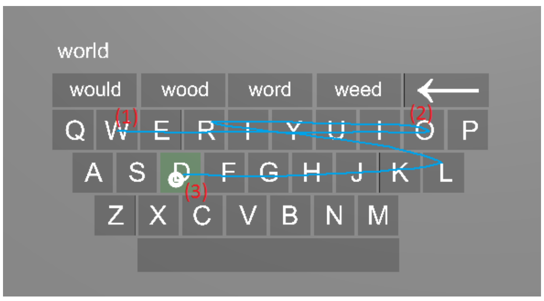


Figure 3-2. For (Controller/Hand/Head)+Swype, to type the word 'world' a user needs to follow these three steps: (1) Moving the cursor to the first letter 'W' and performing a selection action to indicate the start of the Swype process; (2) then Swyping the letters one by one; and (3) Performing another selection action on the last letter (in this case 'D') to indicate the end.

Section 3.2.2 Head

Head-based pointing (or simply Head) is analogous to the Controller, but instead of a handheld device, only the HMD is used. A ray is extended from the HMD position towards the viewing direction into the virtual environment. The ray intersects the keyboard at a point and a blue cursor is given as a prompt (see Figure 3-1b).

Head+Tap. Figure 3-1b shows an example of how a user completes a Head+Tap action. To enter a word, the user needs to move the cursor using their head to the corresponding letter. A letter selection is made via an outside-inside fashion [145] like a nod action. To begin the process, the user moves the cursor to the target letter; then a button representing an action appears above the letter after a wait time of 500 ms. The user now needs to move the cursor to the button and after moves it back to the target to perform the selection (see Figure 3-3a). The user needs to do the action for selecting each letter, suggested word, space, and backspace.

Head+Swype. Selection is like Head+Tap. To type a word, the user needs to perform the selection action on the first letter, then moves the cursor over the component letters, and finally finishes typing by doing the second selection action on the last letter (see Figure 3-3b).

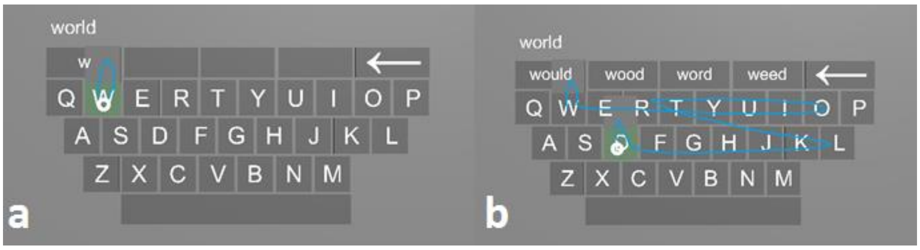


Figure 3-3. An example of typing the letter ’w’ (a) and the word ’world’ (b) in the Head approach.

Section 3.2.3 Hybrid

Head-based Pointing + Hand gesture (or simply Hybrid) is a HoloLens-like text input approach. Both implementations of Hybrid+Tap and Hybrid+Swype are analogous to the Head+Tap and Head+Swype, respectively. The only difference is that Hybrid uses a hand gesture (like a palm closing) to indicate a selection. Palm closing gesture was chosen because it can be accurately recognized by Meta 2.

Section 3.2.4 Hand

This approach enables users to interact with the virtual keyboard with their hands only. The positions of the palm and hand gestures (i.e., grabbing) are captured via the front camera of the HMD. That is, we use the palm mid-air position to indicate the cursor's position that acts as the hand-based 'pointing' (or simply Hand). Users move the cursor according to their hands around the virtual keyboard (see Figure 3-1c).

Hand+Tap. Figure 3-1c shows how a user completes a Hand+Tap. Selection is indicated by a palm closing gesture. The user selects a letter by moving the cursor using their hand to the corresponding letter and then selects it by doing a palm closing action. The user should do this to select either a letter, suggested word, or space/backspace. Either left or right hand can be used in this method.

Hand+Swype. Selection is analogous to Hand+Tap. To Swype a word, the user needs to do a first selection gesture on the initial letter of the word to indicate the start, then moves the cursor over the other letters, and finally needs to do the second selection gesture on the last letter to indicate the end of the Swype process. To select a word suggestion, delete a letter or add a space, the user needs to move the cursor to the corresponding area, and then do the selection gesture.

Section 3.2.5 Commonalities and Differences Between Swype and Tap

When entering text, it is common for the system to suggest some recommended words based on the typed letters. We have also included the use of these suggested words. Both Swype [92] and Tap (using Symspell [81]) used Damerau–Levenshtein distance algorithm and the same library [298]; as such, the word suggestion performance should not affect the text entry performance.

For Tap, because we do not know whether the user has finished entering the word, we cannot automatically add the best suggestion word into the sentence. All word suggestions appear in the selection blocks (see Figure 3-4a, on top of the keys). They are updated every time the user makes a change (i.e., adding or deleting a letter). To select a suggested word, the user needs to choose it from the corresponding selection block. Hitting the space key will append a space after the input. Backspace deletes the last input, which can be a complete word or a single letter.

For Swype, since there is a second selection action to indicate the end of entering a word, the system automatically adds the best word suggestion into the text field with four other possible words in the selection blocks (see Figure 3-4b). If the best word suggestion is the intended word, the user can confirm it by Swyping on the next word. If the best suggestion is not the intended word, the user selects the desired word from the selection blocks. The system also automatically appends a space after a word has been input. A delete action deletes the whole word that is last entered.

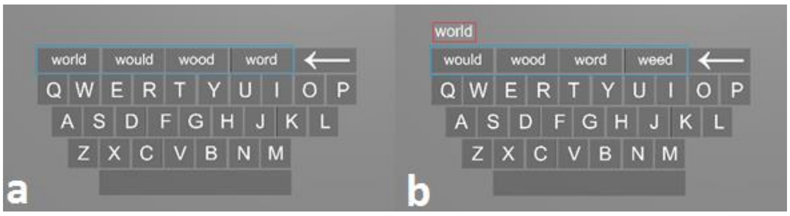


Figure 3-4. The blue areas show the word suggestions for Tap (a) and Swype (b). In addition, for Swype, the best matched word is automatically added into the input field (the red area).

Section 3.3 Empirical Study

We conducted an experiment at a university lab with the four Pointing Methods (Head, Hand, Hybrid, and Controller) and two Input Mechanisms (Swype and Tap) to assess their relative performance (speed and error rates) and user preference (workload, motion sickness, user experience, and immersion level).

Section 3.3.1 Participants and Apparatus

Twenty-four unpaid participants (eight males and four females in each of the two groups) between the ages of 18 to 28 (mean = 21) were recruited randomly from the local university campus through a database of participants. All participants were familiar with the English alphabet because the language of instruction at the university in English but there were not native alphabet users—English was not their first language. Nineteen participants had some limited experience with AR HMDs—they had either seen and/or interacted with them. They all had normal or corrected-to-normal vision and did not have any difficulties moving their arms and heads. The experiment was conducted using a Meta 2 AR HMD connected to a Windows 10 machine running Unity3D. A standard desktop computer was used; it had an i7 CPU, 16 GB RAM and a Nvidia GeForce GTX 1080Ti GPU. Figure 3-5 shows the experimental setup.

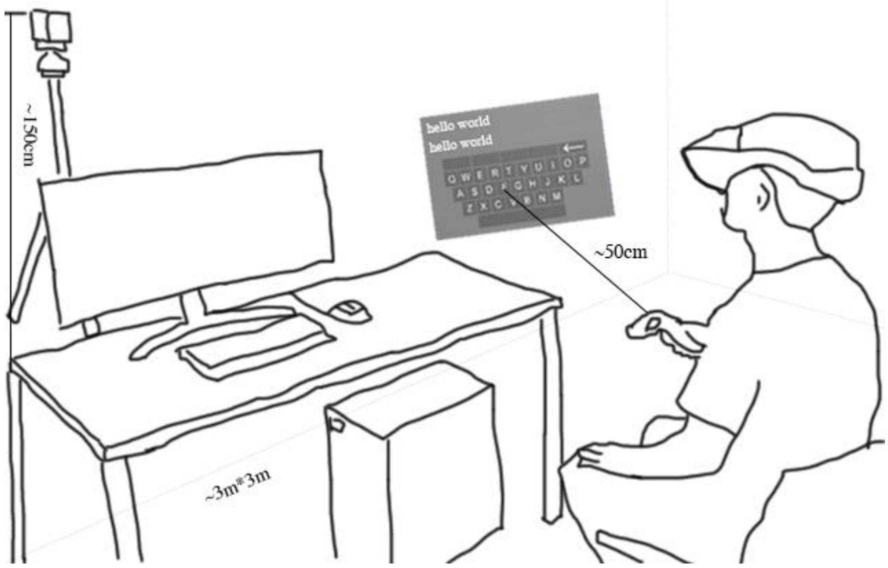


Figure 3-5. This figure shows the experimental setup. The HTC Vive optical trackers were placed at 1.5m high and had a tracking space with 3 × 3m^2. The keyboard is roughly 0.5m away from the participant which is recommended by the developers of the Meta 2 [299].

Section 3.3.2 Design

The experiment followed a mix design approach, with one between-subjects independent variable, Input Mechanisms (Swype and Tap), and one within-subjects independent variable, Pointing Methods (Head, Hand, Hybrid, and Controller). The dependent variables were performance (speed and accuracy) and users' subjective feedback (workload, motion sickness, user experience, immersion). Each Input Mechanism was tested on 12 participants (that is, 12 for Swype and 12 for Tap). For each Pointing Method, participants needed to complete eight phrases which were randomly sampled from the MacKenzie phrase set [167]. To avoid learning effects, we counterbalanced the Pointing Methods. Aside from training phrases, we collected 768 trials (24 participants × 4 Pointing Methods × 8 phrases).

Section 3.3.3 Procedure

To ensure that both groups have equal text entry ability in the actual experiment stage, participants were separated into two groups (Swype and Tap) based on their performance on a standard desktop PC from a pre-test. Before the experiment, participants were told the goal of the investigation and the conditions that were to be tested. The order of the conditions was balanced across participants. In all conditions, participants were instructed to enter the text phrases as quickly and as accurately as possible. Error correction was allowed by using the backspace key. Before each condition, the Pointing Method was explained to the participants and they practiced two warm-up phrases. After the warm-up phrases, participants needed to complete eight phrases for each condition. The conditions were separated by a 5-minute break during which participants filled out the NASA TLX questionnaire [107], Motion Sickness Assessment Questionnaire (MSAQ) [87], Slater-Usoh-Steed Questionnaire (SUS), and User Experience Questionnaire (UEQ) [149]. After the experiment, we interviewed participants and asked them to comment on the techniques. The whole experiment lasted approximately one hour for each participant.

Section 3.3.4 Results

We analyzed the data using a two-way mixed ANOVA with Pointing Methods (Controller, Head, Hand, and Hybrid) as the within-subjects variable and Input Mechanisms (Swype and Tap) as the between-subjects variable. Bonferroni correction was used for pairwise comparisons and Greenhouse-Geisser adjustment was used for degrees of freedom for violations of sphericity. Because of our sample size, the significance threshold was set at $p < .01$ in our analyses.

Text entry rate was measured in Words Per Minute (WPM), with a word defined as five consecutive letters, including the space character. For Swype, we use the following formula

Equation 3-1 Swype technique text entry speed

$$WPM = \frac{|T|}{S} \times 60 \times \frac{1}{5} \tag{1}$$

Where S was the time (in seconds) from the time when the user triggered the first start action to the last action. $|T|$ was the number of characters in the transcribed text.

For Tap, we use the following formula

Equation 3-2 Tap technique text entry speed

$$WPM = \frac{|T| - 1}{S} \times 60 \times \frac{1}{5} \tag{2}$$

Where S was the time (in seconds) from the time of the first to the last key entered, and $|T|$ was the number of characters in the transcribed text.

The error rate was calculated based on the standard typing metrics [250], where the total error rate (TER) = not corrected error rate (NCER) + corrected error rate (CER).

Text Entry Performance

Table 3-2 shows the results from the 2-way mixed ANOVA. Figure 3-6 shows the mean text entry speed among the eight techniques. In general, for Pointing Method, Controller achieved the best results for both Tap (M = 14.6, SD = 0.85) and Swype (M = 13.68, SD = 1.88) and Head had the worst performance in both Tap (M = 5.62, SD = 0.64) and Swype (M = 7.94, SD = 1.36). Figure 3-7 shows the details of the TER and NCER for all methods. Hand caused the highest error rates in TER for both Tap (M = 6.48%, SD = 1.80%) and Swype (M = 5.01%, SD = 4.70%) as well as NCER again for both Tap (M = 3.82%, SD = 2.04%) and Swype (M = 0.75%, SD = 0.92%). Head+Tap achieved the lowest TER (M = 1.06%, SD = 1.23%) and NCER (M = 0.48%, SD = 0.82%) while Controller+Swype achieved the lowest TER (M = 1.24%, SD = 1.44%) and NCER (M = 0.00%, SD = 0.00%).

To see if there was significant effect of Pointing Methods for either Tap or Swype, we employed a one-way repeated ANOVA. For Tap, the test yielded a significant effect of Pointing Methods ($F_{2.137,23.503}$ = 39.971, p < .001). Pairwise comparison revealed significant differences between Controller - Head, Controller - Hybrid, Controller - Hand (all p < .001). For Swype, the test yielded a significant effect of Pointing Methods ($F_{1.974,21.719}$ = 89.375, p < .001). Post-hoc pairwise comparison revealed significant differences between Controller - Head (p < .001), Controller - Hybrid (p < .001), Controller - Hand (p < .001), and Head - Hybrid (p < .01).

Table 3-2. Two-way mixed ANOVA test results for text entry performance. Significant results where $p < .01$ are shown in green and $p < .001$ in dark green.

	WPM	TER	NCER
Pointing Methods	$F_{2.247,49.428} = 125.890$, $p < .001$	$F_{3,66} = 15.798$, $p < .001$	$F_{3,66} = 11.760$, $p < .001$
Pointing Methods × Input Mechanisms	$F_{2.247,49.428} = 7.225$, $p < .01$	$F_{3,66} = 2.468$, $p = .083$	$F_{3,66} = 9.174$, $p < .001$
Input Mechanisms	$F_{1,22} = 5.227$, $p = .032$	$F_{1,22} = .055$, $p = .817$	$F_{1,22} = 18.623$, $p < .001$
Post-hoc Pointing Methods	Controller - Head ($p < .001$), Controller - Hybrid ($p < .001$), Controller - Hand ($p < .001$), Head - Hybrid ($p < .001$)	Controller - Head ($p < .01$), Head - Hybrid ($p < .01$), Head - Hand ($p < .01$)	N/A

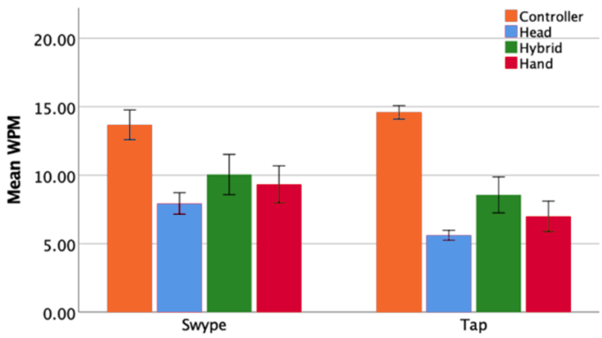


Figure 3-6. Mean WPM for each Pointing Method grouped by Swype and Tap. Error bars indicate ±2 standard errors.

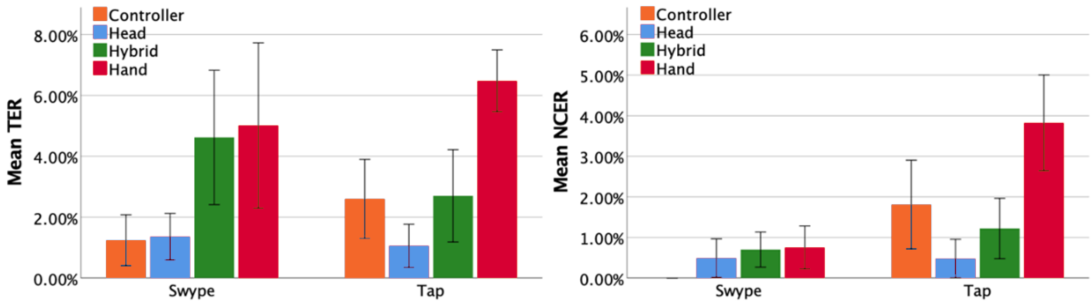


Figure 3-7. Mean TER (a; left) and NCER (b; right) in % among all methods. Error bars indicate ±2 standard errors.

User Preference

SUS. The SUS counts for Hand+Swype (M = 1.08, SD = 1.62) were the highest but the lowest for Controller+Tap (M = 0.17, SD = 0.39). Figure 3-8a shows that the mean immersion score from SUS questionnaire for Hand+Swype (M = 4.11, SD = 0.80) was the highest and Head+Tap (M = 3.25, SD = 1.14) the lowest. There was no significant difference for immersion between the Pointing Methods ($F_{3,66}$ = 3.199, p = .029), Pointing Methods × Input Mechanisms ($F_{3,66}$ = .308, p = .820), and Input Mechanisms ($F_{1,22}$ = .419, p = .524).

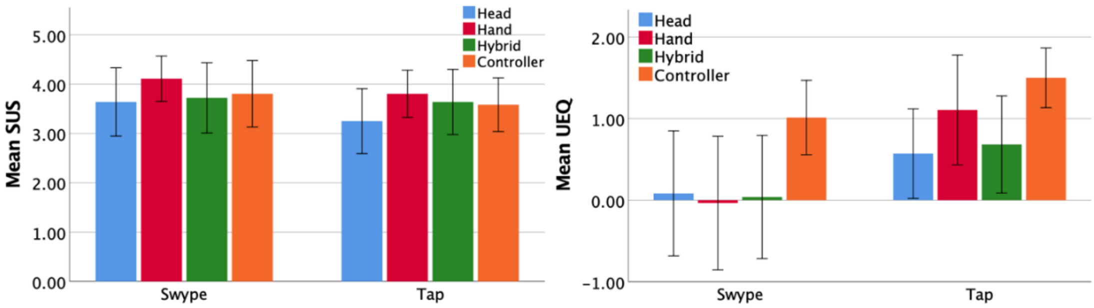


Figure 3-8. Mean immersion score from SUS questionnaire (a; left). Mean user experience score from UEQ (b; right). Error bars indicate ±2 standard errors.

UEQ. The scales for UEQ were adjusted between -3 (very bad) to 3 (excellent). For the average score, ANOVA tests showed a significant effect of Pointing Methods ($F_{3,66}$ = 9.295, p < .001), but insignificant for Pointing Methods × Input Mechanisms ($F_{3,66}$ = 1.183, p = .322). There was no significant effect of Input Mechanisms ($F_{1,22}$ = 3.306, p = .083) where the average experience score for Tap was 0.965 (SD = 1.01) and for Swype 0.275 (SD = 1.27). Post-hoc pairwise comparisons revealed significant differences between Head - Controller (p < .001) and Hybrid - Controller (p < .01). Figure 3-8b shows the details of the mean UEQ for all methods.

Regarding each UEQ subscale (see Figure 3-9), ANOVA tests yielded a significant effect of Pointing Method, Input Mechanisms, or Pointing Methods × Input Mechanisms on attractiveness, perspicuity, efficiency, and dependability. However, no significant effect was found for novelty and stimulation. Table 3-3 shows detailed results of the ANOVA tests. As can be seen from the Figure 3-10, the controller was rated above average to excellent when compared to the benchmark scores while the other three Pointing Methods were rated between bad and above average.

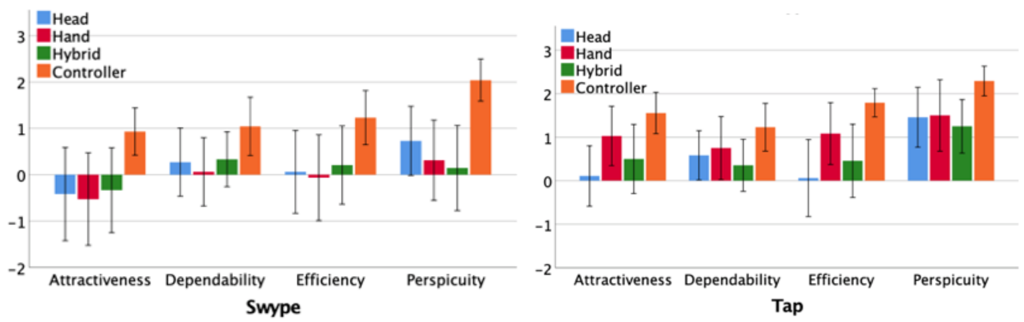


Figure 3-9. Mean UEQ subscales for each Pointing Method for Swype (a; left) and in Tap (b; right). Error bars indicate ±2 standard errors.

Table 3-3. ANOVA test results for UEQ subscales. Significant results where p < .01 are shown in green and p < .001 in dark green. Novelty, Stimulation, Input Mechanisms, Pointing Methods × Input Mechanisms have no significant result and therefore not shown for better clarity.

	Efficiency	Perspicuity	Dependability	Attractiveness
Pointing Methods	$F_{2.244,49.357} = 10.141$, $p < .001$	$F_{3,66} = 16.170$, $p < .001$	$F_{3,66} = 5.054$, $p < .01$	$F_{3,66} = 10.701$, $p < .001$
Post-hoc Pointing Methods	Head - Controller (p < .001), Hand - Controller (p < .01), Hybrid - Controller (p < .01)	Head - Controller (p < .001), Hand - Controller (p < .001), Hybrid - Controller (p < .001)	Hybrid - Controller (p < .01)	Hand - Controller (p < .001), Hybrid - Controller (p < .01)

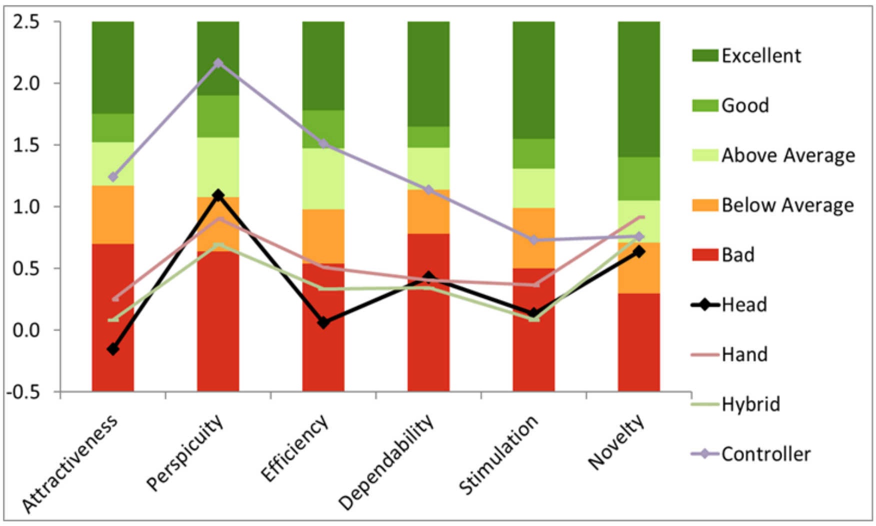


Figure 3-10. UEQ ratings of our tested Pointing Methods (Head, Hand, Hybrid, and Controller) with respect to comparison benchmarks.

Motion sickness. Regarding overall motion sickness, Controller+Tap was rated the best (M = 14.53%, SD = 4.92%) and Hybird+Swype (M = 30.09%, SD = 18.54%) the worst. ANOVA tests yielded significant differences between Pointing Methods

($F_{2.694,59.262}$ = 5.662, p < .01); however, no significant effect was found for Pointing Methods × Input Mechanisms ($F_{2.694,59.262}$ = 1.942, p = .138) and Input Mechanisms ($F_{1,22}$ = 4.435, p = .047). Pairwise comparisons did not reveal any significant differences.

Regarding the MSAQ subscales (gastrointestinal, central, peripheral, and sopite-related), there was a significant effect of Pointing Methods ($F_{3,66}$ = 4.979, p < .01) on central. However, post-hoc pairwise comparison yielded no significant difference. In terms of sopite-related motion sickness, the ANOVA test yielded significant differences between Pointing Methods ($F_{3,66}$ = 8.406, p < .001), but not between Pointing Methods × Input Mechanisms ($F_{3,66}$ = .808, p = .067). Post-hoc pairwise comparison showed a significant difference between Head - Controller and Hybrid - Controller (all p < .01). No other significant effects were found. Figure 3-11 shows MSAQ subscales scores.

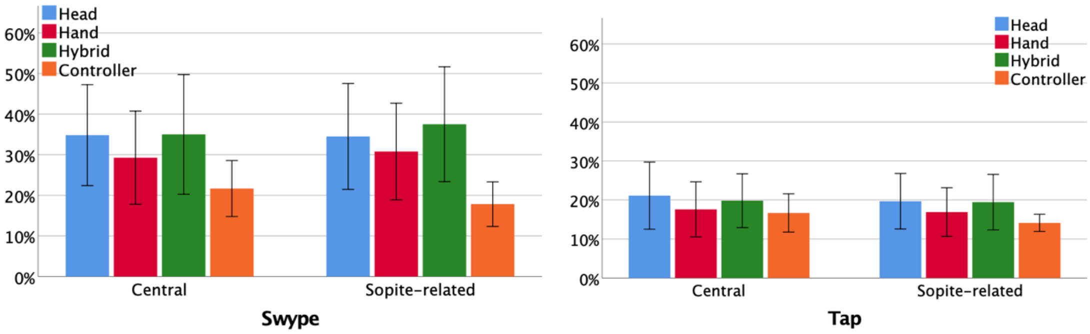


Figure 3-11. MSAQ subscales for each Pointing Method in Swype (a; left) and in Tap (b; right). Peripheral is not shown as no significant difference was found. Error bars indicate ±2 standard errors.

NASA-TLX Workload. For overall task workload, Controller+Tap was rated the best (M = 33.92, SD = 19.44) and Hybrid+Swype (M = 75.19, SD = 12.84) the worst. An ANOVA test showed significant differences for Pointing Methods ($F_{3,66}$ = 26.063, p < .001) on overall workload, but not for Pointing Methods × Input Mechanisms ($F_{3,66}$ = 3.990, p = .011) and Input Mechanisms ($F_{1,22}$ = 5.724, p = .026). Post-hoc pairwise comparisons revealed significant differences between Head - Controller, Hand - Controller, Hybrid - Controller (all p < .01). Regarding each workload subscale, ANOVA tests yielded at least one significant effect for Pointing Methods on all workload subscales except for performance. Details of results of the ANOVA tests can be seen in Table 3-4 and of the workload subscales in Figure 3-12.

Table 3-4. ANOVA test results for NASA-TLX workload subscales. Significant results where $p < .01$ are shown in green and $p < .001$ in dark green. Non-significant results are omitted for clarity.

	Pointing Methods	Input Mechanisms	Post-hoc Pointing Methods
Mental	$F_{3,66} = 4.813, p < .01$	$F_{1,22} = 3.571, p = .072$	N/A
Physical	$F_{3,66} = 22.021, p < .001$	$F_{1,22} = 5.081, p = .034$	Head - Controller, Hand - Controller, Hybrid - Controller (all $p < .001$)
Temporal	$F_{3,66} = 6.975, p < .01$	$F_{1,22} = 8.175, p < .01$	Hand-Controller ($p < .01$)
Effort	$F_{3,66} = 13.045, p < .001$	$F_{1,22} = 4.867, p < .038$	Head - Controller, Hybrid - Controller (both $p < .01$), Hand - Controller ($p < .001$)
Frustration	$F_{3,66} = 6.004, p < .01$	$F_{1,22} = 8.537, p < .01$	Hand - Controller ($p < .01$)

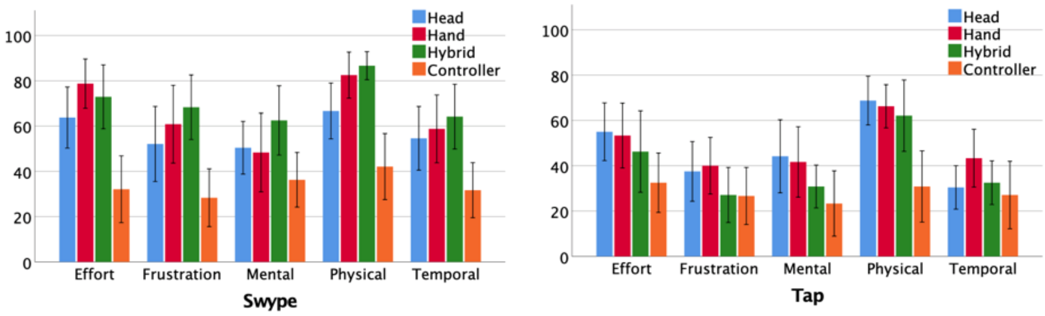


Figure 3-12. Workload subscales for each Pointing Method using Swype (a; left) and Tap (b; right). Performance is non-significant and not shown for better clarity. Error bars indicate ±2 standard errors.

Section 3.4 Discussion

In this section, we first discuss task performance of the combination of each Pointing Method (Head, Hand, Hybrid, and Controller) and Input Mechanism (Swype and Tap), and then user subjective feedback for each combination.

Section 3.4.1 Task Performance

Controller+Tap achieved an average of 14.6 WPM which is comparable to results in other pointing method + tap approaches from research in VR and non-VR domains [175,244,251]. Controller+Swype achieved an average of 13.68 WPM which is also comparable to some results in VR (e.g., in [294], their participants were able to achieve 15.75 WPM). Our results also indicate that Controller outperformed all the other device-free methods. However, when compared to a physical keyboard, which has been shown to be able to support fast text entry of around 45 to 67 WPM [138], Controller-based input seems still not fast enough for heavy text entry activities. It may not be necessarily an issue with any pointing method in particular but that AR in general may not support long periods of intensive text entry. For short text entry tasks like sending short messages via social media chat applications, a technique based on Controller+Swype could work well.

Head+Tap has led to an average of 5.62 WPM using the outside-inside approach but this is only the half of input speed of Head Pointing using a button to make selections (about 10 WPM) [251,294] and is also slower than the dwell-based Head Pointing technique (around 8 WPM) [294]. Unlike Hand and Hybrid, both of which can use hand gestures for letter/word selection, there are currently no optimal methods, except dwell, for purely Head Pointing-based approaches for text entry with a QWERTY keyboard layout. If users have to use Head Pointing-based approaches, an alternative approach that exists in the literature is to use a circular layout like a technique called RingText [287] which has been shown to be faster than dwell QWERTY.

We have observed that text entry performance (both speed and accuracy) for Head and Hybrid are affected by the hardware (e.g., tracking cameras and feasible tracked area), software (e.g., gesture detection algorithms), and users' physical capabilities and predispositions (e.g., how long and how stable they can keep their hand in mid-air). In the context of AR, the area that is tracked by the cameras tend to be limited and because of this the users must lift their hands in the mid-air, which may further cause hand tremor and arm fatigue quickly, making it challenging for many users. The detection algorithm provided by the Meta SDK seems to have issues. We have discovered that, when the users move their hand out of the tracking area accidentally or intentionally, the algorithm sometimes thinks that their hands are performing a palm closing gesture—i.e., a false positive recognition, and assumes a selection is made while in fact, the users are not doing anything. Because of this, during the experiment, we had to remind users to keep their hand within the tracking area.

As for Input Mechanisms, the experimental results suggest that for users who are new to Swype and Tap, the Swype technique have the same text entry speed as Tap and cause lower NCER than Tap. If users prefer lower errors in the transcribed text, they should use Swype instead of Tap.

Section 3.4.2 User Preference

In the following discussion, we discuss each Pointing Method and Input Mechanism based on the subjective feedback and our observations from the experiment.

Workload

Controller outperformed all the other methods for Physical and Effort workload and exceeded Hand for Temporal and Frustration. As such, a Controller-type of input seems to be a good first option if a lower workload is important for users. Our observations also show that our participants complained that Hand and Hybrid were too tiring because of the need to keep their hands in mid-air in a consistent and stable basis. Due to the limitations of the Meta 2 headset's tracking area, the users cannot place their hands in a more relaxing pose. It is worth pointing that this issue is not just confined to the Meta 2 but it is a widely report issues for AR devices. Although Head did not have this problem, participants complained about minor neck pain and fatigue. One solution could be to use a device with an eye-tracking device installed (i.e., gaze input [191]), if the cost is not an issue and the eye tracker can provide accurate and stable performance. Thus, when a controller is not around, users could consider a Head approach when hand fatigue is a big concern. They should consider a Hand approach when arm fatigue is less of an issue.

Swype techniques resulted in a significantly higher temporal and frustration workload than Tap. Surprisingly, Swype and Tap have the same level of mental workload even though Swype requires users to remember and type all letters in one continuous Swype action to complete the words. It is worth noting that although our participants were not native alphabet users, they were still able to mentally keep track of the words that they needed to type using Swype with relative proficiency, but this had come with higher frustration and temporal workload, which may not be the case with English native speakers. In general, if the workload is a critical factor of the text entry technique, a Swype-style text approach should not be considered due to its high workload demand in both temporal and frustration workload.

Motion Sickness

Results indicated no differences for the overall sickness among the tested techniques. For each subscale from motion sickness assessment questionnaire [87], the Controller approach was found to be less annoying, drowsing and tiring than Head and Hybrid techniques because it did not need our participants to use head rotations. This means that a ray-casting enabled controller should be preferred if available. Additionally, users should consider a Hand approach when the controller is not around.

For Input Mechanisms, our results indicate that Tap causes the same level of motion sickness as Swype. The selection of which Input Mechanism to apply should consider other aspects (e.g., workload) as they both have no effect on motion sickness.

Immersion

There were no significant differences between the difference combinations of Pointing Methods and Input Mechanisms for immersion, which indicates that text entry in AR has no significant impact on immersion. Overall, users should consider other factors (e.g., workload) to decide which technique to use.

User Experience

For the user experience subscales, Controller provided a significantly better user experience in efficiency and perspicuity than the other methods. It also gave better dependability than Hybrid and received higher ratings in attractiveness than Hybrid and Hand. When we compare these pointing approaches with the benchmark scores [235], only Controller is found to have received an above average to excellent rating while Head, Hand, and Hybrid are rated bad to below average. For the Input Mechanisms, we found that Tap and Swype have no significant difference on user experience.

In summary, the Controller offers the best user experience and as such, if a ray-casting enabled controller is available, it should be used as a first choice. Otherwise, users should consider other user experience measurements such as workload to decide which alternative Pointing Methods to use.

Section 3.4.3 Recommendations for Text Entry in AR HMDs

The recommendations derived from our experiment can be divided into two groups based on their goals:

Performance. Based on the results, we suggest that users should use a ray-casting enabled handheld device since it can lead to a good text entry performance and it is capable of other tasks, like manipulating virtual object [294,295]. Device-free methods should be considered in addition to speech recognition, if available, when device-free is the only option. On the other hand, if the environment is noisy and users are in a public space, which can potentially bring privacy concerns [280], we suggest using one of the device-free approaches based on user experience. Of the two Input Mechanisms, Swype should be considered first since it has a higher text entry rate and a lower not corrected error rate than Tap.

Experience. We suggest that a handheld device should be the preferred option because it has low workload and motion sickness but provides a better user experience. However, if no such devices are around, the following can be considered. If users have difficulty holding their hands constantly and consistently in mid-air, Head-based pointing can be considered as an alternative. Hybrid can be used if arm and neck fatigue is not a concern and there is enough space for users to lift and hold the arms mid-air. If users' neck fatigue is a concern and users have ample space for hand interaction, the Hand approach could be chosen instead. This is also because a natural hand interaction allows users to perform tasks in both the real and virtual environment at the same time [12]. Of the two Input Mechanisms, Tap should be chosen since it generates lower workload (for both temporal and frustration).

Section 3.4.4 Limitations and Future Work

This research has some limitations. The experiment was tested with a Meta 2 AR HMD. We chose it because it had one of widest field-of-view and, like other AR devices, it has some issues in tracking hand motions and gestures. We used three countermeasures to minimize issues that this could have caused: (1) We chose one of the simplest gestures (closing palm) which the Meta 2 provided and of which it had a reliable tracking performance; (2) To avoid potential environmental noise factors that may affect tracking performance, we did tests to ensure the environment would not cause any tracking issues; and (3) We allowed users to familiarize themselves with the device and techniques via warm-up practices. Given this, the AR device chosen in our study is still suitable for our purposes and the results we obtained are still quite relevant to AR systems. In the future, when AR devices have improved tracking performance, it will be useful to explore other combinations of pointing and selection methods for entering the text that is accurate and fast.

We observed that with the number of phrases that our participants had to type, some of them felt that their hand and arm got tired, especially for the Hand and Hybrid approaches. Future research can explore possible ways to minimize arm/hand fatigue for these two types of approaches. Similarly, our experiment involved 12 participants in each group (24 in total), which according to Caine [36] is one of the most common sample sizes within HCI research. Given our sample size, we used the alpha value of 0.01 to ensure that any replication could likely achieve similar results [1]. In the future, it will be useful to evaluate if performance and user experience can improve with larger sample size and longer experimental sessions, for example, 1-2 sessions over consecutive 4-5 days like PizzaText [295], RingText [287] in VR scenarios and WrisText [93] in smartwatch scenarios.

Additionally, our evaluation experiment was conducted in a lab environment where the background is somewhat, but not fully, controlled to be clean and easy for the front camera to track the hand motions and gestures. Future work can consider experimenting with more realistic environments, e.g., in a park or a shopping mall with people walking in front of the camera. This future research can be informed by the results of this current experiment.

Finally, as mentioned in the discussion section, the selected Pointing Methods and AR devices in general may not be suitable for long text entry sessions and heavy text editing of documents. Although AR devices are usually meant for short text entry sessions (like for sending short messages), it is worthwhile to explore and develop new techniques that will support text entry activities that are more involved and last longer. For instance, easily and widely accessible devices like smartphones, which have been reported to support users to type 50 WPM when they are sitting [50] and about 30

WPM when they are walking [90], can be part of this exploration. Also, voice input techniques, such as SilentVoice [77] which can mitigate privacy issues and work well in noisy environments, are also valuable and can be useful for some text entry activities. Further research is needed because both smartphones or SilentVoice have their inherent technical and usability issues and, if we are to develop new techniques that linked them to an AR system, these issues need to be overcome.

Section 3.5 Conclusion

In this work, we empirically and systematically investigated the combination of four pointing methods (Head, Hand, Hybrid, and Controller) and two input/selection mechanisms (Swype and Tap) that can be used for text entry in augmented reality (AR) head-mounted displays (HMDs). We run an experiment with eight techniques that resulted from their combinations to assess their relative performance and user preference. In general, the results show that the best pointing method is a ray-casting enabled handheld device, but its use is dependent on specific criteria and limitations (e.g., ray-casting enabled controller is not always available for AR systems, or users cannot hold it in a stable basis). Future AR systems may be commonly used for both indoor and outdoor scenario, but a ray-casting enabled controller may not be ideal for outdoor situations. Therefore, a device-free efficient text entry method is still a more practical and cost-efficient solution because it only requires the HMD to be able to track a user's hand or head motions. On the other hand, user preference such as workload and user experience must be considered also. Between the two selection mechanisms that we explored, Swype and Tap, our results show that Swype is as fast as Tap for users who are new to Swype. But Swype brings increased workload (i.e., temporal and frustration). For lighter workload during text entry activities, users can use Tap. Our research is a first to explore the combination of most common pointing methods and selection mechanisms and can provide strong foundations for future research in text entry for augmented reality systems.

Section 3.6 Summary

In short, current motion-based interactions were worse than the controller-based interaction. In addition, through this exploratory study, we confirm the existence of the following issues: (1) hand/arm and lack of boundary awareness are crucial problems for hand motions, (2) dwell could lead to bad performance for head-based interaction. These issues should be addressed in order to improve the motion-based interaction performance and experience for HMDs. Based on the findings from the literature review as well as the exploratory study, we summarize the research directions of this thesis in the next chapter.

Chapter 4 Core Challenges and Research Questions in the Design Space of Motion-based Interaction for HMDs

In this thesis, the term motion-based interaction (MbI) is defined as "interaction that relies on the changes in acceleration, orientation, the velocity of the user's body part(s), where there is no need for direct contact with a pre-defined button or interactive surface." Based on the literature review (see Chapter 2) and the exploratory study (see Chapter 3), we have first identified a list of challenges in the design space of motion-based interaction in a big picture. Then, we have selected the Core Challenges that need to be addressed in this thesis from the list and explained why other challenges are not selected. At last, for each Core Challenge, we have proposed the corresponding Research Question, as well as the user study that was used to address the Core Challenge. Moreover, we also explained the results that we got from the study (see Figure 4-1).

Section 4.1 Challenges of Motion-based Interaction for HMDs

From the previous two chapters, we have explored several challenges in the design space of MbI for HMDs. The main challenges that we aim to address in this thesis are: (1) boundary awareness for hand-based interaction, (2) efficient hands-free head-based interface for HMDs, (3) efficient and feasible full-body interaction for general tasks with HMDs, and (4) accessible full-body interaction for applications in HMDs.

Other challenges are: (1) motion-induced sweating inside the HMD, (2) mid-air hand/arm fatigue, (3) input latency, (4) tracking dropouts, (5) limited 3D workspace, (6) limited gesture vocabulary, (7) hand-based interaction for distant object selection, (8) selecting the ideal gesture set for HMDs, (9) designing foot-based interaction for HMDs, etc. However, these challenges are not covered in this thesis due to the following reasons: (1) beyond the scope of the thesis (e.g., sweating in HMDs): motions could make users start sweating in HMDs, which can only be solved by attaching external fans to HMDs; (2) supported by guidelines (e.g., gesture vocabulary): for the issue of limited gesture vocabulary (e.g., foot-based interaction [265]), guidelines suggest that the more gestures provided to the users, the more difficulties they may face in using the system (e.g., remembering/recalling the gestures,

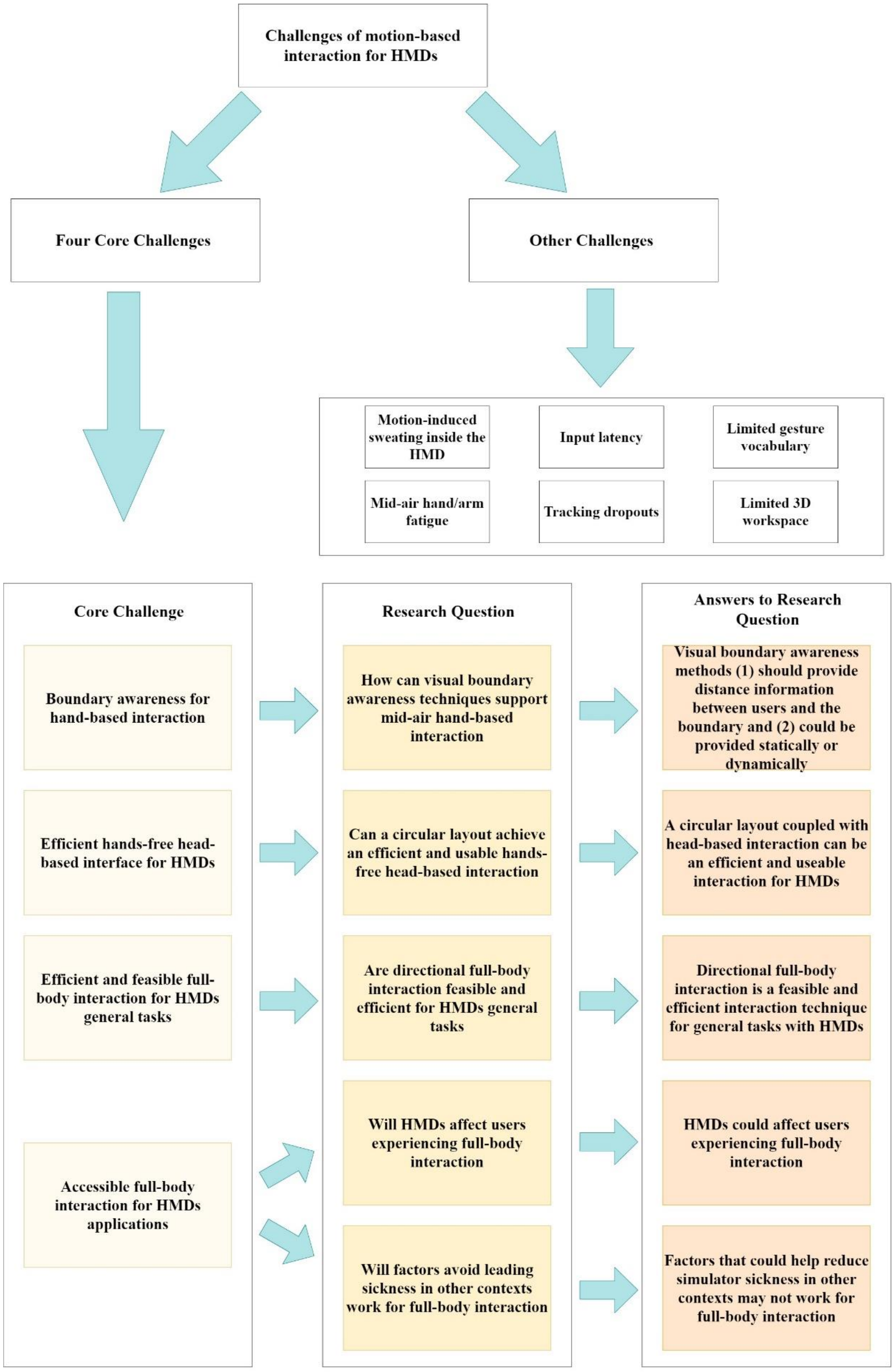


Figure 4-1. Challenges, Research Questions, and Answers to Research Questions in Motion-based Interaction for HMDs.

gesture recognition) [32], and (3) have been studied already (e.g., the ideal gesture set for HMDs): the challenge of proposing user accepted gestures has been explored in several investigations via user elicitation study (e.g., selection and manipulation tasks [199]).

Section 4.2 **Challenge 1: Boundary Awareness for Hand-based Interaction**

As mentioned in Chapter 2, boundary awareness has been observed in early works with motion tracking devices such as Leap Motion [10, 27] and Kinect [8] that unavoidably had to have a restricted tracked area due to technical limitations. Our exploratory experiment has confirmed that boundary awareness is also a problem for mid-air hand-based interaction for HMDs (see Chapter 3). A typical situation in HMDs is that users tend to gradually move their hand from the chest level to a lower level due to the tiredness of the interaction, and eventually go out of or leave the tracking volume (or area) that the devices' sensor(s) can capture, but the users may not have a conscious awareness that their hands are no longer tracked.

A possible solution to deal with this challenge is to apply visual feedback to the HMDs. Visual systems have been used for boundary awareness of the user leaving the play area in many HMDs. For instance, Oculus's *Guardian*[2] system displays an in-application translucent mesh grid when users get near play-area borders they defined—i.e., when the user gets too close to the edge of a boundary. Like the *Guardian* system, HTC VIVE's *Chaperone*[3] also displays visual grids to indicate the boundaries for the users.

However, how to design a visual boundary awareness system for mid-air hand-based interaction remains unknown. As such, **Research Question 1** asks: How can visual boundary awareness techniques support mid-air hand-based interaction?

To address Core Challenge 1 and answer Research Question 1, we have conducted a user study with an object translation task. Object translation task is selected because

[2] https://developer.oculus.com/documentation/native/pc/dg-guardian-system/
[3] https://support.steampowered.com/kb_article.php?ref=6281-TOKV-4722

it is the most fundamental interaction for HMD based environments [31,32]. In addition, the object translation task requires users to keep their hands in mid-air to select objects and move them to various target locations with many repetitions. These repeated mid-air motions could lead to boundary awareness issues for hand-based interaction, which could help us understand the usefulness of our proposed visual techniques.

Chapter 5 presents how we explore Core Challenge 1 and Research Question 1. We first examine the challenges that the user might face when interacting with HMDs without boundary information. Then, we propose two preliminary solutions to visualize the interaction boundary of HMDs and provide them statically and dynamically and evaluate them against the benchmark in the object translation task. Overall, our results suggest that visual boundary awareness methods could positively affect the user's subjective feelings during hand-based interaction. Therefore, our answer to Research Question 1 is that visual boundary awareness methods should provide information on the distance between users and the boundary. In addition, the boundary information can be provided both statically and dynamically.

Section 4.3 **Challenge 2: Efficient Hands-free Head-based Interface for HMDs**

Dwell-based interaction is the most used device-free and hands-free technique for head-based interaction. However, existing work outlined in Chapter 2 showed that the dwell-based technique has certain limitations, such as: (1) a long dwell time may decrease performance, but a short dwell time can cause false-positive selections and errors [127], (2) a pre-set dwell time always "pushed" users to select a target key and quickly move to the next one, a process that can be stressful and error prone [142], and (3) keeping the pointer static for a while to avoid selecting unintentional keys could further lead to eye and neck fatigue [229]. In this context, it is essential to explore hands-free head-based interaction for HMDs that is not dwell-based.

A feasible solution for addressing this challenge is through the use of alternative layout approaches. In our exploration, as an example, we investigated a circular (radial) interface. Circular interfaces have been used for system control tasks (e.g., menu selection [54,226,227]) and daily tasks like text entry [174]. It could be used together with head-based input with an inner and outer circle design [174]: (1) the outer circle hosts items and users reach the target item through a go-and-hit fashion, avoiding the use of dwell-technique, (2) the inner circle does not contain any items so it can be used as a relaxing region, and hence avoiding eye and neck fatigue.

However, the efficiency and usefulness of this design remains unexplored. As such, **Research Question 2** asks: Can a circular layout achieve an efficient and usable hands-free head-based interaction?

We have conducted a user study with the text entry task to **address Core Challenge 2 and answer Research Question 2**. We selected the text entry as the interaction task, not only because of the importance of text entry task for HMDs (see Section 2.3) but also because the efficiency and usefulness of the proposed method can be evaluated with the standard typing metrics [250] and be compared to other standard methods (e.g., dwell QWERTY).

Chapter 6 presents how we explore Core Challenge 2 and answer Research Question 2. First, we explore the feasibility of applying a circular keyboard layout with two concentric areas for text entry that is both dwell-free and hands-free for HMDs. Then, we have compared the text entry performance of our technique, RingText, with four other possible pure head-based methods—dwell circular, swype circular, dwell QWERTY, swype QWERTY. The results show that RingText outperforms them; it has led users to achieve a significantly higher text entry rate and close to a significantly lower total error rate. To further explore its performance, we have conducted a 4-day study with two daily sessions and 10 participants to evaluate the learning effects of RingText on speed and error rates. The results show that after eight practice sessions even novice users can achieve an average text entry speed of 11.30 WPM while expert users can achieve 13.24 WPM in the last session. These results suggest that a circular layout could achieve an efficient and useable hands-free head-based interaction. According to the above findings, the answer to Research

Question 2 is that a circular layout coupled with head-based interaction can be an efficient and useable interaction for HMDs.

Section 4.4 **Challenge 3: Efficient and Feasible Full-body Interaction for General Tasks with HMDs**

As mentioned in Chapter 2, full-body interaction has so far been studied in very limited ways in HMDs due to feasibility issues (e.g., motion sickness). Hence, the third challenge and fourth challenges focus on designing feasible full-body interaction for HMDs. Specifically, the third challenge aims to enable full-body interaction in the task domain (e.g., 3D manipulation, system control, navigation [32]). This type of interaction could avoid the pitfalls of hand-based interaction (i.e., arm/hand fatigue). However, the feasibility (e.g., motion sickness during the HMD use) and efficiency (i.e., speed and accuracy) of full-body interaction for HMDs remain underexplored.

A feasible solution to address the third challenge is to combine directional full-body motions with a compass radial interface, as it allows the same distance to items position around the user's body. Directional full-body motion-based interfaces (four cardinal directions and four intercardinal directions) have been used for dancing-like exergame [213]. These eight directional motions can be used to complete system control tasks (e.g., menu selection) when mapped with a compass radial style interface [297].

However, feasibility and efficiency of this design for HMDs remains unknown. As such, **Research Question 3** asks: Are directional full-body interaction feasible and efficient for HMDs?

To address Core Challenge 3 and answer Research Question 3, we have conducted a user study with the menu selection task and compared the proposed method with commonly used commercial methods (i.e., hand-based interaction and hybrid—head+hand interaction). The menu selection task is chosen because it is a universal task in 3D applications [32]. In addition, HMD users are often required to interact with one or more menus: from basic operations of application selection to video games

[45,65]. Further, the feasibility and efficiency of the proposed method can be evaluated with the standard metrics (task completion time and error rate [32]) and be compared to other standard motion-based interactions (e.g., hand-based interaction).

Chapter 7 presents how we explore Core Challenge 3 and answer Research Question 3. To prove that this type of interface and interaction works, we present DMove, a directional full-body interaction for HMDs that is both hands- and device-free. It uses directional walking to interact with virtual objects. To use DMove, a user needs to perform directional motions such as moving one foot forward or backward. We first investigate the recognition accuracy of our method and the social acceptance of this type of interaction, together with users' comfort ratings for each direction. We have found that (1) the proposed recognition method is very accurate—100% for 8-block DMove and 98.06% accuracy for 16-block DMove; (2) users prefer to use DMove in front of familiar people and indoor scenarios (like their home or office); (3) users felt more discomfort when moving towards directions that they cannot see.

Then, we optimize its design and conduct a second study to compare DMove in task performance and user preferences (workload, motion sickness, user experience), with two approaches—hand-based interaction and hybrid-based (head+hand) interaction for menu selection tasks. Our results show that (1) DMove has an equal task completion time as Hand and Hybrid and a lower error than Hand when using a current consumer HMD and (2) DMove is preferred by users because it has a low workload score but high usability and novelty scores. Based on the above findings, our answer for Research Question 3 is that directional full-body interaction is a feasible and efficient interaction technique for general tasks with HMDs.

Section 4.5 **Challenge 4: Accessible Full-body Interaction for Applications in HMDs**

Literature suggests that users' experience could be significantly different in HMDs and traditional 2D displays (e.g., users frequently suffer motion sickness when interacting with HMDs but not from traditional 2D displays; see Chapter 2). However, there was limited research on studying the effect of display type on full-body interaction.

Therefore, **Research Question 4** asks: Will HMDs affect users experiencing full-body interaction?

Literature also suggests that simulator sickness could be higher in concurrent multi-tasking applications [293] and using the first-person viewing perspective [187]. However, these factors are studied in a flight simulator or a VR racing game. The findings may not be applicable to full-body interactions. As such, **Research Question 5** asks: Will sickness mitigation factors in other contexts work for full-body interaction?

To address Core Challenge 4 and answer Research Question 4, we have conducted two user studies with the exergame as our task. Exergame, a combination of “motion-based exercise” and “gaming”, is used because (1) it is the most commonly used application of full-body interaction both for industry (e.g., Ring Fit Adventure[4], Kinect Sports[5], Dance Central[6]) and academic research [30,41,85], (2) it is a suitable platform to address the accessibility issue of HMDs (i.e., motion sickness) since exergame would require variety full-body motions during the game, which could increase the risk of sickness, especially for HMDs, and (3) it provides a more significant interest within the Human-Computer Interaction community since exergames represent a promising approach for various population groups (i.e., children [3], young individuals [4], and older adults [5]) to promote regular exercise in unmotivated or inactive target groups [6,7].

We address Core Challenge 4 and answer Research Question 4 and 5 by evaluating the effect of task mode (Chapter 8) and viewing perspective (Chapter 9) on full-body interaction in HMDs. In addition, compare the performance and experience of full-body interaction in HMDs and the benchmark (i.e., large display—50-inch 4K TV). Overall, our results suggest that HMDs could result in changes in physiological feelings (Chapter 8) and lead to a better game experience but also a higher sickness (Chapter 9). Hence, our answer for Research Question 4 is that HMDs

[4] https://ringfitadventure.nintendo.com/
[5] https://marketplace.xbox.com/en-GB/Product/Kinect-Sports-Season-Two/66acd000-77fe-1000-9115-d8024d5309d6
[6] https://www.dancecentral.com/

could affect users experiencing full-body interaction. Regarding Research Question 5, it seems that factors that could help reduce simulator sickness in other contexts may not work for full-body interaction.

Section 4.6 Summary

In summary, this chapter lists four Core Challenges and five Research Questions that are addressed in the rest of the thesis. In addition, it also explains the reasons why other issues are not covered.

The next chapter aims to answer the Research Question 1 (i.e., how can visual boundary awareness techniques support mid-air hand-based interaction?) and address the Core Challenge 1 (i.e., boundary awareness for hand-based interaction). It first conducts a formative study to gather the information that users needed for boundary awareness, and then it describes the development of visual boundary awareness techniques for mid-air hand-based interaction. Finally, it presents an experiment comparing the proposed visual boundary awareness techniques against the benchmark (where no boundary information is provided) with respect to object translation tasks regarding their performance and experience.

Chapter 5 Visual Methods for Boundary Awareness for HMDs

Section 5.1 Introduction

Hand-based interaction is one of the most commonly used interaction methods in head-mounted displays (HMDs) [171] (e.g., Meta 2, HTC VIVE, Oculus Quest), because it is assumed to be natural, practical, and easy to use. The proliferation of reasonably-priced depth cameras and sensors has warranted the investigation of natural user interfaces that are often based on mid-air hand interactions [114]. Currently, most AR HMDs have enabled mid-air hand interaction, but the supported tracked interaction volume is relatively small and limited. Due to this small tracked area, users often observe that the virtual object may not be responding to their gestures during regular interaction (see Figure 5-1a for a typical scenario). Such a situation could lead to unnatural and inaccurate interaction experience in different broad interaction scenarios (e.g., AR remote collaboration [83,270]) in specific tasks (e.g., hand-based text entry in AR HMDs [158]). One way to avoid or mitigate this issue is by allowing users to see via explicit visual cues the tracked interaction area (see Figure 5-1b for an example of such method). By knowing the boundary, it might enhance the performance of hand-based text entry technique in HMDs, avoid wasting time in remote collaboration [83], enhance the remote learning experience in training like telemedicine [270].

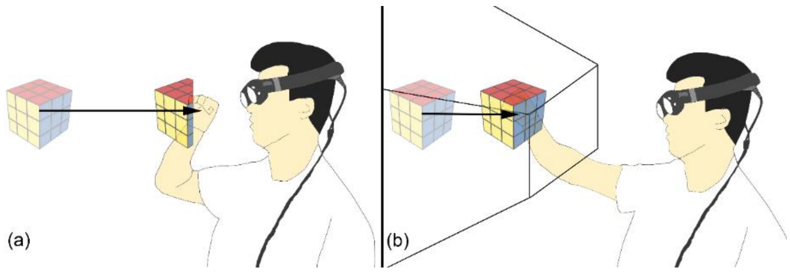


Figure 5-1. (a; left) (1) The user is trying to drag the object closer to himself. (2) The object partially disappears once the user's hand is outside the HMD's interaction boundary. Because the user cannot see the interaction boundary, it can lead to confusion and errors. (b; right) By showing the interaction boundary, the user can interact with the virtual object in HMDs and know when to stop his movement to avoid going outside of the tracked interaction area.

This research begins with a formative study to examine the challenges that the user might face when interacting with HMDs (i.e., AR) without boundary information. Then, based on the participants' comments, our observations, and interviews, we propose two preliminary solutions to visualize the interaction boundary of HMDs. One solution is using an off-body indicator, which mimics the proprietary guardian boundary system that is used by Oculus Rift HMDs. This visual technique displays a transparent colored surface to remind the user of the boundary. The other solution is using an on-body indicator, which mimics the hand state notification system that is utilized by Meta 2 AR HMDs. This technique displays a coordinate system on the users' hands to remind the user of the interaction boundary. To understand how to provide such boundary awareness methods and their usefulness, we explore our preliminary solutions to be provided statically—i.e., the system always displays the boundary awareness information, and dynamically—i.e., it displays the information only when it is necessary.

Our investigation of boundary awareness methods in HMDs began with one of the most common and essential mid-air interaction tasks—object translation [32]. To understand what the best way is for showing boundary awareness in HMDs, we conducted a controlled experiment to assess the accuracy and efficiency of boundary indicators for mid-air hand interactions with HMDs. More specifically, we investigated the following two research questions.

RQ1: How accurately and efficiently can users interact with the system in dynamic tasks (e.g., translating virtual objects) when they cannot see the interaction area?

RQ2: How do boundary awareness methods affect the user's subjective feelings of translating virtual objects in HMDs?

The contributions of the chapter include: (1) the first systematic exploration of visual methods for boundary awareness in HMDs, and (2) results of a user study comparing different visual boundary awareness methods for interacting with virtual objects in these systems.

Section 5.2 Related Work

Our current work builds on prior research on mid-air interaction and its related issues such as hand tracking, gesture recognition, and users' awareness of the boundary of tracked areas.

Section 5.2.1 Mid-air Interaction

Koutsabasis and Vogiatzidakis [140] indicated that mid-air interaction has the following characteristics: (a) touchless interaction, (b) real-time sensor tracking of (parts of) the user body, (c) body movements, postures and gestures that need to be identified and matched to particular user intentions, goals, and commands. In the following subsections, we describe motion tracking devices/sensors, gesture recognition techniques, and the use of mid-air hand interaction in AR HMDs.

Low-cost Motion Tracking

In 2006, one of the earliest commercial mass production motion tracking products released was the Wiimote controller by Nintendo, which uses an accelerometer and optical sensors to track the user's hand movements. Later, the Sixth Sense [184] presented the first affordable, wearable mid-air gestural interface that enables on-demand augmentation of the physical world with digital information, which can be manipulated via hand gestures. Since then, more and more affordable 3D depth cameras such as Kinect, Leap Motion, and Intel's RealSense have been created to support users' interaction in games or other interactive systems using their bodies to leverage the naturalness of hand and body movements for interaction [296].

Gesture Recognition

Gesture-based interaction alongside other natural methods such as speech improves the efficiency and accuracy of the interactions, and reduces the training time and error rates [33,111,162]. Most prior studies on gesture-based recognition are based on the use of one or more RGB cameras [25]. For instance, Dani et al. [53] have proposed a low-cost approach that uses only one monocular RGB camera to enable hand pointing gesture detection and fingertip localization for mobile VR devices. Similarly, Jain et al. [128] presented a low-cost framework that works with just one RGB camera to manipulate objects in mid-air. Kinect, a device that contains an RGB camera and a depth camera, has been widely used for gesture recognition studies. Researchers [220,221] have developed a novel distance metric, the Finger-Earth Mover's Distance (FEMD), to recognize gestures represented from zero to nine and using other arithmetic symbols with the data provided by a Kinect. Inspired by FEMD, Wang et al. [269] proposed a novel superpixel earth mover's distance metric for hand gesture recognition. Reyes et al. [165] presented a novel feature weighting approach within the Dynamic Time Warping framework for gesture recognition using depth video data. Combining RGB image and depth image to recognize gestures not only improves the accuracy of the gesture recognition but also allows one hand to overlap with the face or the other hand [21]. In short, with the recent advances of low-cost depth cameras and RGB cameras, many algorithms and techniques (see [42] for a recent review) have been developed to enable gesture recognition for mid-air interaction.

Mid-air Interaction in AR HMDs

There are three main types of interaction approaches for AR HMDs—controller-based, hand-based, hybrid-based (i.e., head pointing and hand gestures) [286]. However, only hand-based input is the most commonly used interaction method for wearable AR HMDs (e.g., HoloLens, Meta 2, Project North Star, and Magic Leap 1) since it is considered intuitive, natural, and cost-effective [33]. In commercial AR HMDs (like Magic Leap 1), users need to perform the following actions to select an object that is close to them. They need first to hover the hand over the virtual object and then perform a grab gesture to select the object [286].

Section 5.2.2 Boundary Awareness

Issues

According to Bowman et al. [33], current natural interactions (like mid-air hand interaction) provide little additional productivity but make the task more complicated and unnecessarily cumbersome. The main limitations of mid-air interaction in AR HMDs include limited precision with direct input on intangible surfaces [257], arm fatigue [114], and unnatural way of selecting a distant object [33].

In this work, we focus on one limitation of mid-air hand interaction that we refer to as boundary awareness (or lack of it), which is an issue that can occur in motion tracking applications that rely on any type of sensor. For instance, for mid-air interaction, in particular, the user’s hand can easily go out of or leave the tracking volume (or area) that the devices’ sensor(s) can capture, but the user may not have a conscious awareness that their hands are no longer tracked [183] (see Figure 5-1a above). This has been observed in early works with motion tracking devices such as Leap Motion

[52,183] and Kinect [48] that unavoidably had a restricted tracked area due to technical limitations.

For AR systems, lack of boundary awareness could confuse, frustrate and discourage users towards the system because misinterpreted gestures would likely lead to unintentional actions and unresponsiveness for gestures that fall outside of the range and might lead to the users believe that the system recognition is flawed and unusable, thereby leading to an unpleasant experience. For instance, it might affect the text entry accuracy and performance of hand-based text entry techniques (i.e., a text entry technique that involves hand gestures) [158]. It might unnecessarily waste collaboration time due to loss of the hand tracking (due to fewer trackable features in the field-of-view). [83] reported this in a mock-up Boeing 737 cockpit when using a handheld AR to perform a remote collaboration with tasks like placing annotations, drawing, and live imagery (e.g., of hand gestures). Lack of boundary awareness might also affect other remote collaboration training situations (e.g., remote procedural training of telemedicine [270]).

In short, these above issues become a major problem for interactions where gestures require a wider area of motion [183] and especially for dynamic tasks (e.g., translating virtual objects).

Solutions

Boundary awareness remains a crucial challenge for recent tracking technologies such as Leap Motion and Kinect due to their cameras' limited field-of-view. One solution, as proposed in [152], is to use multiple devices at the same time to increase the tracked area. However, this is not feasible for HMDs as the sensors are fixed and mounted on the HMDs. In addition, because AR HMDs are meant to be mobile devices that enable users to move in both indoor and outdoor environments [66,156], setting up multiple depth sensors around the user is not a feasible solution for these AR devices. AR HMDs, unlike standard tracking devices like Leap Motion, is a combination of a tracking and display device, which can not only track users' hands but also provide visual feedback to the users. Therefore, in this chapter, we propose and evaluate an alternative solution to allow users to notice the tracking boundary by (1) showing the tracking boundary all the time, or (2) displaying the tracking boundary when their hands are about to leave the device tracking area. To the best of our knowledge, our study represents the first attempt to explore this issue of boundary awareness in HMDs.

Section 5.3 Formative Study

We could not find any prior work that has focused on boundary awareness in HMDs. To guide our design, we carried out a formative study to observe and identify challenges faced by users when interacting with AR HMDs with no explicit boundary awareness.

Section 5.3.1 Formative Study: Method

We recruited six participants (two females) from a local university, whose ages ranged from 18 to 27. During the one-hour study, we observed participants experiencing a variety of mid-air hand interaction tasks (e.g., manipulating virtual objects, sushi cat, HoloQuarium), while no boundary awareness was provided. After a tutorial, participants interacted with the AR HMD while following a thinking-aloud protocol. They were asked to talk about what they saw, what challenges they had, and possible improvements by having a boundary awareness method to guide their interaction explicitly.

Section 5.3.2 Formative Study: Findings

Our formative study led to three main findings that were extrapolated from participants' comments, our observations during their interaction, and post-experiment interviews.

(1) *Visualizing the boundaries*. During the study, participants had to cope with the system when there was no response to their gestures. In most cases, non-responsiveness was caused by the lack of awareness of the device's tracking area because their hands would stray outside of it. Participants were confused because they were unsure whether it was because of something that they did wrong. This led to 'uncomfortable feelings' and led them to question their ability to work with AR devices in general. This finding led us to hypothesize that if users could be made aware

of the tracked area (e.g., via some type of visualization), the cases of non-responsiveness would likely be reduced.

(2) *Distance to the boundary*. We wanted to investigate the issue of boundary awareness further and asked participants further questions. From the interviews, they indicated that it might be helpful to show how far between their hands were away from the boundary of the tracked interaction area (e.g., *P3*: ‘I could be careful of moving hands when I must interact the object near the boundary’). By knowing this, they could prevent their hands from hitting or going outside.

(3) *When to show the boundary*? Although visualizing the boundary seemed necessary, participants also argued that knowing the boundaries may not be that useful when there would not be risks of moving their hands outside the boundary. This was reasonable because the visual field-of-view (FoV) of HMDs is not large, and having additional visual information would increase the amount of information shown.

Section 5.4 Evaluated Boundary Awareness Methods

Findings from the formative study allowed us to propose the following boundary awareness techniques. The testing platforms were all developed and run in Unity3D. We have summarized the advantages and disadvantages of our visual methods for boundary awareness in Table 5-1.

Section 5.4.1 Static Surfaces (SS)

This condition provides a visualization of the interaction area in the form of planes or borders (Figure 5-2a). The surfaces are shown in blue (i.e., RGB color (0,0,128)) but with 40% opacity to allow users to still see through them. Blue is selected because it works well in indoor environments with white walls [14], which is our experimental environment setting. The area surrounded by the surfaces represents the interaction area. Moving the hand outside the interaction volume leads to tracking issues by the AR headset. The advantages of this method include: (1) allowing users to notice the boundaries easily; and (2) providing such information constantly. On the other hand, the disadvantages of this method include: (1) users have to infer the distance between the hand and the boundary; and (2) because it is visible at all times during interaction, it adds extra visual clutter that may occlude the view of other objects of interest.

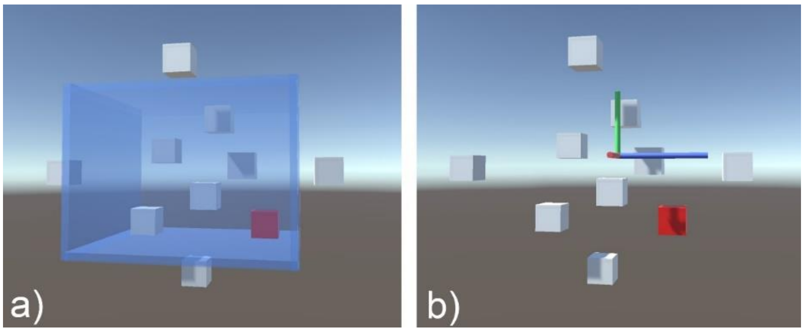


Figure 5-2. Design of static boundary awareness methods. (a) Static Surfaces (SS) that displays the interaction volume with colored transparent surfaces. (b) Static Coordinate Lines (SCL) that displays the distance to the closest interaction boundary in x-, y-, z-axes via coordinate system.

Table 5-1. Summary of the advantages and disadvantages of visual methods for boundary awareness that were tested in our study.

Techniques	Advantages	Disadvantages
SS	(1) allows users to notice the boundaries easily; and (2) provides this visual information constantly	(1) users have to infer the distance between the hand and the boundary; and (2) may occlude the view
SCL	(1) provides distance information between the hand and the boundaries; (2) provides this information constantly; and (3) fewer visual objects in the scene when compared to Static Surfaces	(1) visualizes the boundaries in an indirect way
DS	(1) helps visualize the boundaries in a clear way; and (2) provides such information dynamically and as such it does not occlude the interaction space when users' hands are far from the boundary	(1) users have to infer the distance between the hand and the boundary; and (2) there is still some degree of occlusion when users' hands are close to the boundary and the visuals are activated
DCL	(1) provides distance information from the hand to the boundary; and (2) the scene is clearer than (i) Static Coordinate Lines as the lines only appear when users' hands are close to the boundary; and (ii) does not occlude the view	(1) visualizes the boundaries in an indirect way

Section 5.4.2 Static Coordinate Lines (SCL)

In this approach, as long as the user's hand is inside the interaction volume, the distance between the users' hand to the volume's surfaces is shown through a 3D coordinate axis. The position of the coordinate center follows the hand position. The length of the line(s) indicates the distance to the boundaries (see Figure 5-2b). The advantages of this method include: (1) providing distance information between the hand and the boundaries via simple visuals (in this case lines); (2) providing such information constantly; and (3) there are fewer visual objects in the scene than SS. The disadvantage of this method is that it indirectly visualizes the boundaries.

Section 5.4.3 Dynamic Surface(s) (DS)

This condition visualizes the surface(s) when the user's hand only gets very close (i.e., 1.5cm) to the corresponding boundary (see Figure 5-3). Otherwise, it is analogous to the Benchmark method (no visuals are given). The advantages of this method include: (1) visualizing the boundaries in a clear way; and (2) providing such information dynamically and as such, it does not occlude the interaction space when users' hands are far from the boundary. The disadvantages include: (1) users have to infer the distance between the hand and the boundary; and (2) there will still be some degree of occlusion when users' hands are close to the boundary and the visuals are shown.

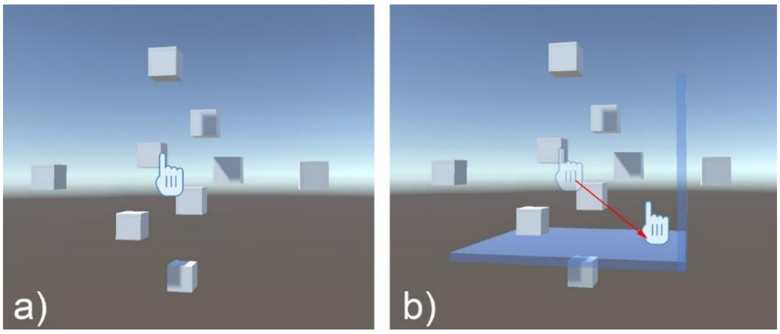


Figure 5-3. Design of Dynamic Surface(s) (DS) boundary awareness method. (a) The boundaries are not shown if they users are 1.5cm away the interaction boundary. (b) When users are about to move outside the interaction volume, at about 1.5cm to the surface, DS would highlight the corresponding surface(s) to let users be aware of the situation.

Section 5.4.4 Dynamic Coordinate Line(s) (DCL)

This condition is analogous to the SCL; the only difference is that the system only visualizes the coordinate line(s) when the user's hand gets very close (i.e., 1.5cm) to the corresponding boundary. Like DS, DCL does not show any visual elements for boundary awareness when the users' hands are outside the interaction area (see Figure 5-4). The advantages of this method include: (1) providing distance information from the hand to the boundary via simple visual lines; (2) the scene is clearer than (i) SCL as lines only appear when users' hands are very close to the interaction boundary, and (ii) the surfaces approach as line approach does not occlude the view significantly. Its disadvantage is that it is an indirect way to visualize the boundaries.

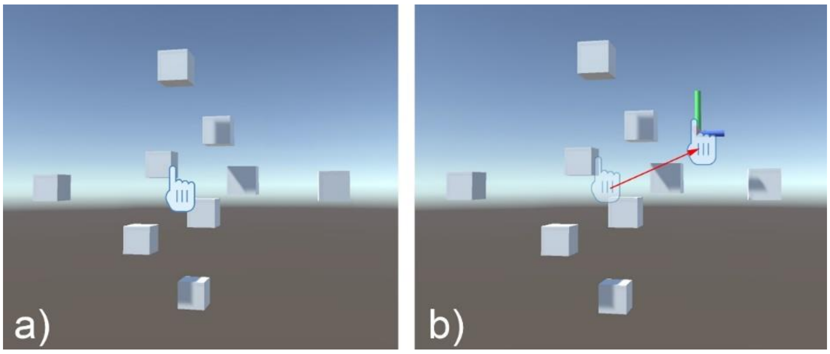


Figure 5-4. Design of Dynamic Coordinate Line(s) (DCL) boundary awareness method. (a) The coordinate lines are not shown if they users are 1.5cm away from the interaction boundary. (b) When users are about to move outside the interaction volume, at 1.5cm to the boundary, DCL would highlight the corresponding coordinate line(s) to let users be aware that they may possibly be exiting the area.

Section 5.4.5 Benchmark

This condition does not provide any visual feedback of the tracking boundaries and represents the case of how users currently interact with commercial HMDs. This approach acts as the benchmark when there is no boundary information provided to the users. It helps us to understand how users would perform and feel when there are visual cues provided to allow for a comparative analysis with the other four conditions.

Section 5.4.6 Tested Environment

The interaction volume is 25cm (width) × 20cm (length) × 16cm (height) and is placed at 42cm in front of the user as Magic Leap 1 can only display virtual items about 40cm away from the user. Users could only perform interaction when their hand is inside the interaction volume. Figure 5-5 shows the tested scenes together with the corresponded technique. There are eight cubes placed inside the interaction volume as target objects and four are outside the interaction volume (12 cm away from the surface and is outside the actual visual FoV of the Magic Leap 1) as target translation locations. Visual support is added to help user complete the task in two ways: 1) changing the color of the cube to green when the user's hand is hovering over a cube, this color would disappear when the player makes a successful selection, and 2) displaying an arrow to point out where the target location is when the user selects the cube successfully.

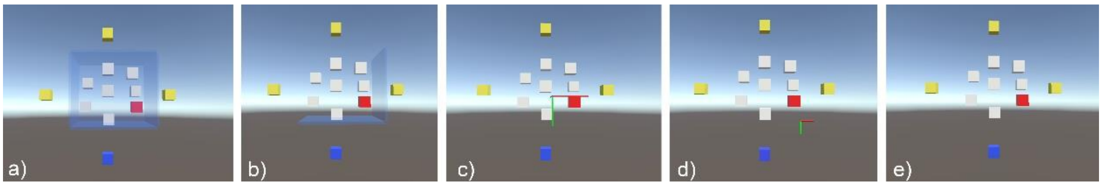


Figure 5-5. Experiment setting for each boundary awareness technique. a) Static Surfaces, b) Dynamic Surface(s), c) Static Coordinate Lines, d) Dynamic Coordinate Line(s), and e) Benchmark. Note the default Unity3D background was not visible during the experiment.

Section 5.5 Experiment: Object Translation

To better understand what the best way is to notify users that they are moving their hands outside the tracking boundary, we looked at user performance and preference for one common and important mid-air interaction—object translation [32]. We conducted a controlled experiment investigating ***RQ1*** (How accurately and efficiently can users interact with the system in dynamic tasks [e.g., translating virtual objects] when they cannot see the interaction area) and ***RQ2*** (How do boundary awareness methods affect the user's subjective feelings of translating virtual objects in HMDs) to explore mid-air translating (dynamic) tasks that would require a more complicated interaction process, from first selecting an object and then moving it to a different location within the AR environment.

Section 5.5.1 Participants and Apparatus

Twenty participants (seven females, average age: 20.2±2.2 years old, all right-handed with an average arm length 71.4±4.1 cm) were recruited from a local university campus. They all had normal or corrected-to-normal (using contact lenses) vision. Fourteen of them had prior experience with AR HMDs, but all were not frequent users. None had prior experience with the AR HMD used in the experiment—Magic Leap 1. The experiment was conducted in a university lab.

Section 5.5.2 Evaluation Metrics

We measured task performance in the form of objective data (speed and accuracy) and collected data describing users' preference to the methods, including subjective feedback (system usability, user experience, workload, arm tiredness, vision tiredness).

Task Performance

The task-completion time was the translation time from the first successful selection of the cube made by the participant to the time when the cube was dragged and dropped at the target location. The error was the number of times the cube hits the boundary as the participant's hand was not rendered by the AR system (i.e., moving outside of the boundary).

User Preference

User Preference was measured by 59 questions compiled from the System Usability Scale (SUS) questionnaire, NASA-TLX workload [107], User Experience Questionnaire (UEQ) [149], Borg CR10 [29], and Computer Vision Syndrome Questionnaire (CVSQ) [238].

Section 5.5.3 Experiment Design and Procedure

The experiment employed a one-way within-subjects design where the independent variable was Technique (SS, DS, SCL, DCL, and Benchmark). The order of the techniques was counterbalanced.

Before the trials started, participants were asked to complete a pre-experiment questionnaire to gather demographic information and were then given three minutes to get familiarized with the Magic Leap 1. Before each condition, they were briefed with the details of the next tested technique. During each condition, a one-minute training session was provided for each participant at the beginning. After each condition, participants were asked to fill in the user preference questionnaires. After the experiment, participants were asked to rank the techniques and give comments on the techniques. The whole experiment lasted about 80 mins.

Section 5.5.4 Task

During the experiment, the system would randomly indicate a target by changing its color. Users could use their index finger to target the cube they want to select and select it by using a palm open gesture. The color of the cube would be changed back to the default color and the target location would appear when the selection of the cube was made successfully. To complete the task, the user would need to drag the cube and drop it on the target location (i.e., hitting the center of the cube in the target location). A wrong selection did not cause any effect while an error (i.e., dragging the cube and hitting the boundary) would stop the cube from moving. Participants had to re-select the target if they performed an incorrect selection or made an error. There was a one-second gap for the next target to be highlighted after a successful translation. Each cube would be moved to all target locations once. Overall, each participant moved 160 targets (32 cubes × 5 techniques).

Section 5.5.5 Results

We first applied a Shapiro-Wilk test to evaluate whether the collected data were normally distributed. Then, unless otherwise specified, we employed a one-way repeated ANOVA with Technique as the within-subjects variable. Bonferroni correction was used for pairwise comparisons and Greenhouse-Geisser adjustment was used for degrees of freedom if there were violations to sphericity in the data.

User Performance

The analysis unveiled that Technique had a significant ($F_{2.733,51.936} = 2.872$, $p < .05$) effect on the task-completion time. Post-hoc tests confirmed a significantly lower time for the DS compared to DCL. As for errors, a Shapiro-Wilk test indicated that the data were not normally distributed, therefore, we conducted a Friedman's ANOVA where the analysis yielded a significant effect of Technique on errors ($\chi^2(4) = 10.539$, $p < .05$). Post-hoc analysis with Wilcoxon singed-rank tests were conducted with Bonferroni corrections, resulting in a significance level at $p < .005$. We found that DS had significantly ($p < .001$) smaller number of errors than DCL. Table 5-2 depicts the mean task-completion time and errors occurred for all conditions.

User Preference

NASA-TLX workload. Table 5-2 depicts the mean mental workload for all conditions. The analysis yielded no significant effect of Technique on overall workload ($F_{4,76}$ = 1.164, p = .334). Regarding the NASA-TLX subscales, the analysis yielded a significant influence of Technique on Mental workload ($F_{4,76}$ = 4.008, p < .01). Post-hoc tests confirmed that SS caused a significantly (both p < .05) lower mental workload than DCL and Benchmark. We did not find any significant effect of Technique on Physical Demand (p = .301), Temporal (p = .582), Performance (p = .464), Effort (p = .778), and Frustration (p = .401) subscales.

Table 5-2. Objective measurement and subjective feedback ratings with significant differences between the Boundary Awareness methods. Significance results are highlighted in green.

Method	Task-Completion Time	Error	Mental Workload
SS	2.19±0.67	14.25±12.52	36.00±18.68
DS	1.99±0.33	9.80±7.37	39.00±18.54
SCL	2.87±1.86	22.45±22.13	43.50±18.07
DCL	3.07±1.56	26.55±25.98	49.50±19.12
Benchmark	3.05±1.75	28.55±41.35	47.50±22.91
p	< .05	< .05	< .01

SUS. The analysis revealed that the Technique had no significant ($F_{4,76}$ = 1.686, p = .162) effect on the system usability. Benchmark (M = 71.5, SD = 13.72) had the highest SUS score while SCL (M = 65.37, SD = 12.23) and DCL (M = 65.37, SD = 13.98) had the lowest.

UEQ. The score for UEQ was analyzed using the excel tool provided by Laugwitz et al. [149] and had been adjusted between -3 (very bad) to 3 (excellent). The analysis yielded no significant influence of Technique on any of the UEQ subscale: stimulation (p = .983), efficiency (p = .702), perspicuity (p = .609), dependability (p = .859), attractiveness (p = .838), and novelty (p = .998). SCL (M = -0.23, SD = 0.20) had the highest UEQ score while Benchmark (M = -0.92, SD = 0.27) had the lowest.

Borg CR10. The analysis yielded no significant effect of Technique on perceived exertion ($F_{4,76}$ = .496; p = .739). DCL (M = 5.33, SD = 2.42) was rated that caused the highest physical fatigue for the participants while SS (M = 4.85, SD = 2.25) and SCL (M = 4.85, SD = 2.46) were rated the lowest.

CVSQ. A Shapiro-Wilk test indicated that the data were not normally distributed. Therefore, we conducted a Friedman's ANOVA where the analysis yielded no significant effect of Technique on perceived visual fatigue ($\chi^2(4) = 5.272$, $p = .261$). SCL was rated the worst (M = 2.35, SD = 4.13) while DCL (M = 3.40, SD = 4.12) was the best. The number of participants reported suffering computer vision syndrome in SS, DS, SCL, DCL, benchmark were 4, 3, 2, 4, 3, respectively. A binary logistic regression test showed that each Technique had the same level likelihood to cause computer vision syndrome ($\chi^2(4) = 1.082$, $p = .897$).

Ranking. The ranking of conditions shows a preference for SS (12 voted SS as the first option) before SCL (15 voted SCL as the second option). Benchmark was selected either the first or the last but mostly placed the last (14 voted it as the last option). Dynamic techniques were equally distrusted in the third and fourth places.

Qualitative Feedback

In general, most participants stated positive comments to Static and Dynamic Surface(s) boundary indicators: "great/good/wonderful" (*P3*, *P13*, *P20*), "easy to know the position and drag the cube" (*P6*, *P19*). However, we still observe a negative comment, "occluded the view" (*P15*). Regarding the Static Coordinate lines boundary indicator, participants indicated that "[it was] difficult to interact with the cubes" (*P3*, *P10*, *P19*). As for Dynamic coordinate line(s) boundary indicator, they stated that "like it" (*P17*), "easy to interact with cubes" (*P2*, *P5*). For benchmark, they stated, "the view is clear" (*P18*) but "extremely easy to move outside the boundary" (*P10*, *P11*, *P19*).

Section 5.6 Discussion, Guideline, and Future Work

Section 5.6.1 Task Performance and User Preference

Task Performance. We found that DS could not only help complete the task faster but also caused fewer errors than DCL. This could be because surface-based boundary awareness is much more apparent, explicit, and obvious than Coordinate Line-based methods. For *RQ1*, boundary awareness methods, in general, did not help to reduce the errors in translation tasks when compared to Benchmark. However, this was highly user-dependent; for instance, *P2* and *P3* had no issue interact with the Benchmark technique (less than 10 errors) while *P19* and *P20* made more than 100 errors.

Moreover, although *P19* and *P20* had many problems interacting with the Benchmark technique, they had no issues interacting with the AR environment with any of the boundary awareness methods, having fewer than 20 errors for all of them.

User Preference. For *RQ2*, boundary awareness methods could positively affect the user's subjective feelings during the interaction as we found that SS led to a significantly lower mental workload than Benchmark. One possible explanation is that users must be aware that they are moving outside of the tracked boundary in Benchmark condition while they did not have such an issue in SS. Interestingly, although SCL presents the tracking boundary all the time, it was not found to have the same effect as SS.

Based on the ranking data, SS is also preferred as the first option. Coordinate line-based methods are preferred by most users.

All in all, based on our results and user feedback, we suggest that in translation tasks, users should choose a surface-based technique (either SS or DS) over Benchmark as the technique could help users to know the boundary visually to guide their interaction. If users feel that their view is occluded and this interferes with their interaction, they could consider a coordinate line-based technique instead.

Section 5.6.2 Guidelines for Boundary Awareness

To our knowledge, this is the first exploration of boundary awareness for AR HMDs. Based on the results and observations of our study, we formulate the following guidelines and discuss implications for the design of boundary awareness methods in AR HMDs.

User-Dependent

Although there was no significant difference between methods on computer vision syndrome (CVS), we suggest that users should experience all available techniques first and avoid the one(s) which can cause computer vision syndrome to provide a better interaction experience. For example, *P15*, who suffered CVS with SS and Benchmark should not consider using them. In addition, participants, who made 117 errors (*P19*) and 153 errors (*P20*) while using Benchmark, should consider the technique(s) that could help them (e.g., SS for *P19* where only 15 errors occurred and DS for *P20* where only six errors occurred). All in all, the boundary awareness method should be tuned to suit the individuals' needs and predispositions.

Providing Boundary Awareness Method by Default

During the phase where participants tried the AR device to get to know it, we observed that novice users tended to over-value the FoV of the AR HMD. They would ignore the FoV of the AR device and assume that the interaction would be the same as what they would typically do during actual tasks. Therefore, we suggest that providing a boundary awareness method at the beginning stage to remind the users about the limited size of the tracked area and FoV of the device. It could be disabled when users think they could do without it.

Section 5.6.3 Limitations and Future Work

The design and results have some limitations, which could frame future research in this area.

Our experiment is limited to the mid-air interaction gestures with one-handed only. Future work can explore whether our findings will also be applicable to two-handed gesture-based interactions where large motions are required. As reported in [183], a gesture that requires a large moving may cause more errors and, therefore, might lead to a different experience.

Several values used in our experiment are pre-defined fixed values due to the lack of related prior work. For instance, we have set the color of the surface(s) blue since it works well in indoor environment with white wall [79]. Future work can 1) implement a dynamic color changing scheme for the surface(s) to suit the background [78,79]; 2) focus on exploring the most suitable values for opacity of the color and the distance for activating the dynamic visual cues for boundary awareness.

In this research, we have investigated the use of boundary awareness methods in translation tasks [32], with visual methods, which is only the starting point for investigating boundary awareness techniques in HMDs. It would be useful to examine the feasibility of boundary awareness methods in other common manipulating tasks in 3D environments (e.g., 3D modeling [51] where the interaction would be more complicated), other AR applications, and even in VR environments (e.g., to compare boundary awareness methods with the one offered by HTC VIVE/Oculus Rift in VR HMDs).

In addition, we have only implemented visual techniques for the boundary awareness problem. Other primary sensory channels [190], such as haptic and auditory, could present feasible and novel solutions but were beyond the scope of the current study. The development of non-visual techniques represents a rich area of future work. For example, audio, haptic, or their combination can be activated when users are about to move their hands outside the tracking boundary. This approach will involve less visual clutter, but more research is needed to understand how well they would work and to determine their optimal parameters.

Section 5.7 Conclusion

In this chapter, we present the first empirical study of visual methods for boundary awareness in head-mounted displays (HMDs). We have first conducted a formative study to understand the challenges that users would face when interacting with HMDs without boundary information. Then, we have introduced four preliminary candidates for boundary awareness that are then compared to the benchmark, where no boundary information is provided, in the common and important mid-air interaction task of object translation regarding task performance and user preference. Based on the results of our experiment, we suggest the boundary awareness method chosen should be user-dependent. We also list two guidelines for the use of boundary awareness methods in HMDs. Because mid-air interaction is an important aspect of current HMDs, issues such as boundary awareness are becoming increasingly critical. Our paper represents a first attempt at exploring and providing low-cost techniques that can improve mid-air interactions for these devices.

Section 5.8 Summary

We can now answer Research Question 1 of this thesis (i.e., how can visual boundary awareness techniques support mid-air hand-based interaction?). Visual boundary awareness methods should provide information on the distance between users and the boundary. In addition, it can be provided both statically and dynamically. Overall, visual boundary awareness methods could positively affect the user's subjective feelings during hand-based interaction.

To answer the Research Question 2 of this thesis (i.e., can a circular layout achieve an efficient hands-free head-based interaction?) and address the Core Challenge 2 (i.e., efficient hands-free head-based interface for HMDs), the next chapter first proposes an optimized circular layout through an iterative process. Then, the proposed technique is compared with four other possible head-based methods to prove its efficiency among other benchmark techniques. Finally, it presents a 4-day study to understand its efficiency after a great amount of training.

Chapter 6 RingText: Dwell-free and Hands-free Interaction for HMDs Using Head Motions

Section 6.1 Introduction

Most virtual reality (VR) head-mounted displays (HMDs) (e.g., Oculus Rift, HTC VIVE, Pico) and some augmented reality (AR) HMDs (i.e., Magic Leap 1) come with a controller device. However, there are cases where users cannot access the controller; for example, the controller is not around, or the users' hands are occupied with other activities. Besides, hands-free input will be useful for users who cannot manipulate a controller at all or with the precision required for text entry. Users who do not possess sufficient hand motor control skills like elderly users or those who have a motor deficiency disease will benefit from a hands-free technique. In this sense, having a technique that does not require users' hands to hold a device for input can come in handy in a variety of situations and for various types of users and HMDs.

Development of efficient text entry methods for HMDs without any dedicated handheld device has remained unexplored. A recent paper [294] reports a head-based text entry technique with dwell time that allows users to achieve an average of 10.59 word-per-minute (WPM) after training for 50 minutes. One limitation observed from their data is that the slowest users cannot improve much, even after having training. Another limitation is the dwell technique itself; it is well-known that dwell-based techniques can limit typing speed because of an imposed waiting time for each character selection. Text entry rates of dwell-based methods are typically between five to ten WPM [168]. By eliminating dwell time and optimizing the layout for selecting not only the letters but also the recommended spelling correction words, it is possible to increase WPM.

In this chapter, we explore the feasibility of applying a circular keyboard layout with two concentric areas for text entry that is both dwell-free and hands-free for HMDs (see Figure 6-1 for a picture of the technique and how it works). We have conducted three studies. The first study evaluates and compares how three possible factors (number of letters per region of the outer circle, size of the inner circle for resetting selection, and alphabet starting position) affect the efficiency of text entry, error rates, workload, and simulator sickness. Informed by both quantitative and subjective data, we then have improved and optimized the best layout (and features) from the first study further by narrowing the letter trigger area, adding a spelling correction feature, and incorporating dynamic, instead of fixed, candidate word regions for fast selection. Unlike other techniques that show the recommended candidate words in a fixed position [130,294], our dynamic candidate regions are designed based on Fitts' law [74] to enable users to choose quickly the desired word suggested by a spelling correction algorithm. In a second study, we have compared the text entry performance of our technique, RingText, with four other possible techniques: dwell QWERTY, dwell circular, Swype circular, and Swype QWERTY—the results show that RingText outperforms them. Finally, we have conducted a 4-day study with two daily sessions and 10 participants to evaluate the learning effects of RingText on speed and error rates. Our last session results indicate that the five novice users can achieve an average of 11.30 WPM (s.e. = 0.80) with 3.29% (s.e. = 0.34%) of the total error rate, and that the five 'expert' users (those who had performed the best in the second study) can achieve an average of 13.24 WPM (s.e. = 0.80) with 2.90% (s.e. = 0.22%) of the total error rate. Our results also show that our technique leads to a high selection rate of the recommended words due to the use of dynamic recommended word regions.

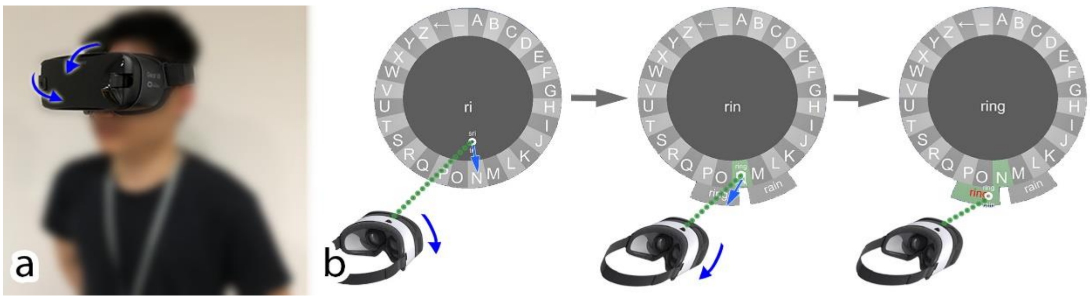


Figure 6-1. (a) Text entry on a mobile head-mounted display through head motions; (b) To finish typing 'ring' after a user has already entered the letters 'r' and 'i', the user selects the letter 'n'. The entered text is shown in the center of the screen; two candidate words are shown in the regions below and on each side of the last letter 'N'. Then the user goes to select the recommended word 'ring' by moving the head down. The design rationale of the technique is to minimize eye and head movements (or distance traveled), but still maintain a reasonably low error rate, of users of mobile virtual reality head-mounted displays.

The contributions of this work include: (1) the first example of a formal evaluation of the circular keyboard layout for text input in HMDs; (2) the first comparison of hands-free text entry mechanisms for both circular and QWERTY keyboard layouts in HMDs; (3) a case for the use of dynamic (rather than static) locations for recommended words—to our knowledge, this is a first case that shows the usefulness of using dynamic locations of these words; and (4) a demonstration of the effectiveness of RingText, a circular layout text entry technique that relies on head motions and uses dynamic locations for recommended words, through a 4-day user study.

Section 6.2 Related Work

In this section, we provide the literature review with respect to text entry for mobile VR HMDs (i.e., the device we used in the experiment); dwell-free text entry techniques; circular layouts; and dynamic vs. fixed positioning and use patterns of candidate words.

Section 6.2.1 Text Entry for Mobile VR HMDs

One of the biggest challenges for mobile VR HMD is to avoid the need of the peripheral devices generally used in stationary VR systems such as keyboards and mice [100] and game controllers [295]. This "accessory constraint" poses extra difficulties for text entry in immersive virtual environments (IVE) and limits the use of not only VR and also AR HMDs.

One possible solution is to use speech-based text entry techniques. Bowman et al. [34] made a comparison among a speech-based text entry, a pen and tablet keyboard metaphor, a one-hand chording keyboard, and pinch gloves, and found that the speech technique is the fastest medium for entering text in IVE at around 14 WPM. A recent speech-based multimodal text-entry system called SWIFTER [205] has claimed to reach an average input rate of 23.6 WPM. Despite their potential use in text entry, one major limitation of speech recognition techniques is that their performance suffers in noisy environments [100]. Furthermore, they can bring privacy problems when the user uses a speech text entry method to input a password or send messages to friends in a public environment, like a bus, coffee shop, or library. This represents a severe shortcoming for mobile VR HMDs which are often operated in an "uninstrumented" environment or public areas.

Other researchers have investigated touchscreen-based text entry techniques [101,136,159] and reported fairly good entry speeds (e.g., 17-23 WPM with a prediction algorithm [159]). However, because users are not able to precisely locate their hands before the first press in IVE [101], the typing process might require extra movements for selecting the target characters. Moreover, since a smartphone might already be used as a display for the mobile VR HMD, an extra touchpad is required for text input, and the use of hands is needed, something that is not possible in situations where users' hands are occupied.

Numerous mid-air typing techniques have been explored for virtual environments including wearable glove-based techniques [34] and motion tracking techniques [291]. Although such techniques enable mobile text entry and some of them allow a fast text entry speed (23 WPM for novice users as reported in [291]), these techniques might require expensive extra sensors or devices like cameras or sensor-equipped gloves. In addition, most of them require a substantial learning curve [100] and may confine users to a fixed location and position.

Current common mobile VR HMDs are designed to be operated using head rotation [101,197] by which users can move the cursor placed in the middle of the view to select target objects. Yu et al. [294] proposed and explored three types of text entry techniques using head-based interaction: Dwell, Tap, and Gesture with text entry speeds of 10.59, 15.58, and 19.04 WPM respectively for novice users after six training sessions. Among them, only their Dwell technique requires no extra device. Further, the input speed of their Dwell technique is not that high even with a prediction and error-correction algorithm (10.59 WPM). For these reasons, one of our key motivations is to propose a more efficient head-based device-free technique for mobile VR HMDs. Our design will eliminate dwell time and avoid the need of using hands (or additional input devices). More importantly, we aim to reduce motion sickness of users by minimizing the need to make large head motions.

Section 6.2.2 Dwell-free Text Entry Techniques

Instead of dwelling on the target for a predetermined duration to trigger a selection [169], dwell-free techniques allow users to enter text on-the-fly. Kristensson and Vertanen [142] investigated an eye gaze dwell-free text entry approach in a non-VR scenario and indicated that dwell-free eye typing could theoretically be significantly faster than existing techniques with a theoretical text entry speed of 46 WPM. Although this result is based on an error-free simulation, it suggests a possible research direction for dwell-free text entry techniques.

Dwell-free typing techniques can be divided into two major groups: gesture-based and selection-based. EyeWrite [281], the first gesture-based eye typing technique, has been shown to be significantly faster, easier to use, and prone to cause less ocular fatigue than the on-screen keyboard [282]. Eye-S [210] allows users to draw letters through sequential movements on nine hotspots and is claimed to reach 6.8 WPM for expert users. A later eye-typing technique, EyeSwipe [145], enables users to glance at the vicinity of the respective characters in the middle of the word but carefully selects the first and last characters of a word using the "reverse crossing" technique. It can reach 11.7 WPM on average for ten participants with 30 minutes of training. This technique is not fully dwell-free since it requires users to look at the hotspot for a pre-defined threshold time to confirm the sequence starting point. Gesture-based techniques are shown to suffer from low-performance issues [209].

Several selection-based dwell-free typing techniques have also been proposed. EyeK [230] allows users to select a character by moving the pointer inside-outside-inside the activation area. The authors have claimed it can achieve an average speed of 6.03 WPM. Filteryedping [203] can filter out unintentionally triggered letters from the sequence of letters swiped by the user and predicts the possible words. This technique is reported to achieve an average text entry speed of 14.75 WPM. One common drawback for most of these selection-based dwell-free techniques is that they might require extra movements to type the word (e.g., inside-outside-inside movements [230]). When used in HMDs this additional movement can increase motion sickness, which instead should be reduced.

There are some recent developments for VR HMD with eye tracking, but the cost of such devices is much higher than the standard HMD. For instance, the price of a FOVE 0 is $599 USD which is seven times higher than the Samsung Gear VR ($76) and also higher than other PC HMDs (i.e., Oculus CV1 – $399). Also, some research (e.g., [106]) suggests that head-based typing is as fast as gaze typing but can induce fewer errors. In line with this, we believe that dwell-free techniques have benefits for head-based text entry, including fast character selection, less error-prone than gaze typing, and high levels of acceptance by mobile VR HMD users.

Section 6.2.3 Circular Layout

Circular Keyboard Layout

The circular keyboard is first designed to work with pen input for desktops and touchscreen phones (e.g., Cirrin [174]). Later circular keyboard styles are designed to work without the stylus. TUP [212] maps the letters at fixed positions around a circle. Users place their finger on the location of the letters for selection. With the aid of a prediction algorithm, novice users can achieve 6-7 WPM.

The circular layout has also been used in gaze typing. pEYEs [120] employed a hierarchical circular interface with gaze-based input and reported a speed of 7.85 WPM for novice users and 12.33 WPM maximum for an expert user. Topal et al. [262] developed SliceType by applying a language prediction model to merge keys of their inner-outer circle layout. Their method can achieve 3.45 WPM for gaze input with 1 second dwell time. Apart from these works, the circular layout is also used in huge wall displays [244], VR with Dual Thumbsticks controller [295], and smartwatch [93,130,290]. So far, the best result for novice users using circular layout is appeared in WrisText [41], participants were able to type as fast as 15.2 WPM at the end of the fifth session.

Hierarchical Marking Menu

A hierarchical marking menu uses a set of multi-level radial menus and “zig-zag” marks to make selections [146]. This design concept has been used in many areas, such as fractal menus for AR HMDs [147] and Swipeboard [43] for smartwatch text entry where users can reach 19.58 WPM after two hours training. However, these examples are not based on a circular layout. Our review shows that there does not seem to be any research that has explored a hierarchical marking menu design with alphabet letters and suggested words using a circular layout.

Section 6.2.4 Placement of Candidate Words

Auto-complete, recommended words, and spelling corrections are commonly used in both research prototypes [130,203,294] and commercial products, like phones and tablets, to show possible words that users are trying to type. These suggested words are typically placed just above the T9 and QWERTY keyboard layouts.

Our review of the literature also shows that not much research has looked at the placement of suggested words for users to choose from. For QWERTY layouts, it is common to find word suggestions to be placed just above [294] or below [203] the virtual keyboard—the assumption seems to be that this placement will lead to fast and accurate selection. In addition, the placement is usually fixed in one region. While fixed placement either above or below the keyboard works for QWERTY layouts, this design may not be the most optimal for other keyboard layouts.

For a circular keyboard layout, placing the candidate words far away from the keyboard [130] makes it difficult for users to check the words and select them. The candidate regions and its selection used in the circular layout on smartwatches are efficient; the user can choose a candidate word by pinching the thumb and index fingers [41] or by pressing a side button [290]. However, these techniques applied in smartwatches are unlikely feasible for hands-free and controller-free HMD text entry scenarios.

Beyond smartwatches, our research points to a lack of research in the design and use of candidate word regions for circular keyboards. Their placement should be such that the user does not need to look back-and-forth between the keyboard and the suggested words, which are updated after each letter entry. Besides, if a cursor or a pointer is used for selection, its placement should aim to reduce the distance between the last selected key on the keyboard and the potential word that the user has in mind. In VR systems when using hands-free and controller-free circular text entry layout, dynamically positioning the suggested words could be a way to minimize the back-and-forth eye movement to check the words and can also reduce the distance (and hence the time) that is needed to make a fast selection. Our technique uses a dynamic location positioning for recommended words and, as described later, results from our experiment show that indeed dynamic placement brings advantages for text entry for circular layouts using head motions for selection.

Section 6.3 RingText

Section 6.3.1 Layout

To achieve dwell-free, our technique divides the boundary of the outer circle into equal size regions to hold the characters (see Figure 6-2 below). The region can potentially hold one or more characters. The inner circle can be regarded as the rest/reset area; users can stay at the center, while their eyes are searching for the next letter. To minimize learning, we have organized the letters based on alphabetical order to leverage users' familiarity with this sequencing.

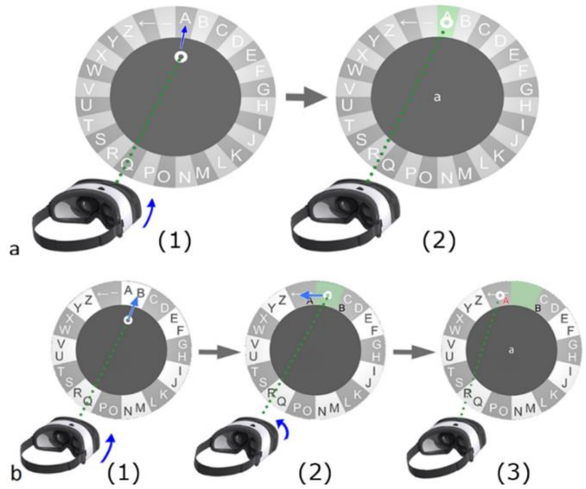


Figure 6-2. Design of the layouts and selection mechanism. (a) The 1 letter per region selection mechanism; and (b) The 2 letters per region selection mechanism. In both cases, a user is selecting the letter 'A'.

Keyboard size was determined in a pilot study with eight participants. We rendered the virtual keyboard far away from the user (8 meters) to avoid the parallax effect [294] and tested the keyboard size with a radius of 5, 5.5 and 6 meters in this preliminary study. We employed the 5.5-meter keyboard in our subsequent studies because of these participants' preference.

Section 6.3.2 Selection Technique

In this section, we describe briefly how our selection mechanisms work. First of all, it is important to note that two keyboard layouts are used in our first experiment. The first layout has only one letter per region and the second has two.

The one letter layout uses a simple go-and-hit selection approach—i.e., as the cursor leaves the center entering the region of a letter, this letter is instantly selected (Figure 6-2a). Since the second layout has two letters per region, the simple go-and-hit does not work. For this layout, we use the following approach: as the cursor leaves the center entering the two-letter region, these letters are split and parallelly placed opposite to each other just outside the current 2-letter region. The user then chooses the desired letter by moving the cursor towards the letter. As the cursor hits the area, the selection is made (Figure 6-2b). The users must move the cursor back to the inner circle to restart the selection process—so to make the process consistent.

We also explored 3- and 4- letter-per-region keyboard designs, which have a selection mechanism similar to the 2 letters per region design. However, participants from our preliminary study believed those two designs to be too complicated to use; besides both led to a high error rate.

Section 6.3.3 Visual and Sound Feedback

Our technique incorporates a sound effect to notify the user after a letter has been selected. To complement the sound, the colour of the region also changes when the cursor enters the region so that the user knows whether the cursor is in the correct region. Also, the colour of the letter also changes for 0.2 seconds to inform the user that the letter has been selected. The typed words are placed at the center so that user can easily see them.

Additional visual feedback is provided for the 2 letters per region layout. That means that once a region is selected the letters within it will move to their respective nearest neighbours. The new position of the letters serves as a visual guide for the user to know to which direction to rotate their head to make the selection (Figure 6-2b (2)).

Section 6.3.4 Advantages of RingText

Our technique leverages the advantages of small head motions such as low cost and higher accuracy when compared to eye gaze [71,106]. Also, as the head moves, the eyes can move along, which might help users to perform faster the visual search of letters (and as described later, to find the recommended words). Further, we make use of head movements to eliminate the need for hand-held input devices; useful for a wide range of mobile scenarios when such devices are unavailable or inconvenient to use; it is actually preferred and suggested to use head pointing (or movement) when a hand-held controller is not available (see [251]). Finally, our layout allows us to reduce selection time through dwell-free selection—selection is made only with small head movements.

We next describe the three studies. The first study explores the factors that can influence typing speed and error rates so that they could be optimized in our technique. Study two then compares the tuned method with four other hands-free methods to evaluate their relative performance. Finally, Study 3 explores the performance of both novice and “expert” users over a longer training period.

Section 6.4 Study One

The goal of this experiment was to evaluate the effect of (1) the number of letters per region on the outer circle, (2) the size of the inner circle for resetting the selection, and (3) the starting position of the letters on speed and error rate. We also evaluated workload and simulator sickness.

Section 6.4.1 Participants and Apparatus

Eighteen participants (13 males and 5 females) between the ages of 18 and 28 (M = 20.83, SD = 2.60) were recruited from a local university campus. All participants were familiar with the alphabet because the language of instruction at the university is English but were not native users. Participants had normal or corrected-to-normal vision and reported an average of 4 for experience with the QWERTY keyboard on a scale from 1 (‘No Skill’) to 5 (‘Expert’). Fourteen participants had previous experience with HMDs before the experiment—they had either seen and/or interacted with them. The experiment was conducted on a 96-degree field-of-view Samsung Gear VR with an S6 Edge+ smartphone. Unity3D was used to develop and implement our proposed head-based text entry technique. Our application also logged the cursor movement data for further analysis (like the heat map of selection areas).

Section 6.4.2 Experiment Design

The experiment used a 2 × 2 × 3 within-subjects design with three independent variables. The first was the number of letters per region (LPR) which had two levels: 1 LPR and 2 LPR. The second was the inner circle size (Center Size) which had two levels: Large (65% of the whole circular layout size—3.575-meters) and Small (55% of the whole circular layout size—3-meters). The last variable was the alphabet starting position: Left, Top, and Right (see Figure 6-3).

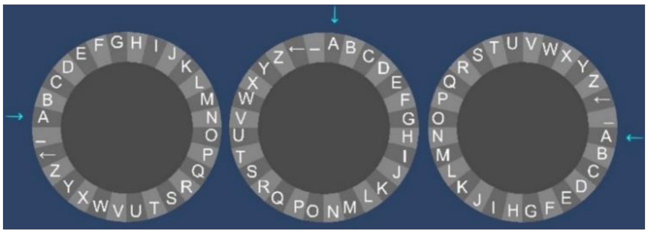


Figure 6-3. Three alphabet starting positions. From left to right: alphabet starting on the left; alphabet starting on the top; and alphabet starting on the right.

LPR and Center Size were counterbalanced; the alphabet starting positions were randomly assigned but also balanced for each condition. All three alphabet starting positions were tested by each participant. Each keyboard layout was randomly tested by six participants.

Each participant transcribed eight phrases for each layout combination. All phrases were randomly sampled from the MacKenzie’s phrase set [167] with no repeated phrases within the session. Each phrase was displayed in the central area. The Gear VR touchpad was applied only for the user to switch to the next phrase. Text entry speed was measured in WPM, with a word defined as five consecutive letters, including spaces. The error rate was calculated based on the standard typing metrics [250], where the total error rate (TER) = not corrected error rate (NCER) + corrected error rate (CER).

Section 6.4.3 Procedure

Before each session, all participants were briefed about the experiment details; then a 1-minute training was provided for the participants before each layout to allow them to familiarize with it. After each layout, the participants were asked to fill the NASA-TLX [107] and simulator sickness questionnaire (SSQ) [131]. Because our technique required frequent neck motions, we also added additional Neck Fatigue questions to SSQ. A 1-minute break was given if the participant felt tired. Before the experiment ended, all participants were asked to choose their preferred layout (LPR × Center Size) and alphabet starting position. This experiment took on average 45 minutes per participant. In total, we collected 18 participants × 2 Center Sizes × 2 LPR × 8 phrases = 576 phrases.

Section 6.4.4 Results

We employed a mixed factorial ANOVA and Bonferroni corrections for pair-wise comparisons. We also used a Greenhouse-Geisser adjustment to correct for violations of the sphericity assumption. Effect sizes were reported whenever feasible (η_p^2).

Text Entry Speed

Figure 6-4 illustrates mean text entry speed for each layout. A 2 × 2 × 3 (LPR, Center size, alphabet starting position) ANOVA tests revealed a significant difference of LPR ($F_{1, 60} = 4.042$, $p < .05$, $\eta_p^2 = .063$, observed power = .507), LPR × alphabet starting position ($F_{2, 60} = 3.254$, $p < .05$, $\eta_p^2 = .098$, observed power = .598) and Center Size × LPR × alphabet starting position ($F_{2, 60} = 4.364$, $p < .05$, $\eta_p^2 = .127$, observed power = .734) on WPM. No other factors were found to have a significant effect on WPM.

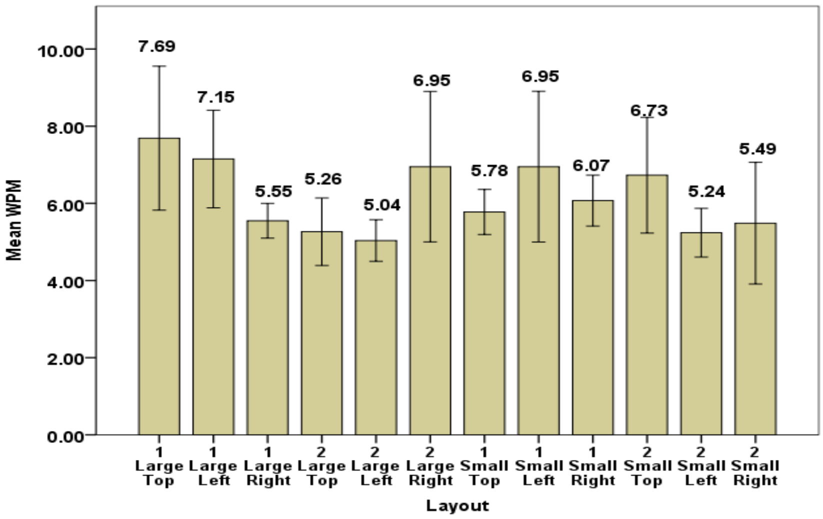


Figure 6-4. Mean text-entry speed across 12 types of RingText layouts. Error bars indicate ±2 standard errors.

Post-hoc pairwise comparisons for LPR indicated that WPM for 1 LPR was significantly higher than 2 LPR ($p < .05$). Post-hoc pairwise comparisons for LPR × alphabet starting position indicated that the text entry rate in 1 LPR Left was significantly higher ($p < .01$) than 2 LPR Left. No other significant differences were found. To test for significant effects on Center Size × LPR × alphabet starting position, we made pairwise comparisons which revealed that participants were significantly ($p < .01$) faster when typing with 1 LPR Large Top than 2 LPR Large Top. Also, participants were significantly faster ($p < .05$) when typing with 1 LPR Large Left than 2 LPR Large Left. Additionally, 1 LPR Large Top led to significantly faster ($p < .05$) speed than 1 LPR Small Top. No other significant differences were found.

Error Rate

Figure 6-5 shows TER and UCER for each layout. ANOVA testes revealed a significant difference of LPR on TER ($F_{1, 60} = 8.601$, $p < .01$, $\eta_p^2 = .125$, observed power = .823), while Center Size had a close to significant effect on TER ($F_{1, 60} = 3.739$, $p = .058$, $\eta_p^2 = .059$, observed power = .477). No other significant differences were found on TER. No main effects were found to be significant on NCER. Center Size × alphabet starting position was the only interaction effect to be significant on NCER ($F_{2, 60} = 3.683$, $p < .05$, $\eta_p^2 = .109$, observed power = .656). Post-hoc pairwise comparisons revealed the Large Left layouts (M = 2.69%, s.e. = 0.80%) had a close to significant ($p = .055$) more NCER than Small Left layouts (M = 0.66%, s.e. = 0.19%).

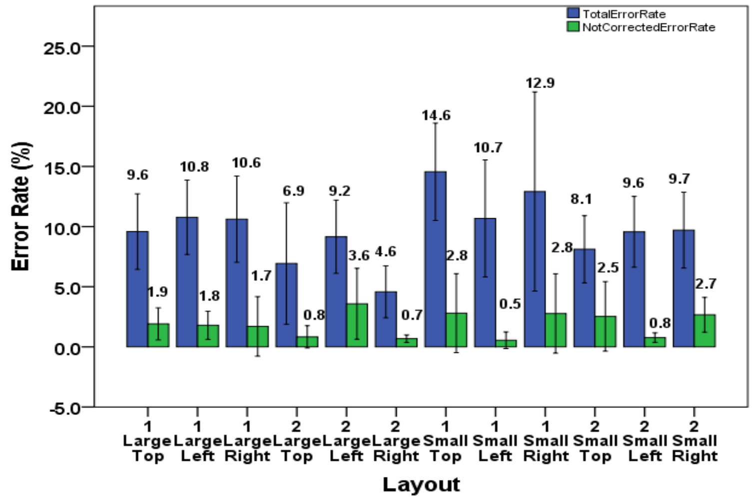


Figure 6-5. Mean TER and NCER across 12 types of RingText layouts. Error bars indicate ±2 standard errors.

Subjective Feedback

NASA-TLX. ANOVA tests showed that there was no significant difference of Center Size ($F_{1,60}$ = 0.003, p = .910, η_p^2 = .000, observed power = .051), LPR ($F_{1,60}$ = 2.021, p = .160, η_p^2 = .038, observed power = .327) and alphabet starting position ($F_{2,60}$ = 0.048, p = .954, η_p^2 = .001, observed power = .056) on the overall workload and its subscales (Mental, Physical, Temporal, Performance, Effort, Frustration). No interaction effects were found either.

Simulator Sickness. ANOVA tests yielded no significant difference of Center Size ($F_{1,60}$ = .265, p = .609, η_p^2 = .004, observed power = .080), LPR ($F_{1,60}$ = .009, p = .923, η_p^2 = .000, observed power = .051) and alphabet starting position ($F_{2,60}$ = .675, p = .513, η_p^2 = .022, observed power = .158) on the overall simulator sickness scores and the subscales (Nausea, Oculomotion, Disorientation). No interaction effects were found on the overall simulator sickness scores and its subscales.

User Preference. Fifteen participants (out of 18) preferred the alphabet to start on the Top; two users on the Right; and one user on the Left. In terms of the layout, seven participants preferred 2 LPR with the small inner circle; six users preferred 1 LPR with the large inner circle; two participants preferred 1 LPR with the small inner circle; three participants selected 2 LPR with the large inner circle.

Section 6.4.5 Discussion

Because all layouts have similar simulator sickness and TLX workload, we discounted the results. We only considered the performance data, users' preference and comments to decide the final layout and select the features that would be optimized and tested in the next experiment.

Overall, 1 LPR was significantly faster than 2 LPR; TER could be potentially solved by a spelling correction algorithm—our results in the next experiments would support this. No significant difference was found between 1 LPR and 2 LPR on NCER. In addition, all participants commented that 1 LPR is much easier to understand and use than 2 LPR. Therefore, we decided to use 1 LPR layout.

Although Center Size only had a close to significant difference on TER, the results showed a reliable trend that a large center should result in lower TER. Thus, we decided to use the large inner circle to minimize the possibility of inducing errors.

Regarding the alphabet starting position, because it did not have any significant difference on WPM and error rates, we chose the alphabet starting at the top based on user preferences. Thus, the final layout we selected was the 1 LPR large center with the alphabet starting at the top.

During the data analysis, we also observed that selecting a letter that was next to the intended one was the main reason why error rates were high. For example, one of our participants wanted to delete an erroneously selected letter. He then moved to the delete letter region, but unintentionally entered the space region twice because the trigger area for the space region and the delete region were very close to each other. To overcome this problem, we decided to narrow the letter region trigger area for the 1 LPR layout; by doing this, we believe it could help reduce the TER and lead to a faster text entry speed than the Dwell Type approach.

Besides, our observations also suggested that if the technique could include a spelling correction method, it would minimize erroneous inputs, thereby reducing the time that participants would need to correct them. As such, it could potentially increase text entry speed.

Section 6.5 Optimized Design

Section 6.5.1 Narrower Trigger Area

Figure 6-6 (a; left) presents the heatmap of the letter triggered locations collected from one participant. It shows that triggered locations are not in the midsections of the border adjacent to the inner circle, but across the whole border areas. Since the trigger areas are very close to each other, users may not find it easy to hit the intended letter region when they are not familiar with the circular layout of RingText, thus leading to error rates that are inevitably high. As shown in Figure 6-6 (b; right), to lower error rates due to accidental erroneous selections, a narrower trigger area for each letter is used (20% smaller than the original size).

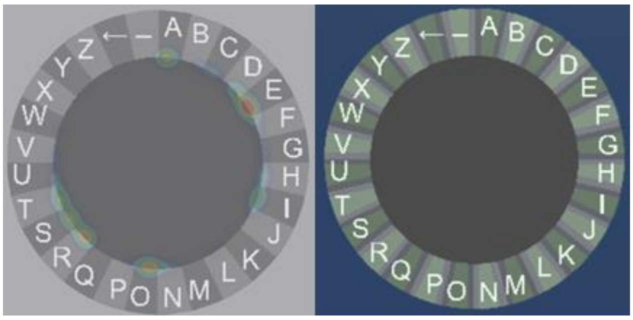


Figure 6-6. (a; left) An example of a heatmap of triggered locations. (b; right) Smaller trigger area of the letter regions used in Study 2 and 3.

Section 6.5.2 Spelling Correction

To further improve the performance of our text entry technique, SymSpell [81] was adopted with a dictionary of the ten thousand most frequently used English words [298]. To predict a word more precisely, we only allowed the algorithm to have its maximum search distance just two letters and return the top two spelling suggestions for the current typed letters. Figure 6-7 shows two examples of recommended words for two sets of letters.

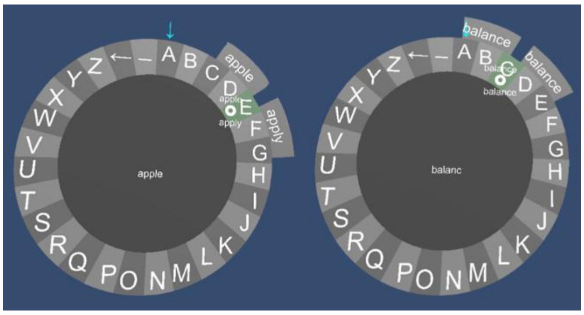


Figure 6-7. Dynamic candidate word locations for the letters ‘C’ and ‘E’ regions. The two results of the spelling correction algorithm are displayed next to the current letter region and close to the cursor to minimize not only eye movement for checking the words but also head movement for rapid selection of the words.

Section 6.5.3 Fixed vs Dynamic Candidate Word Locations

We also explored whether to use a fix location to show the spelling corrections or to have the locations change dynamically so that they would be shown based on the cursor’s location. Suitable fixed locations could be the areas outside the circle, but this approach would force users to look back-and-forth frequently, and this was something we wanted to minimize to lessen simulator sickness. The central area could also be problematic because it might lead to erroneous selections because users would need to rotate their head to cross to other letter regions. Other possible solutions were to use dwell, or to use an additional input device; however, both approaches would go against our design criteria. Moreover, a fixed location within the center area would still require users to move their head or eyes every time they would enter a letter region and want to see whether the word(s) shown were the ones they would need.

Instead of placing the recommended words in a fixed position, a dynamic solution was chosen. Dynamic locations could be based on the current location of the cursor. However, this would also require dwelling or an additional input device for selection. In the end, we decided that the two recommend words could appear just outside of the current letter region and, by implication, next to the location of the cursor (see Figure 6-7). This dynamic solution would minimize not only eye movements to check the words, but also head movements to select a word because of their proximity to the cursor and users’ focal viewpoint. In one way, this represented an extension of our selection technique for letters but applied to select words without the need of dwelling time and an extra device.

Using this approach, the spelling correction would only work when the user entered a letter region. The words would disappear when the user went back into the center area. Similar to selecting a letter (by moving the cursor to the letter region) the user would move the cursor into the word region once. After each selection, the user must go back to the center area. The logic behind this was that after selecting a word, the user would need to go to another letter region.

To further encourage users to select recommended words and improve text entry speed, a space character was automatically added to the end of a word after its selection. This design rationale followed Fitts' law [74]. The completion time was analyzed based on Fitts' law and the formula proposed by Mackenzie [166]

Equation 6-1 Fitts' law

$$MT = a + b\log_2\left(\frac{A}{W} + 1\right) \qquad (1)$$

where *MT* was the average time to complete the movement; *a* and *b* were model parameters; *A* was the distance from movement origin to the target center; and *W* was the width of the target.

In our case, the distance *A* from the current letter, to the word selection region would always be smaller than the distance to reach the "space bar". For *W*, we designed the candidate region to have a broader width than the "space bar" (Figure 6-7), so the completion time to get a space between words from the candidate region, in our layout, would always be smaller than the time to get it from the "space bar" (except from "A" or "<-"). In this way, there was no need for users to hit the space letter region.

Section 6.6 Study Two

The goal of Study Two was to compare five possible hands-free techniques, which were Dwell Circular (DC), Dwell-Free Circular (DFC), Swype Circular (SC), Dwell QWERTY (DQ) and Swype QWERTY (SQ). DFC was our technique that had been optimized based on features described earlier. Figure 6-8 shows examples of using SQ, SC, and DFC to enter the words "hello world". The techniques are described briefly in the next section.

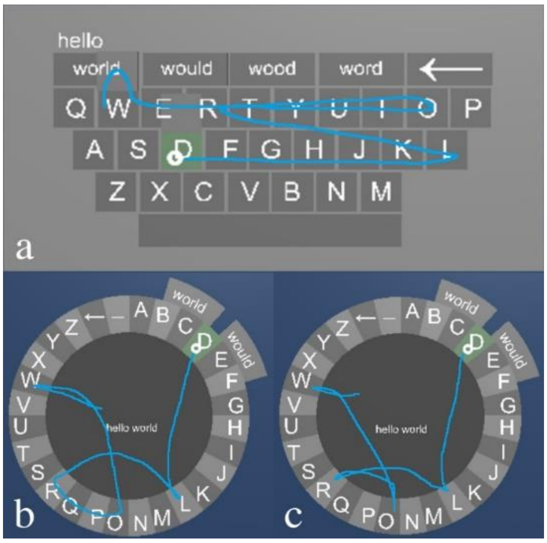


Figure 6-8. (a; top) An example of typing the word 'world' in Swype QWERTY; the interface of the Dwell QWERTY was same but without popup buttons (see the light grey block above letter 'D' and 'W'). (b; lower left) An example of typing a 'world' in Swype Circular. (c; lower right) An example of typing the 'world' in Dwell-Free Circular; the interface for Dwell Circular is the same except that users have to wait for 400 ms to select the letter from the letter regions.

Section 6.6.1 Design of the Testing Techniques

For each layout type, we kept the graphical aspects the same; the only difference was how letters could be selected. Between the circular and QWERTY layouts, we also kept all other parameters the same—e.g., the distance between the user and the keyboard. One difference between them was that the QWERTY layouts had four candidate words where circular layouts only had two. The reason for QWERTY layouts to have four candidate words was because previous research using the QWERTY layout had used four words instead of two.

For SQ, we adopted the method used in [145] for indicating the select action. An example of typing the word 'world' is shown in Figure 6-8a. At the beginning, the user moves the cursor to the target, then a button representing an action appears above the target after a wait time of 400 ms (i.e., the start of a Swype path); after the button appears, the user moves the cursor to the button followed by moving the cursor back to the target to perform the selection. When the user finishes the Swype action, the system provides four recommended words in the candidate regions (See Figure 6-8a, 'world' is the best-recommended word, 'word' is the fourth best-recommended word). The best match is automatically selected if the user starts Swyping the next vocable (e.g., 'world' in Figure 6-8a), however, if the match is not the best the user must select it directly, following the same procedure as selecting a single letter. During the Swype action, only letters are active and selectable, other special characters (e.g., space/delete) are not.

As stated earlier, the design of the dwell-free technique (DFC) was based on the features derived from the first study and described in the previous section. An example of how to use DFC can be found in Figure 6-8c. Of the three circular techniques, SC had a different selection feature; it allowed users to select the next letter (that was different from the last selected letter) without the need of returning to the inner circle—i.e., they could Swype to the next letter. An example of how to use SC is presented in Figure 6-8b.

For two dwell techniques (DC and DQ), we set 400 ms for one letter input and dwell for another 400 ms to make the double input. We adopted 400 ms because any smaller dwell time would be error-prone and larger dwell time would cause a low text input rate. This was consistent with the implementations of dwell techniques in prior research (e.g., [120]).

Backspace deleted the last input, be that a complete word or a single letter. For all techniques, the system would append automatically a space if the word was selected from the candidate regions. Swype-based methods and the spelling correction used the Damerau–Levenshtein distance algorithm for word suggestions. The same dictionary [298] was used among all techniques. SC and SQ applied the Swype algorithm, other three techniques used the Symspell spell-correction algorithm as mentioned in the previous section where we set the algorithm with the max search distance of two to enhance the accuracy.

Section 6.6.2 Hypotheses

We had two hypotheses for this study. Our first hypothesis *H1: DFC should be the fastest technique*. Our second hypothesis *H2: DFC should have the lowest error rate and the error rate should be significantly lower than other techniques*.

Section 6.6.3 Participants and Apparatus

Fifteen participants (10 males and 5 females; aged between 18 to 26; M = 21.4, SD = 2.03) were recruited from the same university campus as in the Study One. None of the participants participated in Study One. Their alphabet familiarity was the same as in Study One since they were the same demographic. All participants had normal or corrected-to-normal vision and reported that they were familiar with the QWERTY keyboard (M = 4.1, from 1 – No Skill to 5 – Expert). Only one participant had no experience with HMD before. This experiment used the same apparatus as Study One.

Section 6.6.4 Procedure and Design

The study followed a within-subjects design with one independent variable: Technique (DC, DFC, SC, DQ, and SQ). The order of the five hands-free techniques was counterbalanced. For each technique, participants needed to enter eight phrases, which were randomly sampled from the MacKenzie's phrase set [167] with no repeated phrases within the same session. Each phrase was displayed at the center of the inner circle for the circular layouts and above the candidate regions for the QWERTY layouts—this was consistent with practices from previous studies. Participants were instructed to type as quickly and accurately as possible. Between sessions, they were encouraged to take breaks if they felt tired. The study lasted around fifty minutes. In total, we collected 15 participants × 5 hands-free techniques × 8 phrases = 600 phrases.

Section 6.6.5 Results

We employed a one-way repeated measure ANOVA and Bonferroni corrections for pair-wise comparisons. We also used a Greenhouse-Geisser adjustment to correct for violations of the sphericity assumption. We indicate effect sizes whenever feasible (η_p^2).

Text Entry Speed

WPM ranged between 6.03 (s.e. = 0.40) for DC and 8.74 (s.e. = 0.53) for DFC (Figure 6-9). ANOVA yielded a significant effect of Technique ($F_{1.507,21.091} = 12.746$, $p < .001$, $\eta_p^2 = .477$, observed power = .975). The pairwise comparisons showed significant differences between DC and DFC, DC and DQ, DFC and DQ, DFC and SC, DFC and SQ, DQ and SQ (all $p < .05$).

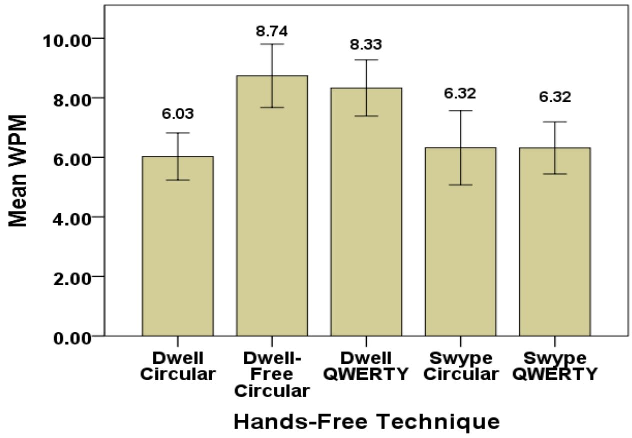


Figure 6-9. Mean text entry speed across the 5 hands-free techniques. Error bars indicate ±2 standard errors. The Dwell-Free Circular technique led to the fastest speed with 8.74 WPM on average.

Error Rate

Figure 6-10 shows the TER and NCER for the five hands-free techniques. Although the difference between each technique seemed large, from the ANOVA test we only found a trend toward a significant effect of the techniques on TER ($F_{2.313,32.376} = 2.652$, $p = .079$, $\eta_p^2 = .159$, observed power = .525). In addition, there was no significant effect of Technique on NCER ($F_{2.282,31.952} = 2.315$, $p = .109$, $\eta_p^2 = .142$, observed power = .464).

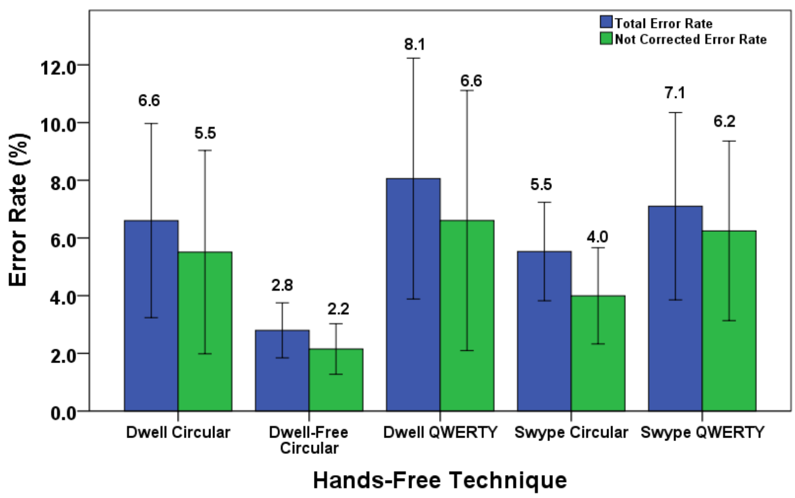


Figure 6-10. Mean TER and NCER across 5 hands-free techniques. Error bars indicate ±2 standard errors. The Dwell-Free Circular technique led to the lowest TER (2.8%) and NCER (2.2%).

Section 6.6.6 Discussion

Our results support *H1* (DFC has outperformed all the other techniques in text entry rate). On the one hand, *H2* is not supported where the difference in TER and NCER between DFC and other techniques is not significant (although the trend seems to be towards significance for TER). On the other hand, DFC has led to the lowest TER and NCER.

Considering that all features, except for the selection mechanism, have been kept the same in the three circular layouts, our findings suggest that the go-and-hit selection seems to be a better approach for a circular layout and that can work well with head-based motions. Surprisingly, the performance of SC is much lower than DFC, even though it can make selections which do not require users to move the cursor back to the inner circle. The reason may be because in DFC users only need to consider whether the candidate regions have the target word and, if they do not, they can directly go back to the inner circle to do the reset and move to the next letter. In SC, on the other hand, users not only need to consider the candidate regions, but they also need to consider whether they should go back to the inner circle or go through the outer circle to select the next letter—this cognitive process would have added extra burden and time for users to make the decision. A closer analysis of the typing process shows cases that users accidentally have typed some letters unrelated to the target word; this might have been caused by the wrong selection during the Swype action as users accidentally move back to the inner circle to select the wrong letter when they had decided to go through the outer circle.

The text entry rate of DQ is in line with the DQ technique tested in [294]. For DQ, some users have commented that 0.4s is very (almost too) short and has made them frustrated and uncomfortable—they have felt that something is pushing them to move to the next letter very quickly in order to avoid unintentional selections—i.e., they have found it not very usable. In contrast, in a non-dwell technique like our DFC RingText, users have felt relaxed, and this might have been the reason that users have been able to achieve a significantly higher text entry rate and close significantly lower TER (but at the same time still feeling comfortable).

Section 6.7 Study Three

Given that our dwell-free technique outperformed other four baseline techniques, we wanted to explore its performance if users could receive some more training for two groups, novices and experts. For the potential expert group, we ordered the participants from Study Two based on their average text entry speed, and invited those participants who achieved a relatively high text entry performance to continue for a 4-day study. For the novice group, we recruited participants who were not involved in either study one or two. The design of the third study followed a similar approach reported in previous works [93,295].

This third study was to last for four days with two daily sessions for each participant. The goal was to measure how well novice and expert users could improve their text entry speed and standard typing metrics [250] through practice over time.

Section 6.7.1 Participants and Apparatus

Ten participants (nine males; aged from 19 to 28, M = 21.6, SD = 3.17) were recruited from the same university campus as the previous two experiments; five of them who achieved a relatively high text entry speed in Study two agreed to join this 4-day study. They formed the potential 'expert' group. The five participants who were not involved in study one and two formed the 'novice' group. These participants had similar visual acuity and alphabetical knowledge as the ones from the previous studies since they represented the same demographic. They reported an average 4 for experience with the QWERTY keyboard on a scale from 1 ('No Skill') to 5 ('Expert'). All participants had some previous experience with HMD before. This experiment used the same apparatus as the previous studies.

Section 6.7.2 Procedure and Design

The study consisted of a series of sessions over four consecutive days, with two sessions per day. In each session, participants needed to complete eight phrases, which were randomly sampled from the MacKenzie's phrase set [167] with no repeated phrases within the same session. Each phrase was displayed at the center of the inner circle. All eight sessions lasted approximately an hour. In total, we collected 640 phrases (10 participants × 8 sessions × 8 phrases).

Section 6.7.3 Results

We employed a mix-design ANOVA with Sessions (from one to eight) as the within-subject variable and Group (novice and potential expert) as the between-subjects variable. Bonferroni correction was used for pair-wise comparisons and Greenhouse-Geisser adjustment was used for degrees of freedom if there were violations to sphericity in the data. We indicate effect sizes whenever feasible (η_p^2).

Text Entry Speed

ANOVA tests yielded a significant effect of Session ($F_{2.592,20.733} = 31.344$, $p < .001$, $\eta_p^2 = .797$, observed power = 1.000) and a close to significant effect of Session × Group ($F_{2.592,20.733} = 31.344$, $p = .058$, $\eta_p^2 = .276$, observed power = .591) on text entry speed. There was a significant effect of Group ($F_{1,8} = 8.127$, $p < .05$, $\eta_p^2 = .504$, observed power = .705) on text entry speed. This suggests that although participants in the two groups had a significant difference in text entry speed, their learning over time was somewhat similar.

Post-hoc pair-wise comparisons revealed significant differences between session 1-4, 1-5, 1-6, 1-7, 1-8, 2-4, 2-5, 2-6, 2-8, 3-8, 4-8, 5-8, 6-8 and 7-8 (all $p < .05$).

Overall, the average speed across all sessions was 10.45 WPM (s.e. = 0.28). In particular, the novice group achieved 8.9 WPM (s.e. = 0.30), while the potential expert group achieved 11.99 WPM (s.e. = 0.34). Figure 6-11 shows the mean WPM by sessions for each participant and the two groups. The average speed for the first session was 8.50 WPM (s.e. = 0.76); it bumped up to 12.27 WPM (s.e. = 0.62) in the last session, with an increase of 44.4%.

In the last session, the potential expert group improved their performance to 13.24 WPM (s.e. = 0.80) from the first session of 10.26 WPM (s.e. = 0.72); the novice group improved to 11.30 WPM (s.e. = 0.80) from the first session of 6.75 WPM (s.e. = 0.72).

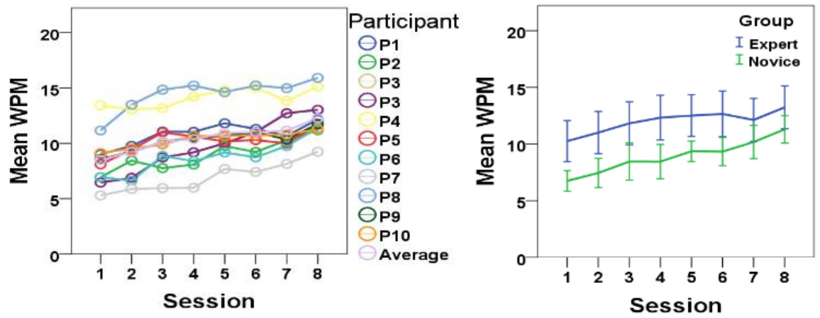


Figure 6-11. Mean WPM using RingText over 8 sessions for each participant (left) and the mean WPM for each group (right). Error bars indicate ±2 standard errors. The graphs show an upward trend for all participants. They also show that participants have not yet reached the peak.

Error Rate

For TER, ANOVA tests yielded no significant effect of session ($F_{7,56} = 1.462$, $p = .200$, $\eta_p^2 = .154$, observed power = .563), Group ($F_{1,8} = .109$, $p = .749$, $\eta_p^2 = .013$, observed power = .060), or Session × Group ($F_{7,56} = .452$, $p = .864$, $\eta_p^2 = .054$, observed power = .182). For NCER, ANOVA tests also yielded no significant effect of session ($F_{7,56} = .574$, $p = .774$, $\eta_p^2 = .067$, observed power = .226), Group ($F_{1,8} = .157$, $p = .702$, $\eta_p^2 = .019$, observed power = .064), or Session × Group ($F_{7,56} = .913$, $p = .504$, $\eta_p^2 = .102$, observed power = .356).

Figure 6-12 shows the mean TER and NCER over eight sessions. Overall, the average TER and NCER across all sessions were 3.10% (s.e. = 0.25%) and 2.25% (s.e. = 0.14%) respectively. In particular, the average TER and NCER for the potential expert group were 2.90% (s.e. = 0.22%) and 2.44% (s.e. = 0.25%), whereas for the novice group they were 3.29% (s.e. = 0.34%) and 2.05% (s.e. = 0.22%).

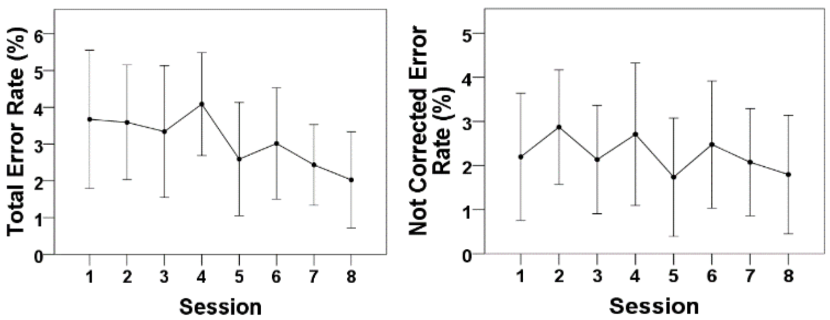


Figure 6-12. Mean TER (left) Mean NCER (right) over 8 sessions. All Error bars indicate ±2 standard errors.

Spelling Corrections Statistics

The total words participants were supposed to type in the experiment were 3261 (excluding the words with length fewer than two letters). Of these words, 2822 (again excluding the words with length fewer than two letters) were selected from candidate regions, including 986 words predicted in advanced and 1836 words corrected in the last letter. For those 2822 words suggested by the spelling correction algorithm, there were 2341 correct selections and 185 wrong selections.

Section 6.7.4 Is RingText Applicable to AR/MR HMDs?

We conducted a small, follow-up experiment at the end of the eighth session to test whether RingText would be applicable to AR/MR HMDs and could lead to similar performance to the VR version. We asked participants to try our technique on Meta 2 AR HMD. Five participants agreed to do the experiment. Thus, we collected 5 participants × 8 phrases = 40 phrases.

The results from these five participants pointed to a positive experience. They were able to achieve an average text entry speed of 12.06 WPM with a low level of TER and NCER (1.82% and 1.44% respectively) on the Meta 2 HMD. This performance was very similar to the results in the last session using the Gear VR device (12.24 WPM, 1.42% TER, and 1.13% NCER).

Based on these results, we can infer that our technique has the potential to lead to comparable results not only in AR but also likely in MR HMDs as well; thus, it is very likely that RingText can be easily adapted to other HMD systems.

Section 6.7.5 Discussion, Limitation, and Future Work

Text Entry Speed and Error Rate. The average speed of RingText across sessions for novice and expert users are 8.9 WPM and 11.99 WPM. Novice users can type 11.30 WPM after 1 hour of practice where expert users can reach 13.24 WPM after 1.5 hours of training (including the time they spent in Study two with circular layouts). This result indicates that RingText outperforms some other dwell-free techniques such as EyeK [230], Eye-S [210], and EyeWrite [281] with 6.03 WPM, 6.87 WPM, 7.99 WPM, respectively. The text entry rate after training is comparable to the speech input

(6-13 WPM) [96,117], and leads to better performance than the head-based dwell method in [294] (10.59 WPM). In terms of word-level TER and NCER, RingText achieved a 3.10% and 2.25% across sessions, which are comparable with the head-based dwell techniques for HMDs reported in [294] (3.79% and 2.46%).

As mentioned before, all our participants are not native alphabet users. It can be argued that given their familiarity with the alphabet, native users could lead to higher text entry speeds than non-native users, similar to the result reported in [295]—this latter group are almost identical to our participants (they are university students within the same age range and whose language of instruction is English but are not native alphabet users). However, future work is needed to confirm whether native users could achieve a significantly better result than non-native alphabet users with RingText.

Overall, significant learning effects were observed in text entry speed, indicating the possibility of even higher text entry speeds with further practice—as Figure 6-11 shows an increasing trend for text entry speeds even in the final session and participants' performance has not peaked yet.

Design of dynamic, non-fixed candidate regions. This work makes the first attempt to combine the circular layout with dynamic candidate regions that are placed just next to the region of the last selected letter. The percentage of the candidate word selections shows that our candidate regions are used very frequently (86.5% of the words have been chosen from the candidate regions). There are three main reasons that explain why our design has led to such high frequent use.

(1) *Minimal checking time*. The time for users to check whether a candidate region had the correct suggested word is reduced as these regions are close to the current letter region which would likely be where the users would be paying to attention to at the moment.

(2) *Reduced travel distance*. Unlike the design in [130], users only need to travel a short distance to hit the region to select a word because the cursor is just next to the candidate regions.

(3) *Space automatically appended*. Users have commented that they have automatically thought of the candidate regions as an easy way to get the space

character. Our observations show that even though in cases when all letters of a word are already entered corrected, participants would move the cursor to the candidate region to select because its distance is often shorter the distance to the letter region of the space character.

An additional option for the hands-free and controller-free scenario. Considering the design guidelines in [251], we recommend the RingText as an additional option for hands-free and controller-free scenarios, since the text entry rate is significantly better than the head pointing dwell techniques and comparable to the speech input [96,117] but with no significant drawbacks in recognition problems and no privacy problems for users when typing in public places. There are several scenarios that people can use RingText; for example, when users receive a message while watching a movie in VR or when they want to send a quick chat text in a VR multiplayer game, they can simply popup RingText and quickly type the message.

Limitations and Future Work. The present research has several limitations, which can also serve a possible direction for future work.

RingText is based on head-pointing so that it might be inappropriate for people who cannot rotate their head—e.g., users with a neck injury. Moreover, we have evaluated in a lab which shows that users have no issues using it in a non-public environment. We have not looked at issues of social acceptability when users want to use it in public places.

It would have been good to use a standardized interface usability survey (like the System Usability Scale) in our first two studies so that we can compare across techniques. This is something that could be done in future studies dealing with new keyboard designs.

RingText shared one limitation with other keyboard design where the default keyboard letters are in lowercase where uppercase letters, symbols, and emoji are required. Future research could explore how RingText would scale up to support uppercase characters and symbols. One possible solution is to use the forward head movement to switch between sub-layouts with different types of characters and symbols. We have tried measuring forward and backward head movements, and current mobile devices can detect these types of motions. It is possible to set a forward acceleration threshold which can be used as an indicator for when users want to switch layouts. Future research is needed to determine how this approach will work.

We have not investigated the optimal size of the trigger area for RingText. Smaller trigger areas of the letter regions can lead to a lower error rate, but it might also result in a lower text entry rate since users may miss the trigger area of the intended letter and must re-enter it to make the selection. Future work is needed to investigate the optimal size(s) of the trigger area to let users select letters quickly without incurring many mistakes. Additionally, we can apply a static decoding method [97] to handle the noise of the input further. This is similar to a method to mitigate the "fat finger" problem in smartphones [268] where users with large fingers may mistakenly select unintended buttons. In our case, it may be possible to use this model to help us understand which letters the user is aiming to type.

As stated earlier, participants in Study Three did not reach peak performance after eight sessions. In similar experiments reported in [93,294,295], their participants had 5-6 sessions and could not reach it either. We designed the experiment with eight sessions assuming that 2-3 extra sessions would have allowed participants to reach a stable text entry rate. It may be of interest to explore if there is a common minimum period of training time that participants need to reach maximum performance with RingText and similar techniques.

Finally, the dwell-time for Dwell technique and the algorithm for Swype technique tested are based on their common implementation. In the future, it may be useful to compare RingText with other variations of these techniques that use some optimized features.

Despite these limitations, our results show the potential use of circular layouts in head-based dwell- and hands-free text entry in HMDs system (e.g., mobile VR HMDs).

Section 6.8 Conclusion

We have provided the first example of a formal evaluation of ring-based text input for head-mounted displays (HMDs) that is both dwell-free and hands-free. Our example technique, RingText, allows users to enter text by making small motions with their head and select letters from a circular keyboard layout with two concentric circles: the outer circle contains letters housed in distinct regions, while the inner circle serves to reset selection and allows users to search for the next letter.

In our first study, we determine the suitable size of the inner circle, the number of letters per region (LPR) in the areas of the outer circle, and alphabet starting position. The results show that 1 LPR leads to a significantly better performance in entry text speed; a larger center area can potentially decrease error rates, and users preferred the alphabet to start from the top. Based on the results, an optimized layout that shows two recommended words placed dynamically next to the cursor is adopted to develop RingText. Then, a first comparative study of hands-free text entry techniques in HMDs has been conducted by comparing the RingText with four other text entry mechanisms. Results show that RingText is the most efficient technique; it has led users to achieve a significantly higher text entry rate and close to a significantly lower total error rate. To further explore its performance, a third study is undertaken with 10 participants doing two daily sessions for four consecutive days. The results of this last study show that after eight practice sessions even novice users can achieve an average text entry speed of 11.30 WPM while expert users can achieve 13.24 WPM in the last session. Because performance over these sessions shows an increasing trend, we believe that there is some place for improvement in their text entry speed with further practice sessions.

All in all, RingText is an efficient technique for text entry in head-mounted displays that do not require users to hold any additional input devices. We hope this work can inform future work on dwell-free and hands-free text entry techniques based on a circular layout for all types of HMDs.

Section 6.9 Summary

According to the above findings, the answer to Research Question 2 of this thesis (i.e., can a circular layout achieve an efficient and usable hands-free head-based interaction?) is that a circular layout coupled with head-based interaction can be an efficient and useable interaction for HMDs. The efficiency of our proposed interface was proved through a comparison with traditional Head+Dwell techniques and 4-day training.

The following chapter aims to answer the Research Question 3 of this thesis (i.e., are directional full-body motion-based interaction feasible and efficient for HMDs?) and address the Core Challenge 3 (i.e., efficient and feasible full-body interaction for general tasks with HMDs). It first investigates the recognition accuracy of our method and the social acceptance of directional full-body motion-based interaction, together with users' comfort ratings for each direction. Then, we optimize its design and conduct a second study to compare DMove to Hand-based interaction and hybrid-based (Head+Hand) interaction for system control tasks (i.e., menu selection).

Chapter 7 DMove: Directional Full-body Interaction for HMDs

Section 7.1 Introduction

Augmented reality (AR) allows users to interact with virtual objects that are overlaid on the physical space via see-through head-mounted/worn displays (HMDs/HWDs). Ordinarily, gestural input [67,259] is preferred to keyboard and mouse. AR HMDs have sensors that can detect head and hand movements [47,180,300]. What these sensors can also capture is body motion (e.g., moving the body forward/backward or left/right) by assuming that the position of the head is the position of the user and that users' head can move along with their body towards a certain direction. Unlike head- and hand-based gestures, body motion is underexplored and thus underutilized in current HMD systems. Body motion can present several benefits compared to hand- and head-based motion. Hand-based motion usually requires users to keep their hands in mid-air which could result in arm fatigue during prolonged interactions [195]; it can also cause inaccurate interactions (e.g., unwanted menu item selection)—for example when users' hands accidentally go off the small tracked area of HMDs. Similarly, HMDs often cause motion sickness and, when using frequent head motions, there is the risk of increased sickness [295]. With body motion, it is possible to avoid arm fatigue and to minimize motion sickness and, as shown later in our results, still allows for high accuracy of interaction and good usability ratings. Our research explores the use of directional body motion to interact with HMDs based on the accuracy of object selection, task completion time, and users' subjective feedback on workload, motion sickness, and overall usability. Our focus in this chapter is on menu item selection, but the results are applicable to other types of interaction and interface.

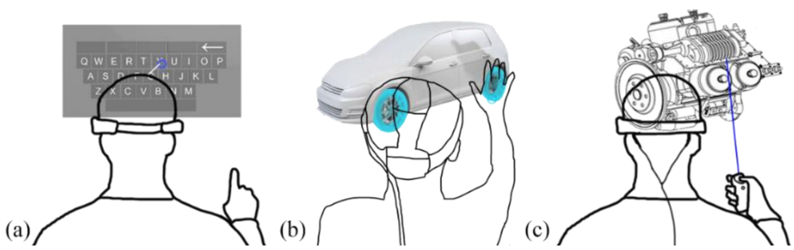


Figure 7-1. Interaction in three commercial HMDs, (a) HoloLens—Head+Hand-based (Hybrid) interaction (b) Meta 2—Hand-based interaction (c) Magic Leap 1—Controller-based interaction.

At present, there are three main commercial AR HMDs—the Meta 2 [47], Magic Leap 1 [300], and HoloLens [180]. Figure 7-1 shows how each device supports users' interaction with the virtual environment. Meta 2 allows hand-based interaction where users need to move their hand to the menu item and confirm the selection by using a hand gesture (i.e., grab). HoloLens uses a hybrid approach for menu selection, where a ray is extended from the virtual camera position towards the viewing direction and into the virtual environment. The end of the ray is akin to a cursor and users confirm a selection by a hand/finger gesture (i.e., hand-tap)—in other words, it requires users to use their head to move the cursor and their hand for selection. This research only considers device-free approaches since they are more flexible than device-based approaches and can be used in more scenarios, environments, and types of HMDs.

In this chapter, we present DMove, an approach to interact with HMDs that is hands-free, does not require handheld devices, and avoids the need to use head motions; instead, it uses directional body movements. In our approach, the system is trained to recognize the possible directional body motions around the user with two distances (Far and Close). Selection is made when the system predicts that the user has made a particular movement. Our approach only needs the sensors that already come in current HMDs, like Meta 2; and unlike Magic Leap, it does not require a handheld device. In the first study, we explore two aspects. The first deals with the feasibility and accuracy of our recognition method, and the second is about assessing users' social acceptance of directional motion-based interactions and their perceived physical and mental comfort levels in each direction. Based on the results, we then optimize our technique and, in a second study, we compare DMove with Hand-based interaction (like what is available to Meta 2 users) and Head+Hand-based interaction (akin to what users do with HoloLens). Menu selection is the chosen task because it is a common activity in all HMDs. Based on the results of the two studies, we are able to extract a set of guidelines for interfaces that are based on directional motions. Also, we present two sample applications that can leverage DMove-type apart from menu selection.

The contributions of the chapter include: (1) a motion direction recognition method that requires no additional handheld devices nor sensors for current HMDs; (2) an optimized directional motion-based interface (DMove); (3) an evaluation of three menu selection methods for HMDs; (4) a set of guidelines for applications that use directional motion-based interactions; and (5) two applications external to menu selection and that use DMove as their interaction interface.

Section 7.2 Related Work

Section 7.2.1 Device-free Interaction in AR HMDs

Mine [182] pointed out that interacting with virtual objects requires (1) a mechanism for the identification of the objects to be selected, and (2) some signal or command to indicate their selection. We next describe two commonly used device-free interactions for HMDs.

Hand-based Interaction

Hand-based interaction is one of the most commonly used selection methods in HMDs [170] because it is assumed to be natural and practical. To perform a selection of a near object [182], users first need to choose the virtual object to be selected by hovering the hand over it and then selecting it by performing a gesture—e.g., in Meta 2 [47] users select the item by making a grab gesture. To select an item that is placed further away from the user, Mine [182] suggests that users can utilize their finger to point at the object followed by a selection gesture. Studies have looked at the finger pointing [22,154], but these techniques require an additional external sensor like Kinect that is placed at a distance to detect and classify the gestures.

In general, hand-based interactions that require users to keep their hands in mid-air are uncomfortable and can be tiring, particularly for HMDs [251]. This is because users are forced to keep their hands within the small area tracked by the sensors. Inaccuracies can often occur when the hands go off the area. In addition to issues with the recognition algorithm and other technical limitations [273], mid-air hand interactions are also sensitive to users' physical abilities which can lead to unpredictable performance.

Head-Pointing

Together with hand-based techniques, head-based interaction has been actively studied in the virtual reality (VR) HMDs [32,46]. It has been widely adopted as a standard way of pointing at virtual objects without using hands or hand-held pointing devices [148]. Instead, it relies on the HMDs' built-in IMU sensors. Recent studies further have explored head-based techniques in both VR [10] and AR [148]. Like techniques based on eye-gaze, using the head may lead users to suffer the ‘Midas Touch’ [126] problem of unintentional selection because head-pointing has this same problem when confirmation of a selection is needed. Researchers have investigated solutions to this problem such as using dwell time [126,202,245,263], adopting gaze gestures [13,62,122,125], applying a second modality such as controllers [148], but these solutions are at times not ideal. For example, having a dwell time can slow performance; gaze requires additional expensive trackers but still suffers from accuracy issues; and not every HMD can track a handheld device, furthermore forcing users to hold a device prevents their hands from being used to manipulate the virtual objects in these systems.

One solution used in commercial HMDs is combining both head and hand-which is referred to as hybrid interaction, which relies on the use of the head to move the cursor to a target and hand gestures to confirm the selection, like it is done with HoloLens [180]. However, this approach still suffers from the limitations of hand-based interaction.

Section 7.2.2 Body Motion-based Interaction

Foot-based Interaction

Alexander et al. [5] suggest that foot-based interactions can be grouped into two categories based on how foot actions are mapped to system commands. *Discrete* foot gesture [49,237,288] are those that are mapped to specific tasks (e.g., locking and unlocking a mobile phone). *Continuous* gestures [105,116,201,215,219,234] are those that are mapped to tasks with a spatial component (e.g., moving in one direction in a space). Although it can add an extra dimension to users' interaction, in general the proposed techniques using users’ feet require additional external sensors. This constraint limits users to fixed environments and within the space tracked by the sensors. Because AR HMDs are meant to allow freedom of movement, the need to have external sensors is not desirable. Our approach avoids this constraint and relies solely on the sensors that already come with current commercial AR HMDs.

Full Body-based Interaction

Body motion direction-based interactions have several advantages. As our results show, they can be accurately predicted by a system that requires minimal training.

They avoid the pitfalls of hand- and head-based interaction. Body motion tends to be natural and does not force users to be in uncomfortable, unnatural positions for long periods (like hand interactions which users must keep their hands in mid-air). Also, as our results show, they do not increase motion sickness despite the need for users to make body movements.

Given their potential benefits, but without the limitations of other types of gestures, we want to explore the use of the motion-based interactions for current consumer HMDs. We also want this type of interaction to be hands- and device-free. As this research shows, our technique DMove is as fast as other methods for menu item selection and also brings a subjectively better user experience.

Besides, full body motion-based interactions can be applied to other domains (e.g., gameplay [85,213]). Further, this type of interaction can encourage physical activity in offices and homes and as such can bring health benefits to their users—e.g., just ten minutes of physical activity can help users gain cognitive and physical benefits [137]. Besides work-related applications, body motion can be used for gaming interfaces. For instance, an exergame leveraging body motion as input has the potential to be utilized to encourage physical activity, so that for example elderly users or children can do exercises in a fun way regularly at home to develop their physical strength [91,258]. At the end of this chapter, we present a sample of exergame that uses motion-based interactions.

Section 7.3 DMove

In this section, we discuss the DMove's motion recognition method and the interface for our Study One.

Section 7.3.1 Motion Recognition Method

We use machine learning to classify the user's motion direction. Instead of classifying it through movement patterns (i.e., changes in the sensors' acceleration values in three dimensions), we identify the gesture (i.e., posture at the end of the movement). This is because the former approach will not always work because some HMDs, like Meta 2, do not allow access to their acceleration data; we want to make this method suitable for all HMDs.

Next, we describe our method in detail. In three dimensions, a spatial position is defined as P = $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ (see Figure 7-2). A spatial path $\boldsymbol{\Pi}$ describes the spatial progression of movement. It is an ordered list of measured spatial positions: $\boldsymbol{\Pi} = (\boldsymbol{P}_0, \dots, \boldsymbol{P}_i, \dots, \boldsymbol{P}_n)$; where P_0 is the starting position, P_i is the position we predict the performed movement, P_n is the position where the user finishes the motion (see Figure 7-2). The values used in our analysis process are described in the following formula:

Equation 7-1 Distance moved from the starting position to a predicted position on a specific axis

$$\Delta \boldsymbol{P} x_{i,0} = \boldsymbol{P} x_i - \boldsymbol{P} x_0 \tag{1}$$

Where $\Delta \boldsymbol{P} x_{i,0}$ is the distance users moved/traveled from the starting position ($\boldsymbol{P}_0$) to the position to be predicted ($\boldsymbol{P}_i$). This formula also applies to $\Delta \boldsymbol{P} z_{i,0}$.

Equation 7-2 Moving speed on a specific axis

$$v x_{j,j-1} = \frac{\Delta \boldsymbol{P} x_{j,j-1}}{\Delta t_{j,j-1}} \tag{2}$$

Where $v x_{j,j-1}$is the current speed of the head along the X-axis. $\Delta \boldsymbol{P} x_{j,j-1}$ and $\Delta t_{j,j-1}$ are the distance and time differences between this frame and the respective last frame. This formula also applies to $v z_{j,j-1}$.

Equation 7-3 Slope calculation based on travelled distance on x- and z- axis

$$\boldsymbol{m} = \frac{\Delta \boldsymbol{P} x_{i,0}}{\Delta \boldsymbol{P} z_{i,0}} \tag{3}$$

Where m is the slope of the line from $\boldsymbol{P}_0$ to $\boldsymbol{P}_i$ in X-axis and Z-axis.

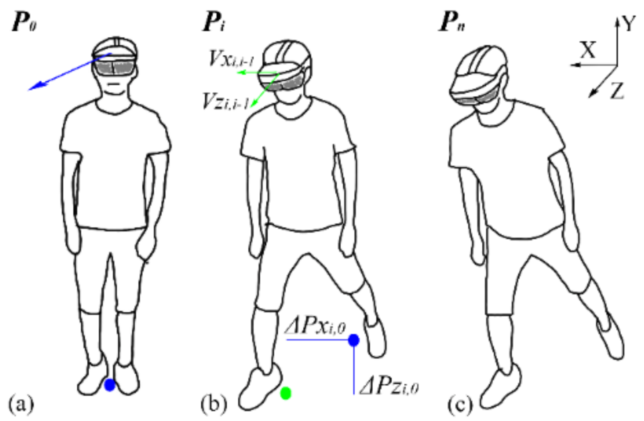


Figure 7-2. An example of a movement. (a) Starting State—A user is ready to move toward the North-East direction. The blue dot is the starting position tracked by the system. (b) Prediction State—The state used to predict the moving direction where the user has nearly finished the movement. The green dot is the end position tracked by the system. The system calculates $\boldsymbol{vx_{j,j-1}}$, $\boldsymbol{vz_{j,j-1}}$, $\Delta \boldsymbol{Px_{i,0}}$, $\Delta \boldsymbol{Pz_{i,0}}$ and then sends the results to the algorithm. (c) End State—A movement is finished.

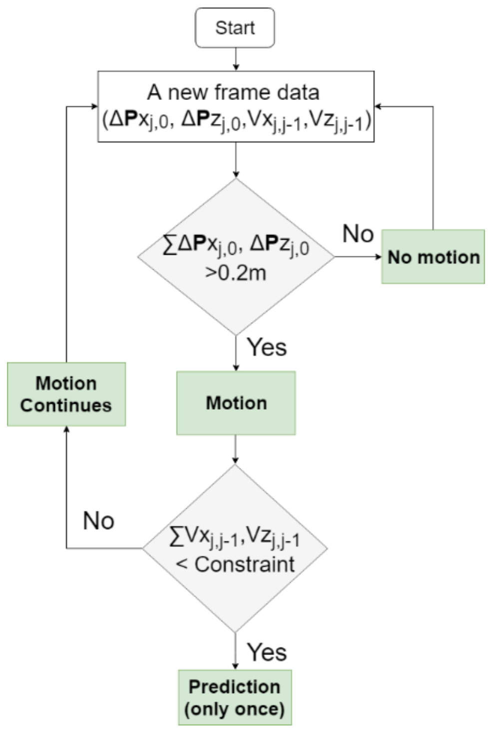


Figure 7-3. Algorithm flowchart for predicting the motion direction; we set the constraint to 0.1 m/s since it works well according to our test trials.

Classification. Tested features are $\Delta \boldsymbol{P}x_{i,0}$, $\Delta \boldsymbol{P}y_{i,0}$, $\Delta \boldsymbol{P}z_{i,0}$, distance traveled between $\boldsymbol{P}_0$ and $\boldsymbol{P}_i$, slope m. Only $\Delta \boldsymbol{P}x_{i,0}$ and $\Delta \boldsymbol{P}z_{i,0}$ are included in our dataset since the features analysis using Weka [104,279] has shown that they are the top two features and all predictions are based on them. We apply the Random Forest classifier provided by Weka for predicting the motion directions. Figure 7-3 shows the algorithm flowchart.

Section 7.3.2 Interface and GUI

We proposed two interfaces that are based on eight direction—East (E), North-East (NE), North (N), North-West (NW), West (W), South-West (SW), South (S), and South-East (SE). Figure 7-4 shows the two designs. The first design is 8-block DMove which each direction has one distance level—No Limit (we suggest at least 20 cm away from the starting position to improve the accuracy); the other is 16-block DMove which each direction has two distance levels—Close (we suggest 30 cm away from the starting position) and Far (we suggest 60 cm away from the starting position). We wanted to use two levels of the distance (Far and Close) around the user because, with two levels, the technique can have more interface items, but this may also affect the prediction accuracy of distinguishing between the two levels. To guide users visually, both interfaces are displayed in front of their view like a GUI where the tiny white point in Figure 7-4 represents the head position.

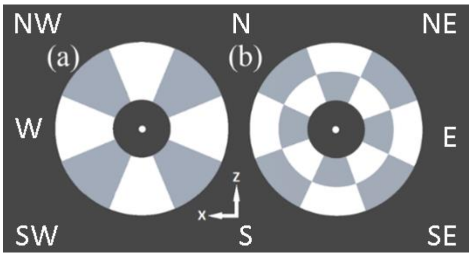


Figure 7-4. (a) 8-block and (b) 16-block DMove interface.

Section 7.4 Study One

In this study, we focused on the accuracy of our motion direction recognition technique. We also investigated the social acceptance of the motions (i.e., in front of whom users would accept to perform these motions and where) and comfort levels (mental and physical) of doing such motions.

Section 7.4.1 Participants and Apparatus

Twelve participants (four females) aged between 17 and 28 were recruited from a local university campus to participate in the study. They all had normal or corrected-to-normal vision. The study was conducted using a Meta 2 AR HMD [47] connected to a standard computer with an i7 CPU, 16 GB RAM and an Nvidia GeForce GTX 1080Ti GPU. We implemented the system in Unity3D. All experiments were conducted in a lab where users cannot be seen from outside.

Section 7.4.2 Design and Evaluation Metrics

The experiment employed a one-way within-subjects design where the independent variable was interface—16-block and 8-block. We were interested in two variables, (1) *Target Direction*—E, NE, N, NW, W, SW, S, SE; and (2) *Target Distance*—Close, Far, and No Limit. Participants were asked to do a training data collection session first for both interfaces and then do the testing sessions. The order of the interface was counterbalanced.

The evaluation metrics used were listed below.

- *Accuracy*. Accuracy was measured based on reproducibility [94] and how stable and scalable the system was against the data collected from a different session. An error was recorded when the classifier failed to predict the correct movement direction.
- *Physical and Mental Comfort*. It quantified how the users' comfort levels (both physical and mental) varied across each Target Direction × Target Distance combination. We used 5-point Likert questions to collect the data.
- *Social Acceptability*. We adopted the questionnaire from [3] to assess in which places and in front of whom users were comfortable doing the motions.

Section 7.4.3 Task and Procedure

The experiment began with the data collection session for each interface where the order of the interface was counterbalanced. The system would ask participants to perform each directional movement five times starting from N followed by the other directions in a clockwise order till the last direction, i.e., NE→E→SE→S→SW→W→NW. For the 16-block DMove, the system would ask participants to do the Target Direction × Close first then Far. For the 8- block DMove, they only needed to do the No Limit movement for each direction. They were asked to let the head follow their body movement in a natural way to help them keep their balance and their head steady. In between conditions, participants were requested to fill out the Physical/Mental Comfort questionnaire.

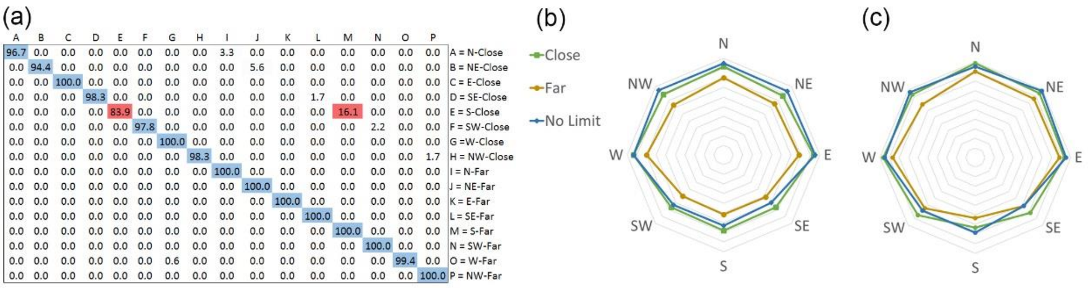
After the data collection session, they did the testing session. The order of interfaces was the same as the data collection session for each participant. However, unlike the data collection session, which had a fixed order for the direction, in this phase, the system randomized the directions. This was done to better assess the accuracy of the system and to avoid participants' muscle memory. Similar to the data collection session, participants had to reach each direction five times.

At the end of the experiment, participants completed the social acceptability questionnaire. The whole experiment lasts around 30 minutes for each participant.

Section 7.4.4 Results

Accuracy

We used 2880 instances collected from the training session to train the model and used another 2880 instances from the testing session to test it. The accuracy, precision, recall, F-Measure for 8-block DMove were all 100% while for 16-block were 98.06%, 98.2%, 98.1%, 98.0%, respectively. As can be observed from the red blocks of the confusion matrix in Figure 7-5a, most of the wrong predictions were in South Close where our recognition method predicted South Close as South Far.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
96.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	A = N-Close
0.0	94.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6	0.0	0.0	0.0	0.0	0.0	0.0	B = NE-Close
0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	C = E-Close
0.0	0.0	0.0	98.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	D = SE-Close
0.0	0.0	0.0	0.0	83.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.1	0.0	0.0	0.0	E = S-Close
0.0	0.0	0.0	0.0	0.0	97.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	0.0	0.0	F = SW-Close
0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	G =W-Close
0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	H = NW-Close
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	I = N-Far
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	J = NE-Far
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	K = E-Far
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	L = SE-Far
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	M = S-Far
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	N = SW-Far
0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	O = W-Far
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	P = NW-Far

Figure 7-5. 16-block DMove Confusion matrix (a). Comfort ratings for each direction for Physical (b) and Mental (c).

Subjective Feedback

The collected data were analyzed using a two-way repeated measures ANOVA with two factors (1) Target Location and (2) Target Distance. Bonferroni corrections were used for pairwise comparisons. For violations of sphericity, we used a Greenhouse-Geisser adjustment for degrees of freedom.

Physical Comfort. Figure 7-5b shows the Physical Comfort ratings of each direction for Target Distance. An ANOVA showed significant effects of Target Direction ($F_{3.029,16.737}$ = 11.130, p < .001) and Target Distance ($F_{1.860,20.458}$ = 13.899, p < .001) on Physical Comfort. However, no significant interaction effect of Target Direction × Target Distance ($F_{14,154}$ = 1.076, p = .383) was found. For Target Direction, post-hoc pairwise comparisons revealed significant differences between N-SW, E-SE, E-S, E-SW, SE-W, SW-W, SW-NW (all p < .05). It also yielded a close significant difference between N-SE (p = .073), N-S (p = .076), SE-NW (p = .053), and S-W (p = .063). For Target Distance, pairwise comparisons revealed a significant difference between Close and Far (p = .001), No Limit and Far (p = .005), but not between Close and No Limit (p = 1.000).

Mental Comfort. Figure 7-5c shows the Mental Comfort ratings of each direction for Target Distance. An ANOVA yield a significant effect of Target Direction ($F_{2.420,26.619}$ = 17.492, p < .001) and Target Distance ($F_{2,22}$ = 8.305, p < .05) on Mental Comfort. However, there was no significant interaction effect of Target Direction × Target Distance ($F_{4.032,44.355}$ = 1.868, p = .132). For Target Direction, post-hoc pairwise comparisons revealed significant differences between N-SE, N-S, NE-SE, NE-S, NE-SW, E-SE, E-S, E-SW, SE-W, SE-NW, S-W, S-NW, SW-NW (all p < .05). For Target Distance, pairwise comparisons revealed a significant difference between Close and Far, No Limit and Far (both p < .05) but there was no significant difference between Close and No Limit (p = 1.000).

Social Acceptability. Participants' overall feelings during the task were rated 4.5 out of 6 (s.e. = 0.195). We calculated the acceptance rate for each given audience and location using the percentage of participants who selected each audience/location in their answers (see Figure 7-6). A Cochran's Q test showed a significant difference between audiences ($\chi^2(5)$ = 20.606, p < .001). Post-hoc McNemar tests (Bonferroni: α-levels from .05 to .004) showed that the acceptance rates for strangers were significantly lower than if participants were alone (p < .004). Also, participants' responses suggested that the location would influence their willingness to use directional motions. A Cochran's Q test showed a significant difference between locations ($\chi^2(6)$ = 39.368, p < .001). Post-hoc McNemar tests (Bonferroni: α-levels from .05 to .004) showed that the acceptance rates for using DMove at home was significantly higher than at a shop or other public places, and on sidewalks (all $p < .004$).

Section 7.4.5 Discussion

Direction Motion-based Interface

Our method showed very good accuracy for identifying the users' movement direction in both 8- and 16-block DMove interfaces. The reason was that the attributes used in our dataset clearly distinguished the movement directions (see Figure 7-7). Participants' subjective feedback indicated that motions toward the South direction led to both physical and mental discomfort. During the experiment, we also observed that each participant had his or her own predisposed way of making directional movements due to their physical attributes—e.g., taller users were able to take a longer step than the shorter users. As such, we believe that using a user's own motion data will likely increase prediction performance because it will consider the physical characteristics of each participant.

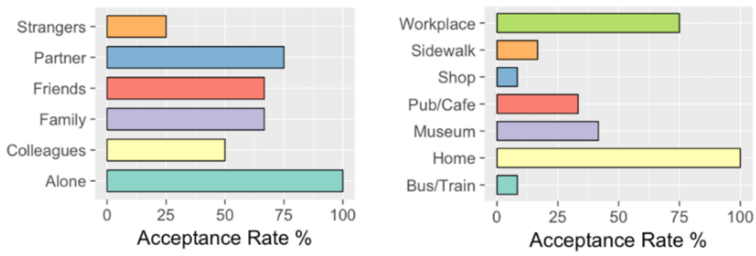


Figure 7-6. Acceptance rates for different audiences (a; left), and locations (b; right).

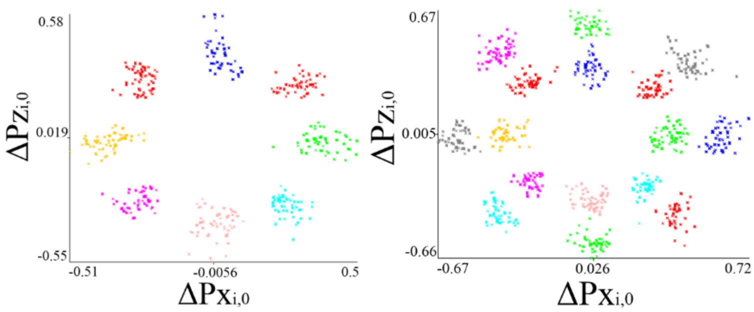


Figure 7-7. 8-block (a; left) and 16-block (b; right) DMove's plot image of $\Delta \mathbf{Px_{i,0}}$, and $\Delta \mathbf{Pz_{i,0}}$, where each color represents a movement direction.

Social Acceptability

According to the results of the social acceptability questionnaire, most participants were quite positive towards a DMove-based interface; only one participant gave a low rating of 3. They were willing to do directional motions alone or in front of familiar people (see Figure 7-6a). They preferred private spaces (such as their home and workplace) rather than public areas (see Figure 7-6b). Based on this feedback, we suggest that a DMove-type of interface should be used in in-door scenarios (i.e., home or workplace) and in front of people familiar to the user.

Optimization

Based on the performance and subjective feedback, we decided to work further with the 16-block interface and optimize it. Since users have difficulty moving towards the S direction, we decided to make some adjustments to S and also SE and SW directions. We removed S and combined the 2-levels SE and SW directions into one single direction each. In this way, users could easily move towards these two (now much larger) directions. After these changes, the DMove interface had 12 items (Figure 7-9b).

Section 7.5 Study Two

In the second study, we explored the use of DMove for menu selection, a very common activity in HMDs. We compared the performance, suitability, and usability of DMove with two device-free interaction methods, Hand-based and Hybrid (Head+Hand), for menu selection because they represent two of the most common, and available ways for selecting menu items in current AR devices. Similar to Study One, we also measured workload, motion sickness, and user experience of the three methods. We only considered device-free approaches because they are applicable to a wider range of scenarios, and types of HMDs.

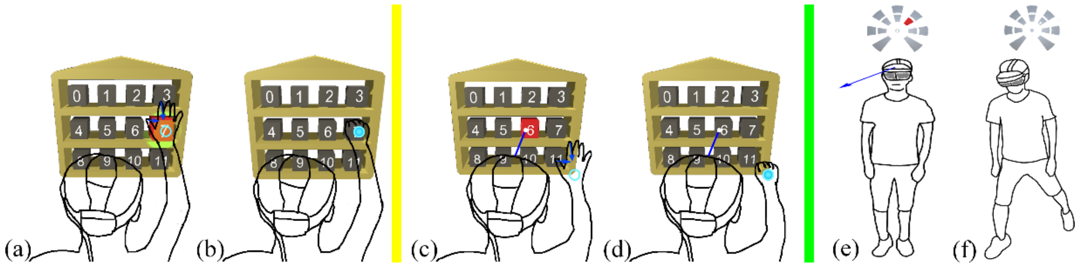


Figure 7-8. Using Hand, Hybrid, and DMove to select an item from the menu. (1) Hand (a) A user needs to move the hand to the target and hover it, (b) and then performs a close palm gesture to select it. (2) Hybrid (c) A user needs to rotate the head to move the cursor to the target, (d) and then performs a palm closing gesture to select it. (3) DMove (e) A user needs to go the NE direction, (f) a selection is made when the user (nearly) completes the action.

Section 7.5.1 Evaluated Conditions

We evaluated the following three Selection Methods for menu selection:

1. *Hand-based interaction (or simply Hand)*. This was similar to what Meta 2 would provide. To select an item in a menu, a user had to move the cursor controlled by one hand in mid-air to hover it on the item and then make a palm closing gesture to confirm its selection. Figure 7-8(1) shows this scenario. Visual feedback, in the form of extra green light and enlarged item, was provided to indicate whether the hand was correctly positioned on the item. A sound would be played to confirm the selection. We modified the code from one of the sample demos provided by Meta Company, the developers of the Meta 2.
2. *Hybrid-based interaction (or simply Hybrid)*. This was analogous to how menu selection was done in HoloLens, where a user had to move the head to control a cursor and position it on an item—selection was confirmed by a hand gesture. The HMD would track the head motions casting a ray to the virtual environment. The end of the ray was akin to a cursor, which served as visual feedback. Hand detection cursor was provided to inform the user of the cursor's state. A sound would be played when a selection was made. Figure 7-8(2) shows an example of this approach.
3. *Directional Motion-based interaction (DMove)*. In this condition, a user had to move their body with one foot towards a direction location that represented a menu item. For any motion performed, the classifier would return the direction and block. A cursor presenting the user's position was provided on the HMD as visual feedback and a sound would be played if a selection was made. Figure 7-8(3) shows an example of how a user would select the NE item.

We designed the menu items (see Figure 7-9) based on official design guidelines [299], which suggested that they should be located at around 0.5m away from the user. However, regardless of this, the users could still adjust the position between them and the menu items to a comfortable distance before the start of the experiment. We used grid menu layout for Hand and Hybrid interaction because both HoloLens and Meta 2 have applications that rely on this type of layout. For example, the developers of Meta 2 provided guidelines and an official application using a grid layout—we followed the guidelines and adapted the application for this experiment. We did not use the grid layout for DMove because it did not represent a natural mapping for around body interactions. Our choice of radial layout was based on feedback from a pilot study and also from previous research [110,153].

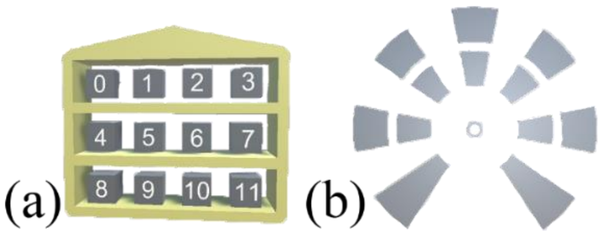


Figure 7-9. (a) Hand/Hybrid—Meta 2 Workspace-like menu interface, and (b) final-DMove interface—optimized based on the 16-block layout with S removed and had one single larger area for SE and SW directions (only 1 level due to users' discomfort with two levels).

Section 7.5.2 Participants and Apparatus

Eighteen participants (six females) aged between 17 and 28 were recruited from the same local university campus as in Study One. They all had normal or corrected-to-normal vision and were right-handed. To avoid biases, none of these participants did Study One. This experiment used the same apparatus and lab location as Study One.

Section 7.5.3 Experiment Design, Task, and Procedure

The experiment followed a 3 × 2 within-subjects design with two factors: Selection Method (Hybrid, Hand, and DMove) and Menu Size (Large—same size as Meta 2 Workspace, and Small—80% of the Large). The combinations of Selection Method × Menu Size were counterbalanced. The whole experiment lasted about one hour for each participant. Before the trials started, the participants were asked to complete a pre-experiment questionnaire to gather demographic information and were informed of the purpose of the study. Since Study One suggested that using the user's dataset could help improve recognition accuracy, we collected data from each user before the first testing session to train our system. This data collection session was conducted in the same way as in Study One but with fewer directions and took just around 2-4 minutes. To balance the conditions, participants were also given up to five minutes of training with both Hand and Hybrid interactions. When participants felt rested and ready, they would proceed to the testing session.

In each session, each block (representing a menu item) would randomly appear once, one by one, for a total of five times. After each session participants completed three questionnaires: NASA-TLX [107], user experience [149], and motion sickness assessment (MSAQ) [87]. We instructed participants to maintain their head steady and in a comfortable position whenever possible. In the end, we asked them to provide comments on each of the interfaces. The experiment returned 3 (Selection Method) × 2 (Menu Size) × 12 (blocks) × 5 (times) × 18 (participants) = 6480 trials.

Section 7.5.4 Results

We analyzed the data using a two-way repeated measures ANOVA with two independent variables, Selection Method (Hand, Hybrid, DMove) and Menu Size (Large and Small). Bonferroni correction was used for pairwise comparisons, and Greenhouse-Geisser adjustment was used for degrees of freedom for violations of sphericity.

Task Performance

Figure 7-10 presents the task completion time and error rate among the six layouts. For task completion time, the ANOVA test yielded no significant effect of Selection Method ($F_{1.197,20.341} = 2.555$, $p = .121$), Menu Size ($F_{1,17} = 1.108$, $p = .307$), and Selection Method × Menu Size ($F_{1.219,20.715} = 1.177$, $p = .303$), which showed that the completion time for each Selection Method was equal. For error rate, there was a significant main effect of Selection Method ($F_{1.506,25.610} = 14.138$, $p < .001$), but no significant main effect of Menu Size ($F_{1,17} = .524$, $p = .479$) and no significant interaction effect of Selection Method × Menu Size ($F_{1.940,32.980} = 2.069$, $p = .144$). Post-hoc pairwise comparison revealed a significant difference between Hand and

Hybrid, Hand and DMove (both $p < .05$); this meant that hand had higher error rates than Hybrid and DMove. There was no significant difference between Hybrid and DMove.

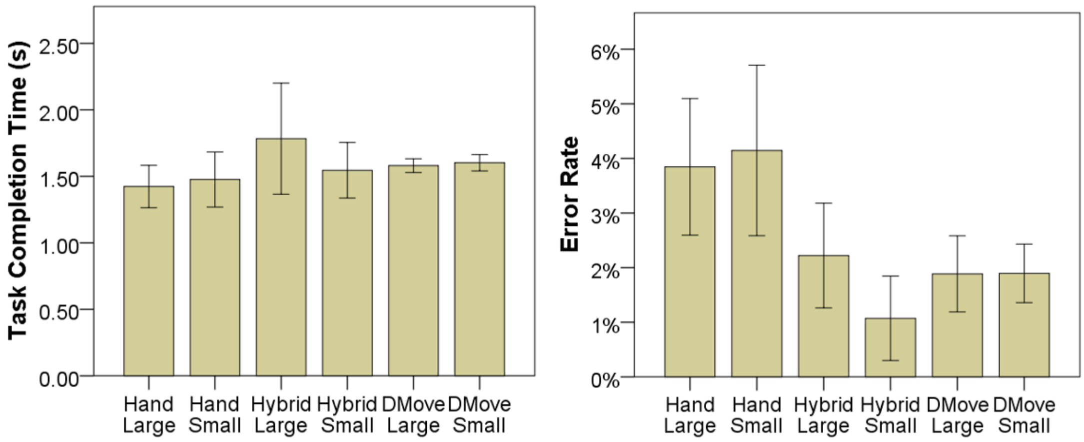


Figure 7-10. Mean task completion time (a; left) and error rate for the six layouts (b; right). Error bars indicate ±2 standard errors.

NASA-TLX Workload

For overall workload, DMove Large was rated the best ($M = 36.63$, $SD = 17.07$) and Hand Small ($M = 47.80$, $SD = 21.13$) was rated the worst. ANOVA tests yielded a significant effect of Selection Method ($F_{1.514,25.732} = 4.676$, $p < .05$), but not of Menu Size ($F_{1,17} = 2.806$, $p = .112$) and Selection Method × Menu Size ($F_{2,34} = .211$, $p = .811$). Post-hoc pairwise comparisons revealed a significant difference between Hybrid and Hand, DMove and Hand (both $p < .05$; see Figure 7-11a).

Regarding NASA-TLX workload subscales, ANOVA tests yielded a close significant effect of Selection Method ($F_{2,34} = 2.947$, $p = .066$) on Mental; a close significant effect of Selection Method ($F_{2,34} = 2.927$, $p = .067$) on Temporal; a close significant effect of Selection Method ($F_{1.544,26.240} = 3.533$, $p = .054$) on Frustration; and a close significant effect of Selection Method ($F_{2,34} = 3.094$, $p = .058$) on Effort. No other significant or close significant effects were found.

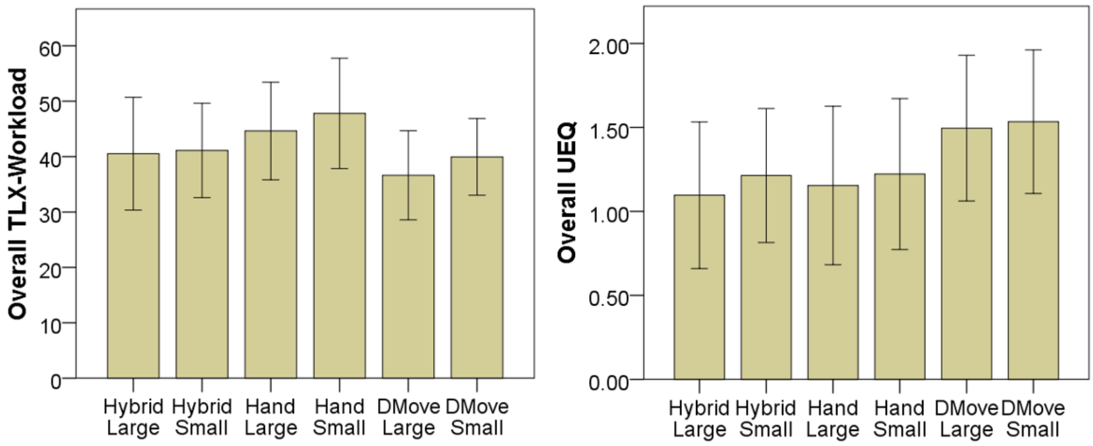


Figure 7-11. Overall NASA-TLX workload (a; left) and overall UEQ scores among all 6 layouts (b; right). Error bars indicate ±2 standard errors.

User Experience

The score for UEQ was adjusted between -3 (very bad) to 3 (excellent). Figure 7-11b shows the overall UEQ score among the six layouts. ANOVA tests yielded a significant effect of Selection Method ($F_{2,34} = 6.371$, $p < .01$), but not of Menu Size ($F_{1,17} = 2.498$, $p = .132$). No significant interaction effect was found on Selection Method × Menu Size ($F_{1.350,22.956} = .202$, $p = .730$). Post-hoc pairwise comparisons showed a significant difference between Hybrid and DMove as well as Hand and DMove (both $p < .05$).

Regarding the UEQ subscales, ANOVA tests yielded a significant main effect of Selection Method ($F_{2,34} = 6.167$, $p < .01$) on attractiveness. The pairwise comparison indicated DMove was more attractive than both Hand and Hybrid (both $p < .05$). There was a significant effect of Menu size ($F_{1,17} = 6.115$, $p < .05$) on stimulation. Post-hoc pairwise comparison showed Small Menu bought more stimulation from users than Large Menu ($p < .05$). No other significant effects were found. DMove outperformed Hand, Hybrid across the UEQ subscales (see Figure 7-12).

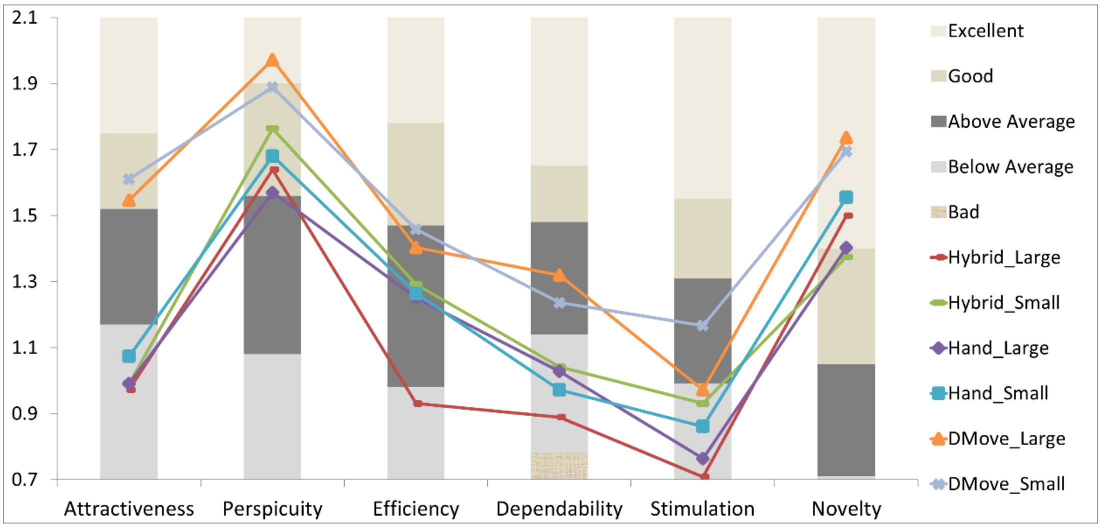


Figure 7-12. UEQ ratings for all 6 layouts with respect to benchmarks.

Motion Sickness

For the overall sickness score, DMove Small was rated the worst (M = 19.29%, SD = 12.55%) and Hand Small was rated the best (M = 16.59%, SD = 8.73%). ANOVA tests yielded no significant effect of Selection Method ($F_{1.207,20.521} = 2.860$, $p = .100$), Menu Size ($F_{1,17} = 1.569$, $p = .227$), and Selection Method × Menu Size ($F_{1.390,23.626} = 1.224$, $p = .297$) on overall motion sickness. Regarding MSAQ subscales (gastrointestinal, central, peripheral, sopite-related), ANOVA tests yielded a significant effect of Selection Method ($F_{2,34} = 4.265$, $p < .05$) on peripheral and a close significant main effect of Selection Method ($F_{1.149,19.532} = 4.022$, $p = .054$) on central. No other significant effects were found. Post-hoc pairwise comparisons showed no significant effect between any Selection Method on peripheral.

Section 7.6 Discussion

In this section, we discuss the reasons why DMove is a strong candidate interface for menu selection based on users' performance and experience for the current AR HMDs.

Section 7.6.1 Task Performance

The results indicated that Hand, Hybrid, DMove have equal selection time, while Hybrid and DMove had lower error rates than Hand. We observed that the high error rate in Hand was due to wrong selection of the item that was next to the intended targets. Although visual feedback was provided (by expanding the size and adding additional highlight color) for the item that the users' hands were currently hovering on, the system's detection time for whether their hands were on the virtual item was slow (1-2 seconds). Faced with this, users chose to trust their spatial knowledge and performed the selection gesture which was often incorrect, and this led to higher error rates. This was not the case for Hybrid which the users' hands were only used to perform a gesture to confirm a selection. Interestingly, we found that Menu Size had no effect on task performance. This might have been because the difference between Large and Small was not big enough to cause a significance. Based on performance alone, we suggest Hybrid and DMove should be considered before Hand for current AR HMDs.

Section 7.6.2 User Preference

NASA-TLX Workload. Regarding the overall workload, Hand was worse than Hybrid and DMove. One reason why participant felt that the overall workload was higher for Hand was that to use it well they had to focus very carefully to gauge where the items were located and the location of the virtual cursor. This process was tiresome. Although there was no difference in physical workload among three methods, participants had arm fatigue in both Hand and Hybrid—several of them said it was too difficult and tiring to keep their hand for long periods or to perform the hand gesture repeatedly to make a selection. In contrast, for DMove there was no need to exercise the visual focus required in Hand because they could rely on their spatial awareness of the location of the items around them to make a quick motion for their selection. So, users should avoid using the Hand approach if they consider workload to be a crucial factor.

Motion Sickness. Our results indicated that performing directional movements in DMove did not result in a higher motion sickness than selecting menu items via Hand and Hybrid. Thus, in terms of motion sickness, we believe DMove was as comfortable as Hand and Hybrid.

User Experience. ANOVA tests showed that DMove provided a better user experience than Hand and Hybrid. As mentioned earlier, we considered Hand and Hybrid because they were used in current the AR HMDs and presumably were thought to be usable.

Our results showed that only DMove was rated above average to excellent while Hand or Hybrid was rated much worse. Although our data samples were not sufficient enough to compare with the benchmarks [235], they still provided a sense of how much more usable DMove would likely be when compared to the other two interfaces. In summary, using DMove results in better user experience than Hand and Hybrid, and if users regard usability and user experience as the most important factors, DMove is the recommended choice.

Section 7.6.3 User Comments

According to Bowman et al. [33], natural interactions (like Hand in our study) provide little additional productivity but actually can make the task more complicated and unnecessarily cumbersome. Hand interaction not only caused some physical discomfort and arm pain (*P7*: *"*my arms are sores after a while*"*) but participants did not like it because of the lack of tactile feedback (*P10*: *"*It feels empty when I use my hand to select the virtual objects, because I don't sense when the action is finished*"*). Physical issues are not easy to solve—the only way is to ask users to rest. The tactile feedback issue could be solved by using a haptic glove. However, it is expensive. In the case of Hybrid interaction, participants seem generally happy with its task performance, but it seems to be bored and may also cause issues like arm muscle tiredness and pain (*P3*: *"*In the end, I felt a bit sleepy and my arms get tired fast*"*). On the other hand, participants have found DMove interesting and very easy to use. Participants suggested that we develop an exergame (like [213]) based on DMove, as eloquently put by *P9*: *"*[DMove] is fun, I would recommend using it as an exergame, it's good for health*"*.

Section 7.6.4 Design Guidelines for DMove Interactions

Guideline 1: Cater to Individual Differences

Based on our findings from Study One, DMove should use an individual's dataset to maintain (100% or close to 100%) accuracy to take into account each user's height, weight, movement speed, and step distance. To account for these factors and to prevent poor accuracy, DMove for general users should be calibrated according to individual physical features and abilities. Besides, we predict a motion just right before a user finishes it by comparing the head movement speed with a pre-set constraint, which should also be tuned to suit the individuals. As our second study show, training the system is easy and fast and needs to be done only once.

Guideline 2: Flexibility, Efficiency of Use, Customizability

The comfort ratings from Study One suggests that the Close level is much easier to reach, and it does not cause discomfort, while directions that users can see—N, NE, NW, E, W are much easier to perform. As such, we suggest putting frequently used items/functions in Close directions and avoid putting them at the directions that users cannot see easily to increase efficiency and usability.

Guideline 3: Not in Front of Strangers and Public Venues

Based on the social acceptance results from the Study One, we recommend using DMove for indoor scenarios such as at home/work environment (or outdoor but when there is nobody around). In addition, we suggest that an interface based on DMove should be used in front of the people users are familiar with instead of strangers.

Guideline 4: Provide Feedback and Keep Consistency with Other Interfaces

Results from Study Two point out two advantages of DMove over Hand and Hybrid. On the one hand, DMove provides users actual tactile feedback when they select an item/function because when placing the foot on the ground they will receive immediate and clear feedback. On the other hand, DMove is an interface that can be considered eyes-free because users can use their spatial awareness and memory to remember where the items are around them. Although it can be eyes-free, we suggest that the menu should always appear as a simple non-obtrusive visual interface on the HMD on-demand, similar to a context menu, whenever users want to use it and so that they do not have to memorize the items of the menu. Similar to what we have done in this research, we suggest that the interface shows the user's movement location—e.g., a simple visual cue like a dot can be used to indicate to which direction they are moving. Visual and/or audio feedback can be included to tell them that a selection has been successfully made.

Section 7.7 Sample Applications

In this section, we present two applications where DMove can be used for not only AR HMDs but also possible for VR/MR HMDs.

Section 7.7.1 Remote Control of an Environment

We developed a prototype application (Figure 7-13) to remotely control electrical appliances and devices in an environment (i.e., home/workplace). There are existing methods for controlling home appliances via voice or a smartphone; however, such methods have limitations—they either are affected by ambient noise [115] or require users to have access to an additional device. DMove does not have any of these limitations. Users can use it to control smart IoT-linked devices such as a TV, lights, air condition, with a DMove-type interface. For instance, when using an AR HMD, a user realizes that the light in the room is too dark (Figure 7-13a), then he/she can take a small step forward, to turn the light on (Figure 7-13b). Further, the user is not limited to turning devices on/off only but can also to interact with a smart TV, for instance, to switch channels by taking a small step leftward and staying at “-” icon to continuously change the channels until the TV shows the desired one. If the items are not in the current interface, users can add a new item and customize its function.

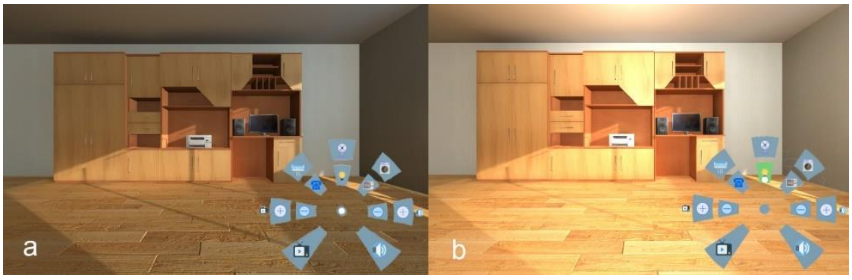


Figure 7-13. An example of a smart environment remote control using an AR HMD; a user realizes the environment is dark (a; left) so he/she uses remote control to switch the lights on (b; right).

Section 7.7.2 Dance Exergame

Our second prototype application is a dance exergame, which can be accessed and played via a DMove-type interface. Such a game can be helpful for users of all ages to entrain themselves while doing exercise and in the process to improve their health [23,112,241,246,283,284]. The game starts with the system randomly activating some blocks (see Figure 7-14). To deactivate a block successfully, the user needs to perform the corresponding directional motion within a time period, which can be adjusted based on difficulty levels. If the user fails to move and tap on the blocks before the time limit expires, the user cannot get points, which are needed to move to other levels. To avoid motion sickness, we allow users to set a time limit per round of gameplay (e.g., about 3-5 minutes akin to the length of a typical song). To make the game suitable for the elderly, one can follow the recommended guidelines provided (e.g., in [85]). In addition, the game can be multiplayer based and be played with friends via an online platform, so it could potentially bring in a social component into the gameplay. Overall, our second prototype is a dance exergame that can be played in an office or home environment with an AR HMD and potentially for a VR system as well.

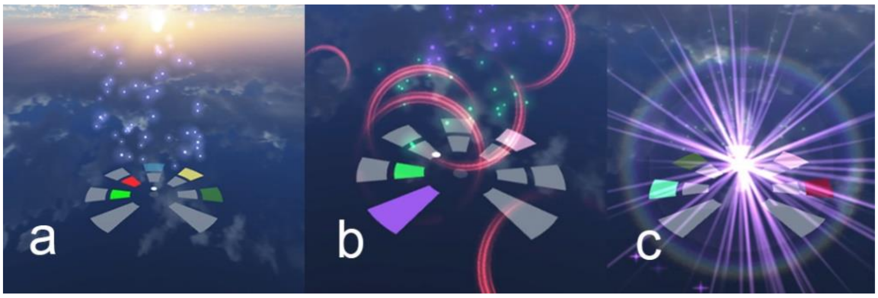


Figure 7-14. An example of a dance game; the DMove interface (a; left), the particle effect when a correct movement is made (b; mid), (c; right) to help users engage with the game.

Section 7.8 Limitations and Future Work

Although DMove does not cause arm and neck fatigue, repeated use in a long period may cause some degree of tiredness in the user's leg or body. On the other hand, the AR HMDs are commonly used by users in standing position. Also, as indicated earlier, standing and moving one's body is often encouraged in today's sedentary society—e.g., standing rather than only sitting while typing. As such, DMove may offer extra benefits in the form of physical activity.

As stated earlier, we have selected the grid menu for Hand and Hybrid interactions based on example applications used in two current AR HMDs. It can be argued that their layout or the items can be made smaller so that they can fit better in the common small field-of-view of AR HMDs or allow faster selection. However, there is usually a tradeoff between smaller menu items and hence smaller layout on accuracy. Our research has not been focused on exploring the ideal size of menu items and this could be a possible line of research to help us develop techniques that require Hand or Hybrid selection of items.

There are several paths to further strengthen DMove. (1) The levels in one direction can be increased to allow for more items. This may be useful because, although the number of items in the radial menu is large enough to meet the needs of applications in AR systems, there can be cases which a large number of items are needed. As such, having more levels will allow more items to be included. (2) It is possible to optimize the layout further—e.g., finding the most suitable distance for each level in one direction instead of pre-defined values (i.e., 30cm) that we used in our study. (3) Since we want DMove to be accessed on-demand, future work can also focus on exploring ways to separate DMove from ordinary moving. We have done some preliminary explorations and one way that is possible for all commercial AR HMDs, for instance, is to use the third dimension (Y-axis) where users can perform an on tiptoe (up/down) action to wake up the DMove. This way, DMove can also be suitable for users with arm/hand disabilities as it does not require hands or any input device.

Section 7.9 Conclusion

In this chapter, we have presented DMove, a device-free and hands-free directional motion-based interaction for head-mounted displays (HMDs) that can be used for a range of applications including menu selection, remote control, and exergame. We first propose a method that can be used for recognizing directional movements in HMDs that does not need any additional external trackers. Then, we conduct a study to examine the accuracy of the proposed method for 8- and 16-block interfaces and also to understand their social acceptability and physical/mental comfort. We then optimize the interface based on findings from the first study and conduct a second study to compare the menu selection performance of DMove with Hand and Hybrid (Head+Hand) approaches.

We have found that (1) Our proposed recognition method is very accurate—100% for 8-block DMove and 98.06% accuracy for 16-block DMove; (2) Users prefer to use DMove in front of familiar people and indoor scenarios (like their home or office); (3) Users felt more discomfort when moving towards directions that they cannot see; (4)

DMove has an equal task completion time as Hand and Hybrid and a lower error than Hand when using a current consumer HMD; and (5) DMove is preferred by users because it has low workload but high usability and novelty.

Based on our results, we list several design guidelines including allowing for customization due differences in users' physical features, placing frequently used items near the user and in the frontal directions, and offering visual and/or auditive feedback—no additional tactile feedback is needed because DMove inherently comes with it, as users can feel when their foot touches the ground.

Section 7.10 Summary

Based on the above findings, we can now answer Research Question 3 of this thesis (i.e., are directional full-body interaction feasible and efficient for general tasks with HMDs?): directional full-body interaction is a feasible and efficient interaction technique for general tasks with HMDs. It is feasible since users are highly receptive to its use and it did not cause a higher sickness than other motion-based interactions. Regarding efficiency, it outperforms Hand-based interaction and is comparable to Head+Hand interaction.

The following two chapters aim to answer the Core Challenge 4 (i.e., accessible full-body interaction for applications in HMDs) and Research Question 4 of this thesis (i.e., will HMDs affect users experiencing full-body interaction?) and Research Question 5 (i.e., will sickness mitigation factors in other contexts works for full-body interaction). It first investigates the effect of tasking mode on full-body motion-based exergame and then explores the effect of viewing perspective on full-body motion-based exergame.

Chapter 8 Assessing the Effects of Tasking Mode in Full-body Motion-based Exergames

Section 8.1 Introduction

Physical inactivity has been identified as the fourth leading cause of death worldwide [151]. In recent years, the idea of using exergames (i.e., video games that are also a form of exercise) to enhance people's health has been promoted by researchers and medical practitioners. Prior studies [15,16,73,185,240] have shown that exergames can increase enjoyment and intrinsic motivation compared to conventional exercises and as such, they can be effective in promoting physical and mental health [207,224].

People are often challenged when attempting to simultaneously accomplish multiple tasks (multi-tasking) due to limitations of how we process information [63]. In the context of games, this challenge can promote users to play them. Since exergames are often used to enhance people's health, researchers have looked at the use of multi-task physical activities as a way to achieve this in different population groups (e.g., elderly [7,44]).

Recently, more and more researchers have assessed the use of Electroencephalography (EEG) to analyze players' physiology feelings and cognitive activities during the gameplay to help to provide a better gaming experience. One of the first studies to deal with games and EEG is [225], their research defines events during gameplay and analyzed the Event-Related Potential (ERP) of the brain when those events are performed. More recently, Monteiro et al. [188] investigated the effect of viewing perspective on players' Arousal-Valence and Focus level. Nacke [193] studied how the use of different kinds of controllers influences the brain during gameplay.

Researchers have also investigated full-body motion-based exergames (e.g., [85,86]). This research has been primarily conducted with common flat displays such as large-screen TV that are placed at some distance for the gamers. Virtual reality (VR) allows a greater degree of immersions and there is a recent trend to use VR for exergames—for example for athletic training [242], fitness training [301], and High-intensity interval training [17]. Although there are a growing number of VR exergames in the market, there is limited research on the feasibility and effect of such games. The advantage of VR is its ability to immerse users in the environment and afford full-body motions. Most of the exergames explored in the recent literature are based on a stationary setting (e.g., on a cycling bike [17]) and to our best knowledge, no study has been done on investigating full-body motion-based exergames in VR, especially focusing on their feasibility and cognitive effects elicited during gameplay.

In this research, we have developed a multi-tasking motion-based video game called KIMove. The game combines the advantages of multi-tasking [7,19] and exercises [39,272] to understand the feasibility of playing full-body motion-based exergames in VR, the effect of multi-tasking on the gamers and the type of responses elicited in players' brains using EEG data collected during the experiment.

Section 8.2 KIMove

To study the effect and gameplay experience of single- and multi-tasks involving hands and feet in VR, we implemented KIMove, a game that was inspired by Beat Saber and Fruit Ninja. The game was implemented in Unity3D and written in C#. It uses Microsoft's Kinect to capture full body motions. We had two versions, one for VR and the other for Large Display, which served as the baseline condition.

The gameplay consisted of performing hand motions in mid-air and foot movements in the form of stepping on the ground through three minutes of game time. There were two types of game objects. Fruits would appear in mid-air for users' hands to hit them, while rectangular prims or cubes would show up on the floor for their feet to step on them.

Objects would appear close to the player and move in a straight line, passing in front (like apples and pears) or going towards (yellow prims) the player. The player's hand and feet had colored balloons attached to them (red, green for each arm and yellow for the legs). The different colors were used to allow fast differentiation of the limbs and also to link the objects to the limbs that should be used to catch and destroy them. The score was given when players successfully eliminate (i.e., catch) the game objects.

We used the Kinect for motion capture and designed the game to be playable at about two meters away from the device which was required for tracking user's limb movements. A door frame was designed as a reminder for the users to be aware of the playing area in the virtual world. Figure 8-1 shows that a player is lifting the left arm to catch the apple while Figure 8-2 presents an example of the player stepping left-wards to stop the foot game object.

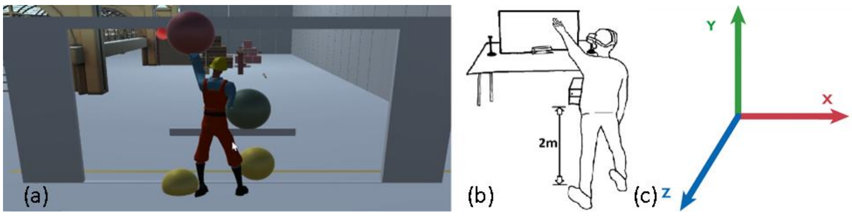


Figure 8-1. A user is trying to kill the arm object (apple) by using the left hand. (a) In a virtual view. (b) In a real-life view. Fruits are passing in front (x-direction) the user during the game. (c) Axis system used by the game.

Figure 8-2. A user is trying to destroy the foot object (cubes) by stepping the left foot on it. (a) In a virtual view. (b) In a real-life view. Cubes moving towards the user in the z-direction during the game. The red line indicates the movement area (3m long) that required throughout the game.

The game has two different game modes: Single-tasking and Multi-tasking. For single-tasking, the game spawns one object in every five seconds so that only one object moving at a time during the game. For multi-tasking, it would present to players multiple concurrent objects to be destroyed by both feet and arms in every five seconds. This means that players were required to perform two tasks in rapid succession and sometimes in parallel. All game objects have the same speed which was 0.2 m/s. These values were chosen after a preliminary study.

Section 8.3 Experiment

Section 8.3.1 Participants and Apparatus

Twelve participants (three females) between the ages of 19-29 (M = 22.42) were recruited from a local university campus to take part in this experiment. Five of them had experience on VR but were all infrequent VR users. We used an Oculus Rift CV1 as our VR device and a 50-inch 4K TV as our Large Display device. Both devices were connected to a standard computer with an i7 CPU, 16GB RAM, and a GeForce GTX 1080Ti GPU. The brainwave signals Alpha (8-14 Hz), Beta (14-30 Hz), Theta (4-8 Hz), Delta (1-4 Hz), and Gamma (30-50 Hz) were measured and collected by the MUSE headset edition 1. A Kinect was used to capture the players' movements.

Section 8.3.2 Experiment Design, Task, and Procedure

To understand the feasibility of playing the exergame in VR HMDs, we conducted an experiment using 2 × 2 within-subjects design. There were two independent variables: (1) Game Mode—Single-tasking and Multi-tasking, and (2) Display (or Device) Type—VR and TV. The order of Game Mode × Display Type combinations was counterbalanced in the experiment. Nacke [193] have shown that playing games with different types of controllers could affect brain activity differently. We were interested in whether Game Mode and Display Type have a similar effect on brain activity.

Before the experiment started, the participants were asked to complete a pre-experiment questionnaire to gather demographic information and were informed of the purpose of the study. Before each session, the participants were taught the game rules and were asked to calibrate the position in the game, and then they were asked to play a 1-min warm-up round to familiarize themselves with the game. Once the warm-up round finished, participants were asked to wear and calibrate the EEG device with the help from a researcher. We only started to record the EEG data when the actual experiment round began and stopped recording once each experiment round had finished. After each session, participants were asked to completed two questionnaires: Game Experience Questionnaire (GEQ) [123], Simulator Sickness Questionnaire (SSQ) [131]. Between sessions, they could rest as much as they want. The whole experiment lasted about 35 minutes for each participant.

Section 8.3.3 Results

We analyzed the data using a two-way repeated measures ANOVA with two independent variables, Display Type (VR and Large Display) and Game Mode (Single-tasking and Multi-tasking). Bonferroni correction was used for pairwise comparisons, and Greenhouse-Geisser adjustment was used for degrees of freedom for violations to sphericity. We reported effect size η_p^2 whenever possible.

The gameplay performance data were recorded in the background during gameplay. We evaluated the data using the Missing Target Rate (MTR) which was the percentage of the objects missed by the users among all objects generated by the system. MTR for foot and arm objects was analyzed separately. For EEG data, we excluded the Delta and Gamma data in the analysis because Delta waves could be affected by blinking and Gamma waves by muscle movements. Therefore, we only analysis the Alpha, Beta, and Theta waves in this study. In details, Alpha power increases have been associated with cortical inactivity and mental idleness. Beta activity is most evident in the frontal cortex and has been connected to cognitive processes, decision making, problem-solving, and information processing. Theta activity seems to be related to creativity, intuition, memory recall, emotions and sensations [193].

Gameplay Performance. Figure 8-3a shows the mean MTR for each condition for foot game objects. ANOVA tests yielded a significant effect of Game Mode ($F_{1,11} = 37.864$, $p < .001$, $\eta_p^2 = .775$), but not for Display Type ($F_{1,11} = 2.628$, $p = .133$, $\eta_p^2 = .193$). There was also a significant interaction effect on Display Type × Game Mode ($F_{1,11} = 7.918$, $p < .05$, $\eta_p^2 = .419$). Post-hoc pairwise comparison showed that participants missed more foot game objects ($p < .001$) in multi-tasking mode (M = 23.5%, s.e. = 2.0%) than single-tasking mode (M = 9.8%, s.e. = 1.8%). No main and interaction effects were found for hand game objects.

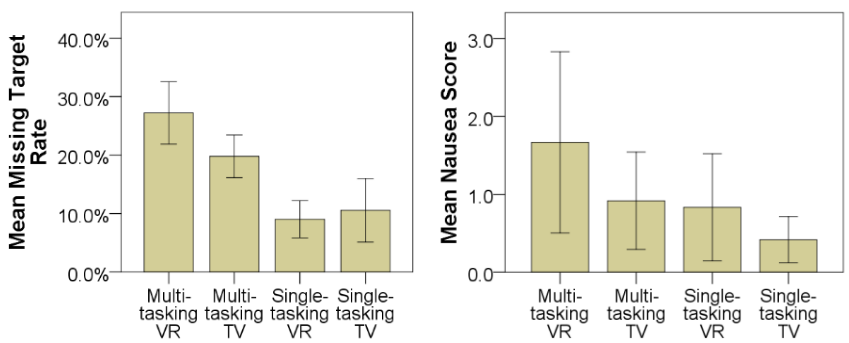


Figure 8-3. (a; left) Mean missing target rate on foot game object. (b; right) Mean nausea score from SSQ. Error bars indicate ±2 standard errors.

Simulator Sickness Questionnaire. Regarding the participants' perceived level of simulator sickness (Nausea, Oculomotor), there was a significant main effect of Game Mode on Nausea ($F_{1,11} = 5.333$, $p < .05$, $\eta_p^2 = .356$), but not for Display Type ($F_{1,11} = 4.115$, $p = .067$, $\eta_p^2 = .272$) and Display Type × Game Mode ($F_{1,11} = .169$, $p = .689$, $\eta_p^2 = .015$). Post-hoc pairwise comparison indicated that participants felt sicker ($p < .05$) when playing in the multi-tasking mode (M = 1.29, s.e. = 0.35) than single-tasking mode (M = 0.63, s.e. = 0.21). Figure 8-3b shows the mean nausea score from SSQ for each condition. No main and interaction effects were found on Oculomotor.

Game Experience Questionnaire. The core GEQ module consists of seven components (Competence, Tension, Sensory and Imaginative Immersion, Flow, Negative Affect, Positive Affect, Challenge). ANOVA tests yielded a significant effect for Game Mode ($F_{1,11} = 7.957$, $p < .05$, $\eta_p^2 = .420$) on Challenge, but not for Display Type ($F_{1,11} = .166$, $p = .691$, $\eta_p^2 = .015$) and Display Type × Game Mode ($F_{1,11} = .617$, $p = .449$, $\eta_p^2 = .053$). Post-hoc pairwise comparison revealed that users felt multi-tasking (M = 1.85, s.e. = 0.18) was more challenge ($p < .05$) than single-tasking (M = 1.38, s.e. = 0.20). Figure 8-4a shows the mean challenge score from GEQ for each condition. However, no other main and interaction effects were found on Competence, Tension, Sensory and Imaginative Immersion, Flow, Negative Affect, Positive Affect.

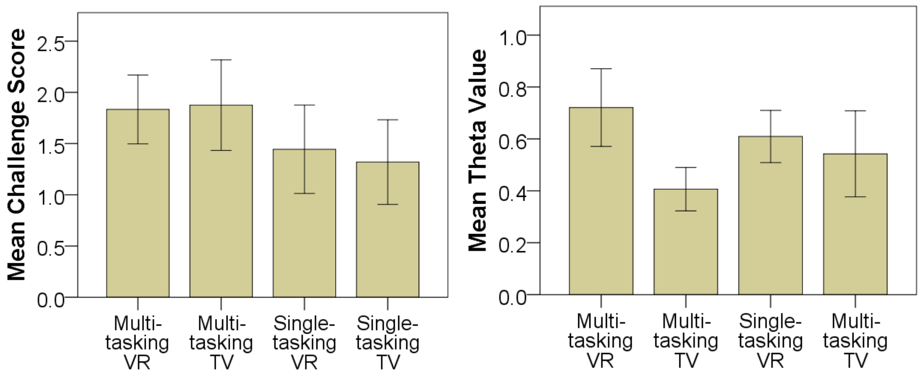


Figure 8-4. (a; left) Mean challenge score from GEQ. (b; right) Mean Theta power during the gameplay. Error bars indicate ±2 standard errors.

EEG. We calculate the mean value of each brainwave signal. Figure 8-4b presents the mean Theta value during the gameplay among the 4 conditions. ANOVA tests showed there was a main effect of Display Type for Theta ($F_{1,11} = 7.415$, $p < .05$, $\eta_p^2 = .403$), but not of Game Mode ($F_{1,11}$ = .031, p = .864, η_p^2 = .003) and Display Type × Game Mode ($F_{1,11}$ = 3.604, p = .084, η_p^2 = .247). No other main and interaction effect were found for both Alpha and Beta waves.

Section 8.4 Discussion

Section 8.4.1 Gameplay Performance

We found that multi-tasking affects the way how participants would decide to eliminate the game objects. From our observation and comments from the participants, they prefer to eliminate the easy option (hand game object) when in a complicated situation (hand and foot game object come in one time).

Section 8.4.2 Simulator Sickness

We found that VR did not generate a higher level of simulator sickness than Large Display. This shows that VR exergames are as feasible as those shown in Large Display for Nausea and Oculomotor. Meanwhile, we found that participants felt sicker when played in multi-tasking mode than single-tasking, suggesting multi-tasking may cause a higher sickness than single-tasking in a full-body motion-based exergame. Therefore, we suggest the future designer should carefully design a game that may consist of a series of multi-tasking tasks, as it may cause a higher sickness.

Section 8.4.3 Game Experience

We found that multi-tasking mode is more challenging than the single-tasking mode, but VR and Large Display share the same level of challenge for participants. Regarding the other GEQ components (Competence, Tension, Sensory and Imaginative Immersion, Flow, Negative Affect, Positive Affect), VR and Large Display have brought similar game experience to participants while single-tasking and multi-tasking also have no effect on their game experience.

Section 8.4.4 EEG

We found a higher mean Theta value for the users when they played the game in VR than Large Display. One possible explanation is that VR might at some point affect the ways participants calculate the spatial position of the game objects. Early studies

[129,139] have shown that Theta power increases during spatial navigation, especially during processing of spatial cues and landmarks, which was also required in our game. We have not found any significant effect of Display Type and Game mode on Alpha and Beta waves.

Section 8.4.5 Limitation and Future Work

The experiment has one issue where each game session only last three minutes, which is a relatively short time period and may lead to a different result. Future work will increase the time for each session, test different game elements (i.e., game object's moving speed). We will also seek an opportunity to examine how feasible for elderly to play VR Exergame. Moreover, we have recorded the gameplay video for each condition, and the next step will focus on the analysis of Event-Related Potential (ERP), which analyses brain waves as an event is happening, helping us to have a deeper understanding of what is happening during gameplay [188]. Also, we will investigate how the EEG metrics related to the subjective questionnaires [186].

Section 8.5 Conclusion

This chapter has explored the effects of display/device type (virtual reality and large display) and game mode (single-task and multi-task) for exergames. Our experiment with 12 young adults indicates that (1) players have the same level of game experience and motion sickness when playing the exergame in either VR and large display; (2) VR has led to increasing Theta power in players' brain; (3) players believe multi-tasking is more challenging and brings a higher of motion sickness than single-tasking; and (4) players have a worse game performance in multi-tasking than single-tasking. For the last two findings, we suggest that if the sickness is crucial for players, they should avoid playing multi-tasking mode, if sickness and performance are not a concern and the players would like to train their hand-foot coordination skills, they should play multi-tasking mode.

Chapter 9 Exploring the Effects of Viewing Perspective in Full-body Motion-based Exergame

Section 9.1 Introduction

Physical inactivity has been identified as the fourth leading cause of death globally [151]. It is now well established that a sedentary lifestyle is a unique risk factor for several diseases such as type 2 diabetes and cardiovascular disease [276], which account for about 30% of global mortality. In recent years, the idea of using interactive computing systems that leverage gamification to promote physical activity has been widely researched [173]. Prior studies [15,73,185,240] have shown that exergames, a type of games that encourage physical activity, can increase enjoyment and intrinsic motivation compared with conventional exercises; as such, they can be effective in promoting physical and mental health [207,224].

Given the advantages of engaging people in long-term and regular physical activity, various non-HMD (like using interfaces such as a flat-screen television/monitor) exergames have been designed to encourage people to be more active [85], promote a positive lifestyle [80] and self-care [68]. Previous literature has shown that exergames could bring physical and mental health outcomes to players. For example, Peng et al. [204] have performed a meta-analysis of energy expenditure in exergames where their main finding suggests that exergames are as effective as traditional physical activities that facilitate light- and moderate-intensity physical exertion. Huang et al. [119] found that exergames can induce positive changes in happiness, perceived energy levels, and relaxation for people who are enthusiastic about doing exercises. Other studies have shown that exergames are as effective as conventional balance training exercises [14,228]. Moreover, the benefits of playing exergames include, but not limited to, improving the quality of life [247], reducing state anxiety [267], as well as improvements in the number of steps taken, standing balance, gait speed, and mobility [82].

Given the recent emergence of affordable head-mounted displays (HMDs), especially for virtual reality (VR), there is limited and only preliminary research on VR exergames. Recently, Barathi et al. [17] have implemented an exercycle game with interactive feedforward method using VR to improve players' performance and maintain intrinsic motivation. Ioannou et al. [124] found that virtual augmented running and jumping in VR could increase intrinsic motivation, perceived competence, and flow. Xu et al. [285] have found that playing exergame in VR would not result in a higher cybersickness than a 50-inch TV. In general, researchers have suggested that VR is useful in promoting physical activity in sedentary and obese children [223], especially to increase their motivation to exercise [179,206]. However, the difference between exergaming with a common display and VR is still largely underexplored, especially regarding their physical and health benefits.

Traditional approaches such as direct observations [196] and subjective measurements [85] are the commonly used methods to measure user experience during games. However, they can be intrusive and not reliable. Psychophysiological methods, such as using electroencephalography (EEG), provide relatively non-intrusive, covert, and reliable measurements of affective states that determine user experience, and this makes them suitable for studying interactive entertainment [217]. Such methods have been used to investigate the effect of controller types [193], viewing angles [189], display types (DTs), and tasking modes [285] on players' brainwave patterns.

Chang et al. [38] and Stoffregen et al. [255] have proved that videogames can carry a significant risk of cybersickness. One solution to reduce it is by seeking the most suitable viewing perspective (VP) (e.g., first-person vs. third-person). For example, Medina et al. [178] found that cybersickness were more pronounced for the first-person viewing perspective (1PP) group than the third-person viewing perspective (3PP) group when performing locomotion walking in navigation tasks in an VR environment. Similarly, Monteiro et al. [187] pointed out that playing an VR racing game in 3PP is less likely to induce cybersickness when compared with playing it in 1PP.

Given the considerations just mentioned, the aim of this study was to investigate the effect of DT (VR and large TV) and VP (1PP and 3PP) on players' exertion, engagement, and overall gaming experience of exergames. To this end, we conducted a first study to select a gesture set for a gesture-based game to make sure that the selected gestures would not affect players' gameplay in both DTs. Afterward, in a second study, we investigated the effect of DTs and VPs when interacting with an exergame.

The current investigation has been guided by the following hypotheses. Because previous research [240] showed that playing an exercycle game with a common flat monitor and VR led to an equal level of burned calories, we hypothesized that:
H1: **a)** There would be no significant differences in gameplay performance (i.e., completing the same number of gestures) among DTs; therefore, **b)** we believe the levels of exertion (%HR_{max}, calories burned, and Borg RPE) should also be the same among the DTs.

H2: VR could result in a higher game experience than Large Display (LD).

Similarly, because prior work that tested different types of interventions showed that 3PP could lead to a lower motion sickness than 1PP [178,187], we predicted that:
H3: During a gameplay of more than three minutes, **a)** 3PP could lead to a lower cybersickness than 1PP in exergames. As for VR, we believe that **b)** it could lead to a higher level of cybersickness than LD.

Section 9.2 Study One

In the interest of removing any possible bias toward a DT, Study one aimed at identifying a set of full-body gestures for the exergame to be used in Study two. That is, we evaluated gestures that would not be affected by DT.

Section 9.2.1 Participants

Twenty-four participants were recruited from a local university campus to participate in this experiment. Because two participants' EEG data were lost due to bad connection between the devices, we recruited another two participants. The final 24 participants (six females) were aged between 19 and 27 (mean = 22.04) years old. Twenty-two played videogames regularly (17 of them played weekly). For the VR group, only two of them were frequent users of VR.

The inclusion criteria of the participants for the study were those who: (1) answered "no" to all Physical Activity Readiness Questionnaire [261], (2) had a resting blood pressure lower than 140/90 mmHg, and (3) had a common (10%~90%) [200] resting heart rate depending on their age and gender.

Section 9.2.2 Instruments

To avoid familiarity with gestures that could potentially affect the selection of gestures, we employed a one-way between-subjects experiment design with 24 participants (six females) equally distributed in two groups where the independent variable was DT—VR and LD. The experiment was conducted at a university lab. We used an Oculus Rift CV1 as our HMD and a 50-inch 4K TV as our LD. Both devices were connected to a standard computer with an i7 CPU, 16GB RAM, and a GeForce GTX 1080Ti GPU. The brainwave signals were collected by a MUSE headset Edition 1. The program was built in Unity3D, and players' gestures were detected by a Microsoft Kinect 2.

The National Aeronautics and Space Administration-Task Load Index (NASA-TLX) [107] is a validated instrument for measuring workload [109], which consists of six subscales that represent independent clusters of variables: mental, physical, and temporal demands, frustration, effort, and performance. It first presents users with a series of pairs of rating scale titles (e.g., effort vs. mental demands) and asks users to choose which of the items was more important to the experience of workload in the task(s) that were just performed. Then, it asks users to rate each workload cluster in a 21-Likert scale [109]. The NASA-TLX has been widely used by universities, industries, and governments [108].

Participants' Rating. Participants needed to rate each gesture via a 7-point Likert scale, ranging from 1, strongly disagree, to 7, strongly agree. A higher score indicated that participants would like to have such a gesture into the final version of the game.

The EEG metric we used for this study was the engagement index, which has been widely used in the research of biocybernetics and automation systems [40,58,75,181,208], is a measurement of how cognitively engaged a person is in a task

[75]. It can be calculated by the formula $E = \frac{\beta}{(\alpha+\theta)}$ [208] where α, β, and θ are averaged value of Alpha, Beta, and Theta waves from the EEG device (i.e., MUSE 1).

Section 9.2.3 Task: Performing the Gestures

Participants needed to perform 12 different gestures in a computer program (Figure 9-1), which was developed by the researchers, with the TV or VR device depending on their assigned group. All gestures were evaluated by rehabilitation doctors we had access to. There were six simple gestures (*Psi*: raising two hands; *Squat*: performing a squat; *Kick*: raising any leg; *Walk*: performing walk-in-place; *Wheel*: performing steering wheel motion; *Zoom*: leaning arms forward and stretching them out), and six complex gestures which were combinations of simple gestures (*Squat*+*Psi*; *Squat*+*Wheel*; *Kick*+*Zoom*; *Kick*+*Wheel*; *Walk*+*Psi*; *Walk*+*Zoom*). For each gesture, instructions were given to participants via a pre-recorded 5-second video (Figure 9-1a). Then, they were requested to repeat each gesture in two 10-second sessions, with 5 seconds of rest in between. The order of the gestures was counterbalanced during the experiment.

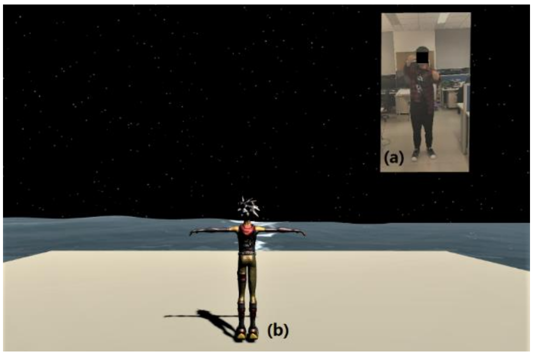


Figure 9-1. Screenshot of Study 1 program. (a) Video display area for participants to follow. (b) A character represents the participant.

Section 9.2.4 Procedure

Before the experiment, participants were told about the purpose of the experiment, given the information sheet to read, and the consent form to sign. Once they agreed to participate, participants were asked to complete a pre-experiment questionnaire to collect demographic data. After the devices used in the experiment were described to them, a researcher helped calibrate the MUSE (to make sure that the MUSE had a good connection with the MUSE application running on a mobile device).

After they understood the process, participants proceeded to play the computer program and perform the gestures. After the experiment, participants needed to complete the post-experiment questionnaire and give comments on the gestures to the experimenter through an interview. The whole experiment lasted about 40 minutes for each participant. The experiment was conducted under the supervision of the experimenter, and the surroundings were cleared of any obstacles to give a safe environment to the participants.

Section 9.2.5 Statistical Analysis

Statistical Package for the Social Sciences (SPSS) version 24 for windows was used for analysis. The Kolmogorov–Smirnov test was used to verify the normality of the data. For NASA-TLX [107] overall workload, we analyzed the data using a univariate analysis of variance (ANOVA) and for its subscales, we employed a multivariate ANOVA to evaluate the effects of DT on gestures that had been performed by participants. For participants' ratings and the EEG Engagement Index, we employed a mix-design ANOVA with gesture (12 gestures) as the within-subjects variable and DT as the between-subjects variable. Bonferroni correction was used for pairwise comparisons and Greenhouse-Geisser adjustment was used for degrees of freedom if there were violations to sphericity in the data.

Section 9.2.6 Results

NASA-TLX

A univariate ANOVA yielded no significant effect of DT ($F_{1,22} = .115$, $p = .737$) on overall workload. A multivariate ANOVA also showed no significant effect of DT on the six NASA-TLX subscales: mental ($p = .442$), physical ($p = .274$), temporal ($p = .421$), performance ($p = .430$), effort ($p = .783$), frustration ($p = .283$). See Table 9-1 for results.

Table 9-1. Means (standard deviations) of national aeronautics and space administration-task load index questionnaire results.

	Overall	*Mental*	*Physical*	*Temporal*	*Performance*	*Effort*	*Frustration*
VR	49.11	40.42	52.92	48.75	45.83	43.33	25.83
	(14.67)	(19.82)	(19.71)	(19.08)	(19.75)	(18.26)	(20.10)
LD	47.42	34.17	61.25	43.75	39.17	41.67	17.50
	(9.15)	(19.29)	(16.53)	(9.08)	(20.87)	(9.85)	(12.70)

Gesture Set

Participants' Ratings. Results of participants' ratings of each gesture can be found in Table 9-2. ANOVA tests yielded a significant effect of Gesture ($F_{5.539,121.848} = 4.288$, $p < .001$) but not of Gesture × Group ($F_{11,242} = .970$, $p = .474$) on the rating scores of the gestures. There was no significant effect of Group ($F_{1,22} = .049$, $p = .826$) on participants' rating of each gesture. Post-hoc pairwise comparisons revealed significant differences between gesture *Psi – Kick+Zoom*, *Psi – Kick+Wheel*, *Walk – Kick+Zoom* (all $p < .05$).

EEG Engagement Index. ANOVA tests yielded no significant effect of Gesture ($F_{11,242} = 1.727$, $p = .175$), Group ($F_{1,22} = 2.619$, $p = .120$), or Gesture × Group ($F_{11,242} = .712$, $p = .726$) on task engagement for each gesture. Results of EEG Engagement Index of each gesture can be found in Table 9-2.

Table 9-2. Means (standard deviations) of participants' ratings and electroencephalography engagement index results of each gesture

	Participants' Rating		*EEG Engagement Index*	
Gesture	*VR*	*LD*	*VR*	*LD*
Psi	5.92 (0.79)	5.67 (1.16)	0.81 (0.48)	0.51 (0.44)
Squat	4.92 (1.38)	4.92 (1.98)	0.76 (0.33)	0.64 (0.27)
Kick	5.08 (1.08)	5.58 (1.38)	0.74 (0.37)	0.52 (0.41)
Walk	5.83 (0.84)	5.67 (1.07)	1.06 (1.48)	0.71 (0.81)
Wheel	5.00 (1.28)	5.25 (1.82)	0.77 (0.37)	0.55 (0.23)
Zoom	5.58 (1.17)	5.17 (1.47)	0.95 (0.75)	0.47 (0.35)
Squat+Psi	5.42 (1.17)	4.67 (1.88)	0.82 (0.83)	0.55 (0.27)
Squat+Wheel	4.83 (1.59)	4.17 (1.70)	0.66 (0.44)	0.48 (0.36)
Kick+Zoom	4.08 (1.51)	5.00 (1.13)	0.65 (0.44)	0.36 (0.32)
Kick+Wheel	4.17 (1.40)	4.25 (1.42)	0.70 (0.58)	0.54 (0.24)
Walk+Psi	5.50 (1.57)	5.25 (1.29)	0.25 (1.01)	0.46 (0.18)
Walk+Zoom	5.50 (1.57)	5.33 (1.37)	0.59 (0.38)	0.46 (0.23)

Section 9.2.7 Discussion

Our results indicated that DT did not affect players' preference of the gestures, their workload, and the engagement index when performing these full-body gestures. We also observed that some gestures might raise issues for future gameplay. We therefore selected the gesture set with the following exclusion considerations: (1) Based on the participants' ratings and comments, we decided to exclude *Wheel*, *Squat+Wheel*, and *Kick+Wheel* gestures since the ratings of these gestures were low. In addition, 20 out of 24 participants complained during the interview that performing these gestures was too hard (e.g., *P9* from the non-VR group: *"*This gesture is too difficult to do*"*). (2) Based on our observations, we decided to exclude *Walk*, *Walk+Psi*, and *Walk+Zoom* gestures since participants could easily go forward instead of walking-in-place when performing such gestures, which could cause tracking issues because, similar to nearly all motion tracking devices, the Kinect 2 we used in Study two only had a limited operational tracking area.

In summary, our exergame in the second study was designed to have four simple gestures—*Psi*, *Squat*, *Kick*, *Zoom*, and two complex gestures—*Squat+Psi* and *Kick+Zoom*.

Since task engagement index was the same, therefore, we hypothesize that:
H4: DT and VP would not affect the EEG Task Engagement Index.

Section 9.3 Study Two

In Study two, we investigated the impact of DT (Large TV and VR) and VP (1PP and 3PP) on gesture-based exergame gameplay performance and experience.

Section 9.3.1 Participants

Another 16 participants were recruited for this study. Because one participant's EEG data were lost due to bad connection, we recruited one more participant. The final 16 participants (five females) included in the data analysis were between the ages of 18 and 28 (mean = 21.75). Ten of them had some prior experience with VR (2 of them interacted with it weekly). Fifteen participants played videogames regularly (12 of them weekly).

We used the same inclusion criteria as Study 1 for this study.

Section 9.3.2 Instruments

The experiment followed a 2 × 2 within-subjects design with combinations of (1) VP—(1PP and 3PP) and (2) DT—(VR and LD). The order of VP × DT was counterbalanced in the experiment.

In addition to the devices used in Study 1, we used a Polar OH1, which has been proved to be able to capture good heart rate (HR) data when compared with the gold standard of HR measurement of an electrocardiography device [113,236], to record participants' heart rate and calorie consumption.

Participants' task performance was evaluated in terms of the percentage of blocks removed (i.e., when the gesture was performed correctly).

Participants' game experience was measured using the 33-item core module of the Game Experience Questionnaire [123]. It consists of seven components: competence, sensory and imaginative immersion, flow, tension, challenge, negative affect, and positive affect.

Cybersickness was assessed using the 16-item Simulator Sickness Questionnaire [131]. It measures a wide range of possible symptoms of cybersickness, including (but not limited to) nausea, eyestrain, dizziness, and vertigo. Each symptom was rated on a severity scale that ranged from 0 (none) to 3 (severe). The scale had an observed Cronbach's α of 0.91. This scale was aggregated to produce two measures of cybersickness (Nausea and Oculomotor) with 27 and 21 points, respectively.

Exertion was evaluated by (1) the average heart rate (%HR_{max}) and was expressed as a percentage of a participant's estimated maximum HR (220 minus age) [6]. (2)

Calories burned and (3) ratings of perceived exertion were measured by the Borg RPE 6-20 scale [28].

Physiological involvement was assessed by the EEG Engagement Index. For details of this measurement, see Study 1: Instruments section.

Participants' preference of the conditions (VR-1PP, VR-3PP, LD-1PP, LD-3PP) was measured by their rankings of the condition from 1 to 4, where 1 stood for the most preferred option and 4 for the least preferred option.

Section 9.3.3 Task: GestureStar Game

Inspired by the commercial exergames *Beat Saber* and *Just Dance*, we developed GestureStar. In GestureStar, players encountered blocks flying toward them every six seconds and were required to make the corresponding gesture to eliminate each block within six seconds; otherwise, they would miss it. One game lasted about eight minutes (one minute for training and seven minutes for the actual experiment). In total, participants were required to perform 10 gestures during training and 70 during gameplay.

As stated earlier, the game had four simple gestures (*Psi*, *Squat*, *Kick*, *Zoom*), and two complex gestures (*Squat+Psi* and *Kick+Zoom*). We employed six different blocks to represent the gestures in the game (Figure 9-2). Figure 9-3a shows a screenshot of the game and Figure 9-3b shows the setup of a player playing the game.

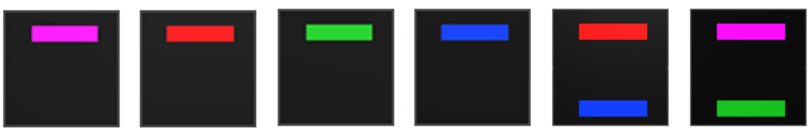


Figure 9-2. The 6 blocks that were used in the game: (left to right) *Kick*, *Squat*, *Zoom*, *Psi*, *Squat+Psi*, *Kick+Zoom*.

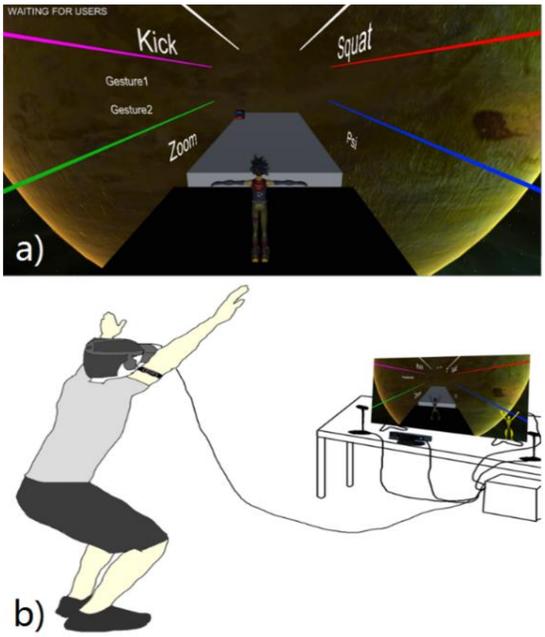


Figure 9-3. A screenshot of the exergame where the name and colored lines in the game work as a reminder of each gesture for the player (a; top) and an example of a participant performing *Psi+Squat* gesture during the game (b; bottom).

Section 9.3.4 Procedure

Participants were briefed of the purpose of the experiment and asked to sign the consent form and complete a pre-experiment questionnaire. Afterward, a researcher helped participants to wear and calibrate the MUSE 1 and Polar OH1. We only recorded EEG and heart rate data for the 7-minute experimental part. After each condition, participants were asked to complete the post-experiment questionnaire. They could rest as much as they want between conditions. After the experiment, they were asked to give feedback and rank each condition. The whole experiment lasted about 1 hour for each participant.

Section 9.3.5 Statistical Analysis

Similar to Study 1, SPSS version 24 for windows was used for analysis. The Kolmogorov–Smirnov test was used to verify the normality of the data. We used the two-way repeated measures ANOVA and Bonferroni correction for pairwise comparisons.

Section 9.3.6 Results

Hypothesis Testing

Analytical results of game performance, exertion (average %HR_{max}, calories burned, and Borg RPE), simulator sickness questionnaire, and EEG Engagement Index can be found in Table 9-3.

Table 9-3. P values of two-way repeated analysis of variance results on game performance, exertion, simulator sickness questionnaire, and electroencephalography task index. Significant results where $p < .05$ are shown in light green, $p < .01$ are shown in green, and $p < .001$ in dark green.

	Completion Rate	*%HR_{max}*	*Calories Burned*	*Borg RPE*	*Nausea*	*Oculom otor*	*EEG Engagement Index*
DT	.468	<.01	<.01	<.001	<.05	<.01	.439
VP	.338	.852	.320	.403	.072	.812	.446
DT×VP	.929	.086	1.000	.333	.300	.510	.303

Details of participants' task performance and exertion for each condition can be found in Table 9-4. No significance was found on task performance between conditions, supporting **H1a**. However, VR had led to a higher %HR_{max} ($p = .005$), calories burned ($p = .001$), and Borg RPE rating ($p = .000$) than LD, not supporting **H1b**.

Table 9-4. Means (standard deviations) of completion rates, exertion, nausea, oculomotor, and electroencephalography engagement index.

	Completion Rate	*%HR_{max}*	*Calories Burned*	*Borg RPE*	*Nausea*	*Oculomotor*	*EEG Engagement Index*
VR_1PP	91.79% (3.28%)	53.60% (6.82%)	42.81 (13.11)	14.50 (1.90)	1.88 (1.54)	2.81 (2.54)	0.36 (0.23)
VR_3PP	92.77% (3.36%)	52.78% (6.33%)	43.81 (13.38)	13.94 (1.12)	2.56 (2.13)	3.13 (2.22)	0.26 (0.43)
LD_1PP	92.41% (5.64%)	50.60% (6.03%)	34.75 (12.07)	11.94 (1.57)	1.69 (1.89)	1.69 (1.82)	0.35 (0.30)
LD_3PP	93.57% (4.92%)	51.25% (5.77%)	35.75 (13.00)	12.06 (1.84)	1.50 (1.41)	1.50 (1.41)	0.38 (0.22)

Table 9-5. P Values of two-way repeated analysis of variance results of the game experience questionnaire. Significant results where $p < .05$ are shown in light green, $p < .01$ are shown in green, and $p < .001$ in dark green.

	Competence	*Sensory and Imaginative Immersion*	*Flow*	*Tension*	*Challenge*	*Negatice Affect*	*Positive Affect*
DT	.588	<.01	<.01	.730	<.01	<.05	.125
VP	.181	.284	.070	.453	.133	.060	.348
DT × VP	.085	<.01	.073	.224	.118	.770	<.05

Table 9-6. Means (standard deviations) of game experience questionnaire subscales.

	Competence	*Sensory and Imaginative Immersion*	*Flow*	*Tension*	*Challenge*	*Negatice Affect*	*Positive Affect*
VR_1PP	2.79 (0.58)	2.66 (0.52)	2.51 (0.69)	0.90 (0.74)	2.09 (0.85)	0.86 (0.84)	3.09 (0.83)
VR_3PP	2.74 (0.69)	2.38 (0.63)	2.26 (0.70)	0.98 (0.75)	2.06 (0.79)	1.10 (0.81)	2.78 (1.03)
LD_1PP	2.54 (0.55)	2.01 (0.51)	1.78 (0.46)	1.00 (0.82)	1.81 (0.85)	1.19 (0.78)	2.64 (0.75)
LD_3PP	2.88 (0.63)	2.09 (0.61)	1.74 (0.43)	0.81 (0.78)	1.50 (0.56)	1.34 (0.76)	2.78 (0.72)

Analytical results of each Game Experience Questionnaire component are shown in Table 9-5. The score for VR was higher than LD regarding challenge ($p = .002$), flow ($p = .004$), sensory and imaginative immersion ($p = .002$), while VR had a lower score regarding negative affect ($p = .023$) than LD. Therefore, the results supported the **H2**. Table 9-6 shows the scores for each component.

No significance was found for Nausea and Oculomotor on VP, not supporting **H3a**. **H3b** was supported since VR had caused a higher level of Nausea ($p = .016$) and Oculomotor ($p = .010$) than LD. Details of the sickness scores for each condition can be found in Table 9-4.

H4 was supported as no significant effect of DT and VP was found on EEG engagement index. Values of EEG engagement index can be found in Table 9-4.

User Preference

Friedman tests yielded a significant difference depending on which version participants preferred $\chi^2(3) = 10.059$, $p = .018$. However, post-hoc analysis with

Wilcoxon signed-rank tests and Bonferroni correction did not reveal any significant difference between conditions, although 63% of the participants selected VR-1PP as their top choice.

Section 9.4 Discussion

Section 9.4.1 Discussion on the Hypotheses

We found support in our results for **H1a**, where participants completed the same number of gestures in both VR and LD conditions. However, **H1b** was not supported, even though the completion rates of the gestures were the same. One possible explanation might be because the weight of the VR HMD that participants had to carry during the VR condition increased the intensity of the exergame, although the Oculus CV1 just weighted 470 g.

We found support for **H2**; that is, playing exergames in VR had a better gameplay experience as it was more challenging, immersive (based on the flow, sensory and imaginative immersion components) to participants, and had fewer negative effects. Interestingly, our findings did not support the results from a previous study [285] in which researchers found that playing a motion-based exergame in VR might have the same level of game experience. One possible explanation might be because the length of our game was much longer than theirs.

Previous studies [178,187] suggested that 3PP could lead to a lower sickness level than 1PP; however, we did not find support for **H3a**. That is, playing an exergame in 3PP did not result in a lower cybersickness level than in 1PP. We hypothesize that since our game demanded a reasonable amount of movement the bone vibration equated in lower levels of cybersickness in both versions equally [271]. Further, in our experiment, participants often focused on a fixed point, so they could better observe the oncoming objects, which equated to the same advantage as 3PP, thus not bringing any special advantage in this scenario. **H3b** was supported, as our data indicated that players felt sicker (both nausea and oculomotor) when playing in VR than LD, which is in line with previous VR studies [4,239].

We confirmed our **H4** that DT and VP did not affect the EEG engagement index.

Section 9.4.2 Practical Implications

Our results indicate that playing a full-body gesture exergame in VR could lead to a higher exertion level than LD (i.e., it burned more calories, and led to a higher $\%HR_{max}$, and perceived exertion level on the Borg RPE). Moreover, playing an exergame in VR can induce not only a higher immersion level but also a lower negative feeling than LD. As such, when players need some exercise, they could be introduced to playing exergames with VR HMDs. However, if players start to get cybersicked quickly, they should play exergames with LD.

For game designers, consideration should be taken with respect to gestures: (1) by not designing and including complex gestures (e.g., *wheel* used in Study 1); (2) by avoiding gestures such as walk-in-place because players might need to move around, which could lead to tracking issues and potentially dangerous situations.

Section 9.4.3 Strengths, Limitations, and Future Work

The strengths of our research include: (1) the gestures used for the exergame were selected systematically (from Study one) to remove any bias toward any particular type of display that could originate from a gesture; (2) the effect of DT (VR and LD) and VP (1PP and 3PP) on cybersickness and exertion in exergames were never previously examined. To our knowledge, we are the first ones to conduct this research; (3) another strength of the study is that it has contributed to the limited research topic of VR on health benefits to its users (e.g., exertion).

There are some limitations to this research. One limitation is that the research involved a relatively small sample (though this is normal in research published in this area [36]). Future work can involve a larger and more diverse group of participants. Moreover, the current version of GestureStar seems only to be a light-intensity game as participants' $HR_{max}\%$ is lower than 64% (see Table 9-4), which is the lower bound of moderate intensity exercises [37]. One possible solution to increase the intensity of the game is by narrowing the wait time for the next block if the player eliminates the current block in advance. In addition, future work can focus on reducing potential nausea and other adverse side effects while increasing the intensity of the VR version of the exergame.

Section 9.5 Summary

Overall, our results suggest that HMDs could result in changes in physiological feelings (Chapter 8) and lead to a better game experience but also a higher sickness (Chapter 9). Hence, our answer for Research Question 4 (i.e., will HMDs affect users experiencing full-body interaction?) is that HMDs could affect users experiencing full-body interaction. Regarding Research Question 5 (i.e., will sickness mitigation factors in other contexts works for full-body interaction), it seems that factors that could help reduce simulator sickness in other contexts may not work for full-body interaction. Our results show consistent results where multi-tasking leads to a higher sickness [293], but viewing perspective does not affect the simulator sickness [187].

Chapter 10 Discussion, Conclusion, and Future Work

Section 10.1 Summary

This thesis has examined the design of motion-based interaction (MbI) for head-mounted displays (HMDs). This thesis first outlines the literature of HMDs, MbI that can be used for HMDs, and formulates four Core Challenges (CC) of motion-based interaction for HMDs that need to be addressed: (CC1) boundary awareness for hand-based interaction; (CC2) efficient hands-free head-based interface for HMDs; (CC3) efficient and feasible full-body interaction for general tasks with HMDs; and (CC4) accessible full-body interaction for applications in HMDs.

Based on these Core Challenges, we further formula the following five Research Questions (RQ):

RQ1 – CC1 – Chapter 5: How can visual boundary awareness techniques support mid-air hand-based interaction?

RQ2 – CC2 – Chapter 6: Can a circular layout achieve an efficient and usable hands-free head-based interaction?

RQ3 – CC3 – Chapter 7: Are directional full-body interaction feasible and efficient for general tasks with HMDs?

RQ4 – CC4 – Chapter 8: Will HMDs affect users experiencing full-body interaction?

RQ5 – CC4 – Chapter 9: Will sickness mitigation factors in other contexts work for full-body interaction?

These Core Challenges and Research Questions are addressed in the context of different research studies. The first study is designed to explore visual techniques for boundary awareness issues for HMDs and comparing them against the non-visual feedback benchmark. The second study evaluates the feasibility of a novel hands-free head-based interaction with a circular interface for HMDs. The third study proposes a directional full-body interaction for general tasks with HMDs and evaluates it against hand-based interaction and hybrid-based (head+hand) interaction. The last challenge was addressed through two studies focusing on designing accessible full-body interactions for applications in HMDs. These two studies investigate the effect of tasking mode and viewing perspective on full-body motion-based interactions for

HMDs and compared their performance and experience against the benchmark (i.e., large display—50-inch 4K TV).

The final chapter of this thesis provides a discussion of Core Challenges and Research Questions listed in Chapter 4, lists a set of design guidelines and takeaway messages, and future work of MbI for HMDs.

Section 10.2 Have Core Challenges and Research Questions Been Addressed?

Section 10.2.1 Challenge 1: Boundary Awareness for Hand-based Interaction

This challenge has been addressed in Chapter 5. We introduced the idea of using visual methods for boundary awareness during interaction for HMDs. In that chapter, we first conducted a systematic formative study to identify the challenges users might face when interacting with HMDs without any boundary awareness information (i.e., how current systems work). Based on the findings, we then propose four methods as our visual solutions: static surfaces, dynamic surface(s), static coordinated lines, and dynamic coordinate line(s). To further explore whether the use of visual technique could make users aware of the tracked interaction area for HMDs, we conducted an experiment with twenty participants to evaluate these four methods against the benchmark (i.e., baseline condition without boundary awareness). Our results show that visual methods for boundary awareness can help with dynamic mid-air hand interactions in AR HMDs.

Chapter 5 also answered RQ1 (i.e., how can visual boundary awareness techniques support mid-air hand-based interaction?). Visual boundary awareness methods should provide information on the distance between users and the boundary to support users in mid-air hand-based interaction. Besides, visual boundary awareness methods can be provided both statically and dynamically.

In summary, the contributions to boundary awareness include:

(1) The first systematic exploration of visual methods for boundary awareness in HMDs.
(2) Results of a user study comparing different visual boundary awareness methods for interacting with virtual objects in these systems.

Section 10.2.2 Challenge 2: Efficient Hands-free Head-based Interface for HMDs

This challenge has been addressed in Chapter 6, with a focus on the text entry task. In that chapter, we presented a case for interaction using a circular layout for HMDs that is dwell-free and does not require users to hold a dedicated input device for letter selection. We have implemented RingText, whose design is based on a circular layout with two concentric circles to support the case. The outer circle is subdivided into regions containing letters. Selection is made using a virtual cursor controlled by the user's head movements—entering a letter region triggers a selection and moving back into the inner circle resets the selection. The design of RingText follows an iterative process, where we initially conduct one first study to investigate the optimal number of letters per region, inner circle size, and alphabet starting location. We then optimize its design by selecting the most suitable features from the first study and creating candidate regions that incorporate two suggested words to appear next to the current letter region (close to the cursor) using a dynamic approach. Our second study compares the text entry performance of RingText with four other hands-free techniques and the results show that RingText outperforms them. Finally, we run a third study lasting four consecutive days with ten participants (five novice users and five expert users) doing two daily sessions and the results show that RingText is quite efficient and yields a low error rate.

The results from Chapter 6 answered RQ2 (i.e., can a circular layout achieve an efficient and useable hands-free head-based interaction?) that a circular layout with head-based interaction can be an efficient interaction for HMDs.

The contributions of this chapter include:

(1) The first example of a formal evaluation of the circular keyboard layout for text input in HMDs.

(2) The first comparison of hands-free text entry mechanisms for both circular and QWERTY keyboard layouts in HMDs.

(3) A case for using dynamic (rather than static) locations for recommended words—to our knowledge, this is the first case that shows the usefulness of using dynamic locations of these words.

(4) A demonstration of the effectiveness of RingText, a circular layout text entry technique that relies on head motions and uses dynamic locations for recommended words, through a 4-day user study.

Section 10.2.3 Challenge 3: Efficient and Feasible Full-body Interaction for General Tasks with HMDs

This challenge has been addressed in Chapter 7. We presented a directional full-body interaction for HMDs that is both hands- and device-free (i.e., DMove). To use DMove, a user needs to perform directional motions such as moving one foot forward or backward with the body also move in the direction. The design of DMove was decided through an experiment while we investigated the recognition accuracy of the motion directions of our method and the social acceptance of this type of interaction together with users' comfort rating for each direction. We then conducted a second study to compare DMove with two other device-free motion-based approaches—hand-based interaction and hybrid-based (head+hand) interaction for menu selection tasks regarding task performance and user preferences (workload, motion sickness, user experience). Our results showed that DMove outperforms (1) hand-based interaction regarding task performance and user experience and (2) hybrid-based interaction regarding user experience.

Results from Chapter 7 confirmed that directional full-body interaction is highly accepted by users and does not cause a higher sickness than other MbI. In addition, it outperforms hand-based interaction and is comparable to hybrid-based (head+hand) interaction. Overall, this answers our RQ3 (i.e., are directional full-body interaction feasible and efficient for general tasks with HMDs?) that directional full-body interaction is a feasible and efficient interaction technique for general tasks with HMDs.

Overall, we have made the following contributions:

(1) A motion direction recognition method that requires no additional handheld devices nor sensors for current HMDs.
(2) An optimized directional motion-based interface (DMove).
(3) An evaluation of three menu selection methods for HMDs.

(4) A set of guidelines for applications that use directional motion-based interactions.
(5) Two applications external to menu selection and that use DMove as their interaction interface.

Section 10.2.4 Challenge 4: Accessible Full-body Interaction for Applications in HMDs

This challenge has been addressed in Chapters 8 and 9, focusing on full-body motion-based exergames. We have addressed this challenge by (1) exploring the differences between performing full-body interaction for HMDs and common displays (i.e., TV) and (2) providing a set of design guidelines.

In Chapter 8, we evaluated the effect of tasking mode (single-tasking and multi-tasking) on exergame with 12 participants and found that multi-tasking could lead to a worse performance, gameplay experience, and a higher sickness than single-tasking. In Chapter 9, we found that playing exergame in HMDs could lead to greater health benefits (i.e., exertion) and provide a much positive gameplay experience than in large displays. Meanwhile, we proposed and evaluated a list of full-body motion-based gestures that is accessible and safe for HMDs.

Chapter 8 suggests that HMDs could result in higher Theta brainwave and Chapter 9 indicates that HMDs could lead to a better game experience but also a higher sickness. Hence, we believe HMDs could affect users to experience full-body motion-based interaction differently in HMDs and common display, which answers RQ4 (i.e., will HMDs affect users experiencing full-body interaction).

Regarding RQ5 (i.e., will sickness mitigation factors in other contexts works for full-body motion-based interaction), we suggest factors that could reduce simulator sickness in other contexts may not work for full-body motion-based interaction. Our results show consistent results where multi-tasking leads to a higher sickness [293], but viewing perspective does not affect the simulator sickness [187].

In summary, we made the following contributions:

(1) The first investigation of display type and tasking mode in full-body motion-based exergame on gameplay performance, experience, and physiological feeling.
(2) The first investigation of display type and viewing perspective in full-body motion-based exergame on gameplay performance, experience, physiological feeling, and exertion.
(3) A list of gestures that can be used for full-body motion-based exergame for HMDs.
(4) A list of design guidelines for designing full-body motion-based exergame for HMDs.
(5) Two standing full-body motion-based exergames for HMDs.

Section 10.3 Design Recommendations

Section 10.3.1 Visual Methods for Boundary Awareness

Providing boundary awareness method by default

During the phase where participants tried the device to get to know it, we observed that novice users tended to over-value the FoV of the HMD. They would ignore the FoV of the HMD device and assume that the interaction would be the same as what they would typically do during actual tasks. Therefore, visual boundary awareness methods should be provided for users at the beginning stage to remind them about the limited size of the tracked area and FoV of the device. It could be disabled when users think they could do without it.

User-dependent

Visual boundary awareness methods should be tuned to suit the individuals' needs and predispositions. One way to do this is to let users experience all available techniques first and select the ones which can bring a better interaction experience. Users should avoid the method that could lead them to a high error rate, computer vision syndrome, or workload.

Section 10.3.2 DMove: Directional Motion-based Interaction for HMDs

Cater to individual differences

Based on our findings, DMove should use an individual's dataset to maintain (100% or close to 100%) accuracy to take into account each user's height, weight, movement speed, and step distance. To account for these factors and prevent poor accuracy, DMove for general users should be calibrated according to individual physical features and abilities. Besides, we predict a motion just right before a user finishes it by comparing the head movement speed with a pre-set constraint, which should also be tuned to suit the individuals.

Flexibility, efficiency of use, customizability

The comfort ratings from our study suggest that the Close level is much easier to reach, and it does not cause discomfort, while directions that users can see—N, NE, NW, E, W are much easier to perform. As such, we suggest putting frequently used items/functions closer to the users and avoid putting them at the directions that users cannot see easily to increase efficiency and usability.

Not in front of strangers and public venues

DMove should be used for indoor scenarios such as at home/work environment (or outdoor but when there is nobody around). In addition, an interface based on DMove should be used in front of the people users are familiar with instead of strangers.

Provide feedback and keep consistency with other interfaces

Although DMove can be eyes-free, we suggest that the menu should always appear as a simple non-obtrusive visual interface on the HMD on-demand, similar to a context menu, whenever users want to use it and so that they do not have to memorize the items of the menu. Similar to what we have done in this research, we suggest that the interface shows the user's movement location—e.g., a simple visual cue like a dot can be used to indicate to which direction they are moving. Visual and/or auditory feedback can be included to tell them that a selection has been successfully made.

Section 10.3.3 Full-body Motion-based Interaction

Avoid multi-tasking situations

Multi-tasking could not only lead to worse performance but also bad experience (e.g., simulator sickness). Hence, when performing full-body motion for HMDs, designers should avoid situations where players need to use their hands and feet to interact with multiple virtual objects separately.

Consideration should be taken with respect to gestures

Game designers should not design and include complex gestures (e.g., *wheel*—turning a wheel motion). In addition, designers should avoid designing gestures such as walk-in-place because players might need to move around, which could lead to tracking issues and potentially dangerous situations.

Warning signs should be provided for standing exergames

We observed that players tend to move around during gameplay, which could lead to potentially dangerous situations (e.g., hitting objects that are in the environment and going out of the safe tracking area) or decrease the recognition performance of the sensors (e.g., tracking may not work when they are too close to or far from the sensors). Therefore, we suggest providing warning signs for standing exergames if users have left (or are about to leave) the calibration position and are too far to keep them protected.

Section 10.4 Summary of Takeaways

Below we summarize the takeaway messages for designers and researchers of the MbI for HMDs:

- For mid-air hand-based interaction, visual boundary awareness methods should be provided by default, and users should have the option to select their desired method.
- Circular layouts with go-and-hit selection style could form an efficient hands-free head-based input for HMDs.
- Directional full-body motion-based interaction is a feasible and efficient interaction technique for HMDs, but its usage should be limited when users are in public venues or front of strangers.

- Visual support should be provided to directional full-body interfaces.
- Frequently used functions in directional full-body interaction should be mapped to the directions that users can easily see.
- Extra consideration should be taken when design full-body interaction, especially for HMDs, because (1) factors that could help reduce simulator sickness in other contexts may not work for full-body interaction, (2) users could perceive differently in HMDs regarding game experience, simulator sickness, and physiological feeling.
- Multi-tasking should be avoided in full-body interaction if multi-tasking training is not a must because multi-tasking could be more challenging to do and can lead to worse performance and higher-level simulator sickness.
- For standing full-body interaction where external motion-tracking devices are used, warning signs should be provided if users leave their position; otherwise, this will affect the gameplay.

Section 10.5 Conclusion

In conclusion, this dissertation has investigated the design of motion-based interaction (MbI) for head-mounted displays (HMDs). It has first identified four challenges in the context of MbI for HMDs (boundary awareness for mid-air hand-based interaction, efficient hands-free head-based interaction, feasible and efficient full-body interaction for general tasks in HMDs, and accessible full-body interaction for applications in HMDs). Then, we have presented solutions to each challenge: (1) visual boundary awareness techniques for hand-based interaction, (2) circular interface for hands-free head-based input, (3) directional full-body interaction with directions mapped to functions/items for general tasks in HMDs, and (4) several recommendations (e.g., gestures and design guidelines) for full-body interaction applications. At last, we have concluded this dissertation with a set of design recommendations and takeaway messages for MbI for HMDs.

All in all, this thesis can act as a starting point for designers who are interested in designing MbI for HMDs. With the rapid advancements of motion-tracking devices and algorithms, MbI can play a significant role in HMDs and are capable of much more than they are currently used.

Section 10.6 Future Work

Several features could further enhance the performance and experience of MbI for HMDs but that could not be implemented due to the time constraint.

We only tested boundary awareness with one-handed gestures and translation tasks [32]. Future work can explore whether our findings will also be applicable to two-handed gesture-based interactions where large motions are required and other tasks in 3D environments (e.g., for 3D modeling [51] where the interaction would be more complicated). In addition, several values used in our experiment are pre-defined fixed values due to the lack of related prior work. Future work can (1) implement a dynamic color changing scheme for the surface(s) to suit the background [78,79]; (2) focus on exploring the most suitable values for the opacity of the color and the distance for activating the dynamic visual cues for boundary awareness.

RingText and other similar circular layouts could also be strengthened in two ways. (1) Currently, RingText only contains 28 items in one level. A technique that could switch the item layer to enrich the interaction is needed. One possible solution that has been tested initially is to use the forward head movement. However, future research is needed to determine how this approach will work. (2) We have not investigated the optimal size of the trigger area for RingText. Future work is needed to investigate the optimal size(s) of the trigger area to let users select letters quickly without incurring many mistakes. One possible solution is to apply a static decoding method [97] to handle the input noise further. This is similar to a method to mitigate the “fat finger” problem in smartphones [268], where users with large fingers may mistakenly select unintended buttons. In our case, it may be possible to use this model to help us understand which letters the user is aiming to type.

There are several paths to further strengthen DMove. (1) The levels in one direction can be increased to allow for more items. This may be useful because, although the number of items in the radial menu is large enough to meet the needs of applications in AR systems, there can be cases in which a large number of items are needed. As such, having more levels (i.e., used to hold items/functions) will allow more items to be included. (2) It is possible to optimize the layout further—e.g., finding the most suitable distance for each level in one direction instead of the pre-defined values (i.e., 30cm) that we used in our study. (3) Since we want DMove to be accessed on-demand, future work can also focus on exploring ways to separate DMove from ordinary moving. We have done some preliminary explorations and one way that is possible for all commercial HMDs, for instance, is to use the third dimension (Y-axis) where users can perform an on tiptoe (up/down) action to wake up the DMove. This way, DMove can also be suitable for users with arm/hand disabilities as it does not require hands or any input device.

Our investigation mainly focuses on healthy young adults. Future work could focus on investigating the motion-based interaction with different target user groups (i.e., middle-aged adults, older adults, disabled users) since different population groups could face unique challenges. For instance, age-related declines are unique challenges for middle-age adults and older adults, previous studies show that reduction include, but not limited to, cognitive abilities [72,266], motor skills [133], muscle strength [35,141]). While physically disabled users could not use walking as an interaction technique, instead, they could use the wheelchair.

Bibliography

1. Alexander Aarts, Joanna Anderson, Christopher Anderson, Peter Attridge, Angela Attwood, Jordan Axt, Molly Babel, Štěpán Bahník, Erica Baranski, Michael Barnett-Cowan, Elizabeth Bartmess, Jennifer Beer, Raoul Bell, Heather Bentley, Leah Beyan, Grace Binion, Denny Borsboom, Annick Bosch, Frank Bosco, and Mike Penuliar. 2015. Estimating the Reproducibility of Psychological Science. *Science* 349. https://doi.org/10.1126/science.aac4716
2. Johnny Accot and Shumin Zhai. 2002. More than Dotting the i's — Foundations for Crossing-Based Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '02), 73–80. https://doi.org/10.1145/503376.503390
3. David Ahlström, Khalad Hasan, and Pourang Irani. 2014. Are you comfortable doing that?: acceptance studies of around-device gestures in and for public settings. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, 193–202. https://doi.org/10.1145/2628363.2628381
4. Hironori Akiduki, Suetaka Nishiike, Hiroshi Watanabe, Katsunori Matsuoka, Takeshi Kubo, and Noriaki Takeda. 2003. Visual-vestibular conflict induced by virtual reality in humans. *Neuroscience Letters* 340, 3: 197–200. https://doi.org/10.1016/S0304-3940(03)00098-3
5. Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting your best foot forward: investigating real-world mappings for foot-based gestures. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 1229. https://doi.org/10.1145/2207676.2208575
6. American College of Sports Medicine. 2010. *ACSM's Guidelines for Exercise Testing and Prescription*. Lippincott Williams & Wilkins, Philadelphia, US.
7. Joaquin Anguera, Jacqueline Boccanfuso, Jean Rintoul, Omar Al-Hashimi, Farshid Faraji, Jacki Janowich, Eric Kong, Yudy Larraburo, Cammie Rolle, Erica Johnston, and Adam Gazzaley. 2013. Video game training enhances cognitive control in older adults. *Nature* 501, 7465: 97–101. https://doi.org/10.1038/nature12486
8. Georg Apitz and François Guimbretière. 2004. CrossY: A Crossing-Based Drawing Application. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (UIST '04), 3–12. https://doi.org/10.1145/1029632.1029635
9. Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3: 121–136. https://doi.org/10.1016/j.cag.2012.12.003
10. Rowel Atienza, Ryan Blonna, Maria Isabel Saludares, Joel Casimiro, and Vivencio Fuentes. 2016. Interaction techniques using head gaze for virtual reality. In *2016 IEEE Region 10 Symposium (TENSYMP)*, 110–114. https://doi.org/10.1109/TENCONSpring.2016.7519387
11. Shiri Azenkot and Shumin Zhai. 2012. Touch behavior with different postures on soft smartphone keyboards. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, 251–260.
12. Ronald T Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments* 6, 4: 355–385.

13. Mihai Bâce, Teemu Leppänen, David Gil de Gomez, and Argenis Ramirez Gomez. 2016. ubiGaze: ubiquitous augmented reality messaging using gaze gestures. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications on - SA '16*, 1–5. https://doi.org/10.1145/2999508.2999530
14. Jéssica Maria Ribeiro Bacha, Gisele Cristine Vieira Gomes, Tatiana Beline de Freitas, Larissa Alamino Pereira Viveiro, Keyte Guedes da Silva, Géssika Costa Bueno, Eliana Maria Varise, Camila Torriani-Pasin, Angélica Castilho Alonso, Natalia Mariana Silva Luna, Júlia Maria D'Andrea Greve, and José Eduardo Pompeu. 2017. Effects of Kinect Adventures Games Versus Conventional Physical Therapy on Postural Control in Elderly People: A Randomized Controlled Trial. *Games for Health Journal* 7, 1: 24–36. https://doi.org/10.1089/g4h.2017.0065
15. Bruce W. Bailey and Kyle Mclnnis. 2011. Energy Cost of Exergaming: A Comparison of the Energy Cost of 6 Forms of Exergaming. *Archives of Pediatrics & Adolescent Medicine* 165, 7: 597. https://doi.org/10.1001/archpediatrics.2011.15
16. Rosa M. Baños, Patricia Escobar, Ausias Cebolla, Jaime Guixeres, Julio Alvarez Pitti, Juan Francisco Lisón, and Cristina Botella. 2016. Using Virtual Reality to Distract Overweight Children from Bodily Sensations During Exercise. *Cyberpsychology, Behavior, and Social Networking* 19, 2: 115–119. https://doi.org/10.1089/cyber.2015.0283
17. Soumya C. Barathi, Eamonn O'Neill, Christof Lutteroth, Daniel J. Finnegan, Matthew Farrow, Alexander Whaley, Pippa Heath, Jude Buckley, Peter W. Dowrick, Burkhard C. Wuensche, and James L. J. Bilzon. 2018. Interactive Feedforward for Improving Performance and Maintaining Intrinsic Motivation in VR Exergaming. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. https://doi.org/10.1145/3173574.3173982
18. Ralph Barnett, Steven Schmid, and Dolores Gildin. 2009. Foot Controls: Riding the Pedal. *The Ergonomics Open Journal* 2: 13–16. https://doi.org/10.2174/1875934300902010013
19. Chandramallika Basak, Walter R. Boot, Michelle W. Voss, and Arthur F. Kramer. 2008. Can Training in a Real-Time Strategy Videogame Attenuate Cognitive Decline in Older Adults? *Psychology and aging* 23, 4: 765–777. https://doi.org/10.1037/a0013494
20. Steffi Beckhaus and Ernst Kruijff. 2004. Unconventional Human Computer Interfaces. In *ACM SIGGRAPH 2004 Course Notes* (SIGGRAPH '04), 18-es. https://doi.org/10.1145/1103900.1103918
21. Michael V. D. Bergh and Luc V. Gool. 2011. Combining RGB and ToF cameras for real-time 3D hand gesture interaction. *2011 IEEE Workshop on Applications of Computer Vision, WACV 2011*: 66–72. https://doi.org/10.1109/WACV.2011.5711485
22. Ana M. Bernardos, David Gómez, and José R. Casar. 2016. A Comparison of Head Pose and Deictic Pointing Interaction Methods for Smart Environments. *International Journal of Human-Computer Interaction* 32, 4: 325–351. https://doi.org/10.1080/10447318.2016.1142054
23. Aimee L. Betker, Ankur Desai, Cristabel Nett, Naaz Kapadia, and Tony Szturm. 2007. Game-based Exercises for Dynamic Short-Sitting Balance Rehabilitation of People With Chronic Spinal Cord and Traumatic Brain Injuries. *Physical Therapy* 87, 10: 1389–1398. https://doi.org/10.2522/ptj.20060229

24. Ali Bigdelou, Loren Schwarz, and Nassir Navab. 2012. An adaptive solution for intra-operative gesture-based human-machine interaction. *International Conference on Intelligent User Interfaces, Proceedings IUI*. https://doi.org/10.1145/2166966.2166981
25. Kanad Biswas and Saurav Kumar Basu. 2011. Gesture recognition using Microsoft Kinect. In *The 5th International Conference on Automation, Robotics and Applications*, 100–103. https://doi.org/10.1109/ICARA.2011.6144864
26. Frederick Bonato, Andrea Bubka, and Stephen Palmisano. 2009. Combined pitch and roll and cybersickness in a virtual environment. *Aviation, space, and environmental medicine* 80, 11: 941–945. https://doi.org/10.3357/asem.2394.2009
27. Kevin Boos, David Chu, and Eduardo Cuervo. 2016. FlashBack: Immersive Virtual Reality on Mobile Devices via Rendering Memoization. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (MobiSys '16), 291–304. https://doi.org/10.1145/2906388.2906418
28. Gunnar Borg. 1982. Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise* 14, 5: 377–381.
29. Gunnar Borg. 1998. *Borg's perceived exertion and pain scales*. Human Kinetics, Champaign, IL, US.
30. Jared N. Bott, James G. Crowley, and Joseph J. LaViola. 2009. Exploring 3D Gestural Interfaces for Music Creation in Video Games. In *Proceedings of the 4th International Conference on Foundations of Digital Games* (FDG '09), 18–25. https://doi.org/10.1145/1536513.1536527
31. 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* (I3D '97), 35-ff. https://doi.org/10.1145/253284.253301
32. Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.
33. Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. 2012. Questioning naturalism in 3D user interfaces. *Communications of the ACM* 55, 9: 78–88. https://doi.org/10.1145/2330667.2330687
34. Doug A. Bowman, Christopher J. Rhoton, and Marcio S. Pinho. 2002. Text input techniques for immersive virtual environments: An empirical comparison. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2154–2158.
35. Marybeth Brown and Eileen M Hasser. 1996. Complexity of age-related change in skeletal muscle. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 51, 2: B117–B123.
36. Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 981–992. https://doi.org/10.1145/2858036.2858498
37. Karissa L. Canning, Ruth E. Brown, Veronica K. Jamnik, Art Salmon, Chris I. Ardern, and Jennifer L. Kuk. 2014. Individuals Underestimate Moderate and Vigorous Intensity Physical Activity. *PLoS ONE* 9, 5: e97927. https://doi.org/10.1371/journal.pone.0097927
38. Chih-Hui Chang, Wu-Wen Pan, Li-Ya Tseng, and Thomas A. Stoffregen. 2012. Postural activity and motion sickness during video game play in children and

adults. *Experimental Brain Research* 217, 2: 299–309. https://doi.org/10.1007/s00221-011-2993-4
39. Yu-Kai Chang, Jeff Labban, Jennifer Gapin, and Jennifer Etnier. 2012. The effects of acute exercise on cognitive performance: A meta-analysis. *Brain Research* 1453: 87–101. https://doi.org/10.1016/j.brainres.2012.02.068
40. Maher Chaouachi and Claude Frasson. 2010. Exploring the Relationship between Learner EEG Mental Engagement and Affect. In *Intelligent Tutoring Systems*, 291–293.
41. Salman Cheema, Michael Hoffman, and Joseph J. LaViola. 2013. 3D Gesture classification with linear acceleration and angular velocity sensing devices for video games. *Entertainment Computing* 4, 1: 11–24. https://doi.org/10.1016/j.entcom.2012.09.002
42. Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. 2017. A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications* 76, 3: 4405–4425. https://doi.org/10.1007/s11042-015-3177-1
43. Xiang “Anthony” Chen, Tovi Grossman, and George Fitzmaurice. 2014. Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST ’14*, 615–620. https://doi.org/10.1145/2642918.2647354
44. Dan Chiappe, Mark Conger, Janet Liao, J. Lynn Caldwell, and Kim-Phuong L. Vu. 2013. Improving multi-tasking ability through action videogames. *Applied Ergonomics* 44, 2: 278–284. https://doi.org/10.1016/j.apergo.2012.08.002
45. Luca Chittaro and Riccardo Sioni. 2019. Selecting Menu Items in Mobile Head-Mounted Displays: Effects of Selection Technique and Active Area. *International Journal of Human–Computer Interaction* 35, 16: 1501–1516. https://doi.org/10.1080/10447318.2018.1541546
46. Rory M.S. Clifford, Nikita Mae B. Tuanquin, and Robert W. Lindeman. 2017. Jedi ForceExtension: Telekinesis as a Virtual Reality interaction metaphor. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 239–240. https://doi.org/10.1109/3DUI.2017.7893360
47. Meta Company. Meta | Augmented Reality. Retrieved August 3, 2018 from https://www.metavision.com
48. Harrison Cook, Quang Vinh Nguyen, Simeon Simoff, Tomas Trescak, and Dean Preston. 2015. A Close-Range Gesture Interaction with Kinect. In *2015 Big Data Visual Analytics (BDVA)*, 1–8. https://doi.org/10.1109/BDVA.2015.7314284
49. Andrew Crossan, Stephen Brewster, and Alexander Ng. 2010. Foot Tapping for Mobile Interaction. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*, 418–422. https://doi.org/DOI: 10.14236/ewic/HCI2010.49
50. Justin Cuaresma and I. Scott MacKenzie. 2013. A Study of Variations of QWERTY Soft Keyboards for Mobile Phones. In *International Conference on Multimedia and Human Computer Interaction*, 8.
51. Jian Cui, Arjan Kuijper, Dieter W. Fellner, and Alexei Sourin. 2016. Understanding People’s Mental Models of Mid-Air Interaction for Virtual Assembly and Shape Modeling. In *Proceedings of the 29th International Conference on Computer Animation and Social Agents* (CASA ’16), 139–146. https://doi.org/10.1145/2915926.2919330
52. Jian Cui and Alexei Sourin. 2014. Feasibility Study on Free Hand Geometric Modelling Using Leap Motion in VRML/X3D. In *2014 International Conference on Cyberworlds*, 389–392. https://doi.org/10.1109/CW.2014.60

53. Meghal Dani, Gaurav Garg, Ramakrishna Perla, and Ramya Hebbalaguppe. 2018. Mid-Air Fingertip-Based User Interaction in Mixed Reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 174–178. https://doi.org/10.1109/ISMAR-Adjunct.2018.00061
54. Matthew Davis, Joseph Gabbard, Doug Bowman, and Dennis Gracanin. 2016. Depth-based 3D gesture multi-level radial menu for virtual object manipulation. In *2016 IEEE Virtual Reality (VR)*, 169–170. https://doi.org/10.1109/VR.2016.7504707
55. Mark S. Dennison, A. Zachary Wisti, and Michael D'Zmura. 2016. Use of physiological signals to predict cybersickness. *Contains Special Issue Articles – Proceedings of the 4th Symposium on Liquid Crystal Photonics (SLCP 2015)* 44: 42–52. https://doi.org/10.1016/j.displa.2016.07.002
56. Morgan Dixon, François Guimbretière, and Nicholas Chen. 2008. Optimal Parameters for Efficient Crossing-Based Dialog Boxes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), 1623–1632. https://doi.org/10.1145/1357054.1357307
57. P. DiZio and J. Lackner. 1997. Circumventing Side Effects of Immersive Virtual Environments. In *Design of Computing Systems: Social and Ergonomic Considerations*. Elsevier, San Francisco, California, USA, 893–896. Retrieved from http://dblp. uni-trier.de/db/conf/hci/hci1997-2.html
58. Michael Dorneich, Stephen Whitlow, Patricia Ververs, Jim Carciofini, and Janet Creaser. 2004. Closing the Loop of an Adaptive System with Cognitive State. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 48: 590–594. https://doi.org/10.1177/154193120404800367
59. Paul Dourish. 2004. *Where the Action Is: The Foundations of Embodied Interaction*. The MIT Press, Cambridge, Mass.
60. Pierre Dragicevic. 2004. Combining Crossing-Based and Paper-Based Interaction Paradigms for Dragging and Dropping between Overlapping Windows. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (UIST '04), 193–196. https://doi.org/10.1145/1029632.1029667
61. M. H. Draper, E. S. Viire, T. A. Furness, and V. J. Gawron. 2001. Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human factors* 43, 1: 129–146. https://doi.org/10.1518/001872001775992552
62. Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the Computer Using Gaze Gestures. In *Human-Computer Interaction – INTERACT 2007*, Cécilia Baranauskas, Philippe Palanque, Julio Abascal and Simone Diniz Junqueira Barbosa (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 475–488. https://doi.org/10.1007/978-3-540-74800-7_43
63. Paul E. Dux, Michael N. Tombu, Stephenie Harrison, Baxter P. Rogers, Frank Tong, and René Marois. 2009. Training Improves Multitasking Performance by Increasing the Speed of Information Processing in Human Prefrontal Cortex. *Neuron* 63, 1: 127–138. https://doi.org/10.1016/j.neuron.2009.06.005
64. Natalia Dużmańska, Paweł Strojny, and Agnieszka Strojny. 2018. Can Simulator Sickness Be Avoided? A Review on Temporal Aspects of Simulator Sickness. *Frontiers in psychology* 9: 2132–2132. https://doi.org/10.3389/fpsyg.2018.02132
65. Paula Edwards, Leon Barnard, Kathlene Leonard, Ji soo Yi, Kevin Moloney, Thitima Kongnakorn, Julie Jacko, and Francois Sainfort. 2005. Understanding

users with Diabetic Retinopathy: factors that affect performance in a menu selection task. *Behaviour & Information Technology* 24, 3: 175–186. https://doi.org/10.1080/01449290512331323189

66. Neven A.M. ElSayed, Bruce H. Thomas, Kim Marriott, Julia Piantadosi, and Ross T. Smith. 2016. Situated Analytics: Demonstrating immersive analytical tools with Augmented Reality. *Journal of Visual Languages & Computing* 36: 13–23. https://doi.org/10.1016/j.jvlc.2016.07.006
67. Barrett Ens, Ahmad Byagowi, Teng Han, Juan David Hincapié-Ramos, and Pourang Irani. 2016. 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*, 99–102. https://doi.org/10.1145/2983310.2985757
68. Anthony Faiola and Hadi Kharrazi. 2010. Diabetes education and serious gaming: Teaching adolescents to cope with diabetes. In *Health Informatics. A patient centered approach to diabetes*. MIT Press, Cambridge, Massachusetts, 151–177.
69. Kevin Fan, Liwei Chan, Daiya Kato, Kouta Minamizawa, and Masahiko Inami. 2016. VR Planet: Interface for Meta-View and Feet Interaction of VR Contents. In *ACM SIGGRAPH 2016 VR Village* (SIGGRAPH '16). https://doi.org/10.1145/2929490.2931001
70. Mingming Fan, Yizheng Ding, Fang Shen, Yuhui You, and Zhi Yu. 2017. An Empirical Study of Foot Gestures for Hands-Occupied Mobile Interaction. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers* (ISWC '17), 172–173. https://doi.org/10.1145/3123021.3123043
71. Wenxin Feng, Ming Chen, and Margrit Betke. 2014. Target reverse crossing: a selection method for camera-based mouse-replacement systems. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, 39.
72. Daniel Ferreira, Rut Correia, Antonieta Nieto, Alejandra Machado, Yaiza Molina Rodríguez, and José Barroso. 2015. Cognitive decline before the age of 50 can be detected with sensitive cognitive measures. *Psicothema* 27, 3: 216–222. https://doi.org/10.7334/psicothema2014.192
73. Samantha Finkelstein and Evan A. Suma. 2011. Astrojumper: Motivating Exercise with an Immersive Virtual Reality Exergame. *Presence: Teleoperators and Virtual Environments* 20, 1: 78–92. https://doi.org/10.1162/pres_a_00036
74. Paul M. Fitts. 1992. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General* 121, 3: 262.
75. Frederick G Freeman, Peter J Mikulka, Lawrence J Prinzel, and Mark W Scerbo. 1999. Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological Psychology* 50, 1: 61–76. https://doi.org/10.1016/S0301-0511(99)00002-2
76. Koumei Fukahori, Daisuke Sakamoto, and Takeo Igarashi. 2015. Exploring Subtle Foot Plantar-Based Gestures with Sock-Placed Pressure Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 3019–3028. https://doi.org/10.1145/2702123.2702308
77. Masaaki Fukumoto. 2018. SilentVoice: Unnoticeable Voice Input by Ingressive Speech. In *Proceedings of the 31st Annual ACM Symposium on User Interface*

Software and Technology (UIST '18), 237–246. https://doi.org/10.1145/3242587.3242603
78. Joseph Gabbard, J. Edward Swan, Deborah Hix, Si-Jung Kim, and Greg Fitch. 2007. Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications. In *2007 IEEE Virtual Reality Conference*, 35–42. https://doi.org/10.1109/VR.2007.352461
79. Joseph Gabbard, J. Edward Swan, Deborah Hix, Robert Schulman, John Lucas, and Divya Gupta. 2005. An empirical user-based study of text drawing styles and outdoor background textures for augmented reality. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, 11–18. https://doi.org/10.1109/VR.2005.1492748
80. Yue Gao and Regan L. Mandryk. 2011. GrabApple: The Design of a Casual Exergame. In *Entertainment Computing – ICEC 2011*, Junia Coutinho Anacleto, Sidney Fels, Nicholas Graham, Bill Kapralos, Magy Saif El-Nasr and Kevin Stanley (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 35–46. https://doi.org/10.1007/978-3-642-24500-8_5
81. Wolf Garbe. 2018. *SymSpell: 1 million times faster through Symmetric Delete spelling correction algorithm*. Retrieved April 1, 2018 from https://github.com/wolfgarbe/SymSpell
82. Jaime A. Garcia, Daniel Schoene, Stephen R. Lord, Kim Delbaere, Trinidad Valenzuela, and Karla Felix Navarro. 2016. A Bespoke Kinect Stepping Exergame for Improving Physical and Cognitive Function in Older People: A Pilot Study. *Games for Health Journal* 5, 6: 382–388. https://doi.org/10.1089/g4h.2016.0070
83. Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. 2012. Integrating the Physical Environment into Mobile Remote Collaboration. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '12), 241–250. https://doi.org/10.1145/2371574.2371610
84. Alireza Mazloumi Gavgani, Keith V. Nesbitt, Karen L. Blackmore, and Eugene Nalivaiko. 2017. Profiling subjective symptoms and autonomic changes associated with cybersickness. *Autonomic Neuroscience* 203: 41–50. https://doi.org/10.1016/j.autneu.2016.12.004
85. Kathrin Gerling, Ian Livingston, Lennart Nacke, and Regan Mandryk. 2012. Full-body motion-based game interaction for older adults. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 1873. https://doi.org/10.1145/2207676.2208324
86. Kathrin M. Gerling, Regan L. Mandryk, and Conor Linehan. 2015. Long-Term Use of Motion-Based Video Games in Care Home Settings. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '15*, 1573–1582. https://doi.org/10.1145/2702123.2702125
87. Peter J. Gianaros, Eric R. Muth, J. Toby Mordkoff, Max E. Levine, and Robert M. Stern. 2001. A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness. *Aviation, space, and environmental medicine* 72, 2: 115–119.
88. Yulia Gizatdinova, Oleg Špakov, and Veikko Surakka. 2012. Comparison of video-based pointing and selection techniques for hands-free text entry. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 132–139.
89. Fabian Göbel, Konstantin Klamka, Andreas Siegel, Stefan Vogt, Sophie Stellmach, and Raimund Dachselt. 2013. Gaze-Supported Foot Interaction in

Zoomable Information Spaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '13), 3059–3062. https://doi.org/10.1145/2468356.2479610

90. Mayank Goel, Leah Findlater, and Jacob Wobbrock. 2012. WalkType: Using accelerometer data to accommodate situational impairments in mobile touch screen text entry. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/2207676.2208662
91. David F. Goldspink. 2005. Ageing and activity: their effects on the functional reserve capacities of the heart and vascular smooth and skeletal muscles. *Ergonomics* 48, 11–14: 1334–1351. https://doi.org/10.1080/00140130500101247
92. Daniil Goncharov. 2018. *Implement the same algorithm "swype keyboard" for .NET and Unity.: Neargye/SwipeType*. Retrieved September 10, 2018 from https://github.com/Neargye/SwipeType
93. Jun Gong, Zheer Xu, Qifan Guo, Teddy Seyed, Xiang "Anthony" Chen, Xiaojun Bi, and Xing-Dong Yang. 2018. WrisText: One-handed Text Entry on Smartwatch using Wrist Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. https://doi.org/10.1145/3173574.3173755
94. Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. 2017. Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17*, 553–563. https://doi.org/10.1145/3126594.3126615
95. Gabriel González, José P. Molina, Arturo S. García, Diego Martínez, and Pascual González. 2009. Evaluation of Text Input Techniques in Immersive Virtual Environments. In *New Trends on Human-Computer Interaction*, José A. Macías, Antoni Granollers Saltiveri and Pedro M. Latorre (eds.). Springer London, London, 1–10. https://doi.org/10.1007/978-1-84882-352-5_11
96. Gabriel González, José P. Molina, Arturo S. García, Diego Martínez, and Pascual González. 2009. Evaluation of Text Input Techniques in Immersive Virtual Environments. In *New Trends on Human-Computer Interaction*, José A. Macías, Antoni Granollers Saltiveri and Pedro M. Latorre (eds.). Springer London, London, 1–10. https://doi.org/10.1007/978-1-84882-352-5_11
97. 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*, 194–195.
98. Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 159–166. https://doi.org/10.1109/VR.2018.8446059
99. Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. Effects of Hand Representations for Typing in Virtual Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 151–158. https://doi.org/10.1109/VR.2018.8446250
100. Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 159–166. https://doi.org/10.1109/VR.2018.8446059

101. 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 - UIST '16*, 49–60. https://doi.org/10.1145/2984511.2984576
102. Jože Guna, Gregor Geršak, Iztok Humar, Jeungeun Song, Janko Drnovšek, and Matevz Pogacnik. 2018. Influence of video content type on users' virtual reality sickness perception and physiological response. *Future Generation Computer Systems* 91. https://doi.org/10.1016/j.future.2018.08.049
103. J. Hakkinen, T. Vuori, and M. Paakka. 2002. Postural stability and sickness symptoms after HMD use. In *IEEE International Conference on Systems, Man and Cybernetics*, 147–152. https://doi.org/10.1109/ICSMC.2002.1167964
104. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11, 1: 10. https://doi.org/10.1145/1656274.1656278
105. Teng Han, Jason Alexander, Abhijit Karnik, Pourang Irani, and Sriram Subramanian. 2011. Kick: investigating the use of kick gestures for mobile interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, 29. https://doi.org/10.1145/2037373.2037379
106. John Paulin Hansen, Kristian Tørning, Anders Sewerin Johansen, Kenji Itoh, and Hirotaka Aoki. 2004. Gaze typing compared with input by head and hand. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, 131–138.
107. Sandra G. Hart. 1986. NASA Task load Index (TLX). Volume 1.0; Paper and pencil package.
108. Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9: 904–908. https://doi.org/10.1177/154193120605000909
109. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Elsevier, 139–183.
110. Khalad Hasan, David Ahlström, and Pourang Irani. 2013. Ad-Binning: Leveraging around Device Space for Storing, Browsing and Retrieving Mobile Device Content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 899–908. https://doi.org/10.1145/2470654.2466115
111. Gunther Heidemann, Ingo Bax, and Holger Bekel. 2004. Multimodal interaction in an augmented reality scenario. In *Proceedings of the 6th International Conference on Multimodal Interfaces*. https://doi.org/DOI: 10.1145/1027933.1027944
112. Hamilton A. Hernandez, Zi Ye, T.C. Nicholas Graham, Darcy Fehlings, and Lauren Switzer. 2013. Designing action-based exergames for children with cerebral palsy. In *SIGCHI Conference on Human Factors in Computing Systems*, 1261–1270. https://doi.org/10.1145/2470654.2466164
113. Imali T. Hettiarachchi, Samer Hanoun, Darius Nahavandi, and Saeid Nahavandi. 2019. Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. *PLOS ONE* 14, 5: e0217288. https://doi.org/10.1371/journal.pone.0217288

114. 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 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 1063–1072. https://doi.org/10.1145/2556288.2557130
115. Karen Ho and Hanley Weng. 2013. Favoured attributes of in-air gestures in the home environment. In *Proceedings of the 25th Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration - OzCHI '13*, 171–174. https://doi.org/10.1145/2541016.2541095
116. Daniel Horodniczy and Jeremy R. Cooperstock. 2017. Free the Hands! Enhanced Target Selection via a Variable-Friction Shoe. 255–259. https://doi.org/10.1145/3025453.3025625
117. Lode Hoste, Bruno Dumas, and Beat Signer. 2012. SpeeG: a multimodal speech- and gesture-based text input solution. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, 156. https://doi.org/10.1145/2254556.2254585
118. Bin Hu and Jiacun Wang. 2018. Deep Learning Based Hand Gesture Recognition and UAV Flight Controls. In *2018 24th International Conference on Automation and Computing (ICAC)*, 1–6. https://doi.org/10.23919/IConAC.2018.8748953
119. Han-Chung Huang, Huynh Van Nguyen, T.C.E. Cheng, May-Kuen Wong, Hsin-Ying Chiu, Ya-Hui Yang, and Ching-I Teng. 2019. A Randomized Controlled Trial on the Role of Enthusiasm About Exergames: Players' Perceptions of Exercise. *Games for Health Journal* 8, 3: 220–226. https://doi.org/10.1089/g4h.2018.0057
120. Anke Huckauf and Mario H. Urbina. 2008. Gazing with pEYEs: towards a universal input for various applications. In *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08*, 51. https://doi.org/10.1145/1344471.1344483
121. Andrew Hutchison. 2007. Back to the Holodeck: New Life for Virtual Reality? In *Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts* (DIMEA '07), 98–104. https://doi.org/10.1145/1306813.1306838
122. 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 - ETRA '12*, 229. https://doi.org/10.1145/2168556.2168602
123. Wijnand IJsselsteijn, Yvonne De Kort, and Karolien Poels. 2008. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven*.
124. Christos Ioannou, Patrick Archard, Eamonn O'Neill, and Christof Lutteroth. 2019. Virtual Performance Augmentation in an Immersive Jump & Run Exergame. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–15. https://doi.org/10.1145/3290605.3300388
125. Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. 2010. Designing gaze gestures for gaming: an investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*, 323. https://doi.org/10.1145/1743666.1743740
126. 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 Empowering people - CHI '90*, 11–18. https://doi.org/10.1145/97243.97246

127. Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2: 152–169. https://doi.org/10.1145/123078.128728
128. Varun Jain, Gaurav Garg, Ramakrishna Perla, and Ramya Hebbalaguppe. 2019. GestARLite: An On-Device Pointing Finger Based Gestural Interface for Smartphones and Video See-Through Head-Mounts. https://doi.org/arXiv:1904.09843
129. Michael J. Kahana, Robert Sekuler, Jeremy B. Caplan, Matthew Kirschen, and Joseph R. Madsen. 1999. Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature* 399, 6738: 781–784. https://doi.org/10.1038/21645
130. Keiko Katsuragawa, James R. Wallace, and Edward Lank. 2016. Gestural Text Input Using a Smartwatch. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*, 220–223. https://doi.org/10.1145/2909132.2909273
131. Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3: 203–220. https://doi.org/10.1207/s15327108ijap0303_3
132. Behrang Keshavarz and Heiko Hecht. 2011. Axis rotation and visually induced motion sickness: the role of combined roll, pitch, and yaw motion. *Aviation, space, and environmental medicine* 82, 11: 1023–1029. https://doi.org/10.3357/asem.3078.2011
133. Caroline Ketcham and George Stelmach. 2001. Age-related declines in motor control. In *Handbook of the Psychology of Aging* (5th ed.). Academic Press, San Diego and London, 267–287.
134. Sehoon Kim, Seungheon Lee, Nupur Kala, Jaesung Lee, and Wonhee Choe. 2018. An effective FoV restriction approach to mitigate VR sickness on mobile devices. *Journal of the Society for Information Display* 26, 6: 376–384. https://doi.org/10.1002/jsid.669
135. Young Youn Kim, Eun Nam Kim, Min Jae Park, Kwang Suk Park, Hee Dong Ko, and Hyun Taek Kim. 2008. The Application of Biosignal Feedback for Reducing Cybersickness from Exposure to a Virtual Environment. *Presence* 17, 1: 1–16. https://doi.org/10.1162/pres.17.1.1
136. Youngwon R. Kim and Gerard J. Kim. 2017. HoVR-Type: Smartphone as a typing interface in VR using hovering. In *Consumer Electronics (ICCE), 2017 IEEE International Conference on*, 200–203.
137. Matthew J. Klein and Christina S. Simmers. 2009. Exergaming: virtual inspiration, real perspiration. *Young Consumers* 10, 1: 35–45. https://doi.org/10.1108/17473610910940774
138. Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–9. https://doi.org/10.1145/3173574.3173919
139. Silvia Erika Kober and Christa Neuper. 2011. Sex differences in human EEG theta oscillations during spatial navigation in virtual reality. *International Journal of Psychophysiology* 79, 3: 347–355. https://doi.org/10.1016/j.ijpsycho.2010.12.002

140. Panayiotis Koutsabasis and Panagiotis Vogiatzidakis. 2019. Empirical Research in Mid-Air Interaction: A Systematic Review. *International Journal of Human-Computer Interaction*: 1–22. https://doi.org/10.1080/10447318.2019.1572352
141. Rumi Kozakai, Fujiko Ando, Heung Kim, Atsumu Yuki, Rei Otsuka, and Hiroshi Shimokata. 2016. Sex-differences in age-related grip strength decline: A 10-year longitudinal study of community-living middle-aged and older Japanese. *The Journal of Physical Fitness and Sports Medicine* 5, 1: 87–94. https://doi.org/10.7600/jpfsm.5.87
142. Per Ola Kristensson and Keith Vertanen. 2012. The potential of dwell-free eye-typing for fast assistive gaze communication. In *Proceedings of the symposium on eye tracking research and applications*, 241–244.
143. Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. 1985. VIDEOPLACE—an Artificial Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’85), 35–40. https://doi.org/10.1145/317456.317463
144. Ernst Kruijff, J. Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, 3–12. https://doi.org/10.1109/ISMAR.2010.5643530
145. Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos Morimoto, and Margrit Betke. 2016. EyeSwipe: Dwell-free Text Entry Using Gaze Paths. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI ’16*, 1952–1956. https://doi.org/10.1145/2858036.2858335
146. Gordon Kurtenbach and William Buxton. 1993. The limits of expert performance using hierarchic marking menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’93*, 482–487. https://doi.org/10.1145/169059.169426
147. 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 - CHI ’18*, 1–14. https://doi.org/10.1145/3173574.3173655
148. Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. 1–14. https://doi.org/10.1145/3173574.3173655
149. Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6
150. Joseph J. LaViola. 2000. A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bull.* 32, 1: 47–56. https://doi.org/10.1145/333329.333344
151. I-Min Lee, Eric J Shiroma, Felipe Lobelo, Pekka Puska, Steven N Blair, and Peter T Katzmarzyk. 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet* 380, 9838: 219–229. https://doi.org/10.1016/S0140-6736(12)61031-9
152. Suk Jin Lee, Yuichi Motai, and Hongsik Choi. 2013. Tracking Human Motion With Multichannel Interacting Multiple Model. *Industrial Informatics, IEEE Transactions on* 9: 1751–1763. https://doi.org/10.1109/TII.2013.2257804
153. Frank Chun Yat Li, David Dearman, and Khai N. Truong. 2009. Virtual Shelves: Interactions with Orientation Aware Devices. In *Proceedings of the 22nd Annual*

ACM Symposium on User Interface Software and Technology (UIST ’09), 125–128. https://doi.org/10.1145/1622176.1622200
154. Chiuhsiang Joe Lin, Sui-Hua Ho, and Yan-Jyun Chen. 2015. An investigation of pointing postures in a 3D stereoscopic environment. *Applied Ergonomics* 48: 154–163. https://doi.org/10.1016/j.apergo.2014.12.001
155. Jia-Wei Lin, Ping-Hsuan Han, Jiun-Yu Lee, Yang-Sheng Chen, Ting-Wei Chang, Kuan-Wen Chen, and Yi-Ping Hung. 2017. Visualizing the keyboard in virtual reality for enhancing immersive experience. In *ACM SIGGRAPH 2017 Posters on - SIGGRAPH ’17*, 1–2. https://doi.org/10.1145/3102163.3102175
156. Alfredo Liverani, Giancarlo Amati, and Gianni Caligiana. 2004. A CAD-augmented Reality Integrated Environment for Assembly Sequence Check and Interactive Validation. *Concurrent Engineering* 12, 1: 67–77. https://doi.org/10.1177/1063293X04042469
157. Gerard Llorach, Alun Evans, and Josep Blat. 2014. Simulator Sickness and Presence Using HMDs: Comparing Use of a Game Controller and a Position Estimation System. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (VRST ’14), 137–140. https://doi.org/10.1145/2671015.2671120
158. Xueshi Lu, Difeng Yu, Hai-Ning Liang, Wenge Xu, Yuzheng Chen, Xiang Li, and Khalad Hasan. 2020. Exploration of Hands-free Text Entry Techniques For Virtual Reality. In *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 6. https://doi.org/In press
159. Yiqin Lu, Chun Yu, Xin Yi, Yuanchun Shi, and Shengdong Zhao. 2017. BlindType: Eyes-Free Text Entry on Handheld Touchpad by Leveraging Thumb’s Muscle Memory. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2: 1–24. https://doi.org/10.1145/3090083
160. Yuexing Luo and Daniel Vogel. 2014. Crossing-Based Selection with Direct Touch Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’14), 2627–2636. https://doi.org/10.1145/2556288.2557397
161. Zhihan Lv. 2013. Wearable Smartphone: Wearable Hybrid Framework for Hand and Foot Gesture Interaction on Smartphone. In *Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops* (ICCVW ’13), 436–443. https://doi.org/10.1109/ICCVW.2013.64
162. Zhihan Lv, Alaa Halawani, Shengzhong Feng, Haibo Li, and Shafiq Réhman. 2014. Multimodal Hand and Foot Gesture Interaction for Handheld Devices. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11: 1–19. https://doi.org/10.1145/2645860
163. Zhihan Lv, Alaa Halawani, Shengzhong Feng, Shafiq ur Réhman, and Haibo Li. 2015. Touch-less interactive augmented reality game on vision-based wearable device. *Personal and Ubiquitous Computing* 19, 3: 551–567. https://doi.org/10.1007/s00779-015-0844-1
164. Kent Lyons, Thad Starner, Daniel Plaisted, James Fusia, Amanda Lyons, Aaron Drew, and E. W. Looney. 2004. Twiddler typing: one-handed chording text entry for mobile phones. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI ’04*, 671–678. https://doi.org/10.1145/985692.985777
165. M. Reyes, G. Domínguez, and S. Escalera. 2011. Featureweighting in dynamic timewarping for gesture recognition in depth data. In *2011 IEEE International*

Conference on Computer Vision Workshops (ICCV Workshops), 1182–1188. https://doi.org/10.1109/ICCVW.2011.6130384

166. I. Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1: 91–139.
167. I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems*, 754–755.
168. Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 357–360.
169. Päivi Majaranta and Kari-Jouko Räihä. 2002. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, 15–22.
170. Shahzad Malik, Chris McDonald, and Gerhard Roth. 2002. Hand tracking for interactive pattern-based augmented reality. In *Proceedings. International Symposium on Mixed and Augmented Reality*, 117–126. https://doi.org/10.1109/ISMAR.2002.1115080
171. Shahzad Malik, Chris McDonald, and Gerhard Roth. 2002. Hand tracking for interactive pattern-based augmented reality. *International Symposium on Mixed and Augmented Reality, ISMAR 2002*: 117–126. https://doi.org/10.1109/ISMAR.2002.1115080
172. Laura Malinverni, Marie-Monique Schaper, and Narcis Pares. 2018. Multimodal methodological approach for participatory design of Full-Body Interaction Learning Environments. *Qualitative Research* 19, 1: 71–89. https://doi.org/10.1177/1468794118773299
173. Regan L. Mandryk, Kathrin M. Gerling, and Kevin G. Stanley. 2014. Designing Games to Discourage Sedentary Behaviour. In *Playful User Interfaces*, Anton Nijholt (ed.). Springer Singapore, Singapore, 253–274. https://doi.org/10.1007/978-981-4560-96-2_12
174. Jennifer Mankoff and Gregory D. Abowd. 1998. Cirrin: a word-level unistroke keyboard for pen input. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, 213–214.
175. Anders Markussen, Mikkel R. Jakobsen, and Kasper Hornbæk. 2013. Selection-Based Mid-Air Text Entry on Large Displays. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson and Marco Winckler (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 401–418. https://doi.org/10.1007/978-3-642-40483-2_28
176. Anna Lisa Martin-Niedecken, Andrea Mahrer, Katja Rogers, Eling D. de Bruin, and Alexandra Schättin. 2020. "HIIT" the ExerCube: Comparing the Effectiveness of Functional High-Intensity Interval Training in Conventional vs. Exergame-Based Training. *Frontiers in Computer Science* 2: 33. https://doi.org/10.3389/fcomp.2020.00033
177. Denys J. C. Matthies, Franz Müller, Christoph Anthes, and Dieter Kranzlmüller. 2013. ShoeSoleSense: Proof of Concept for a Wearable Foot Interface for Virtual and Real Environments. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology* (VRST '13), 93–96. https://doi.org/10.1145/2503713.2503740
178. Eliana Medina, Ruth Fruland, and Suzanne Weghorst. 2008. Virtusphere: Walking in a Human Size VR "'Hamster Ball.'" In *Human Factors and*

Ergonomics Society Annual Meeting. https://doi.org/DOI: 10.1177/154193120805202704

179. Daniel R. Mestre, Virginie Dagonneau, and Charles-Symphorien Mercier. 2011. Does Virtual Reality Enhance Exercise Performance, Enjoyment, and Dissociation? An Exploratory Study on a Stationary Bike Apparatus. *Presence: Teleoperators and Virtual Environments* 20, 1: 14. https://doi.org/10.1162/pres_a_00031
180. Microsoft. Microsoft HoloLens. *Microsoft HoloLens*. Retrieved September 14, 2018 from https://www.microsoft.com/en-us/hololens
181. Peter J. Mikulka, Mark W. Scerbo, and Frederick G. Freeman. 2002. Effects of a Biocybernetic System on Vigilance Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 44, 4: 654–664. https://doi.org/10.1518/0018720024496944
182. Mark R. Mine. 1995. *Virtual Environment Interaction Techniques*. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
183. Brunelli Pinto Miranda, Nikolas Jorge Santiago Carneiro, Tiago Davi Oliveira de Araújo, Carlos Gustavo Resque dos Santos, Alexandre Abreu de Freitas, Jefferson Magalhães de Morais, and Bianchi Serique Meiguins. 2016. Categorizing issues in mid-air infovis interaction. In *2016 20th International Conference Information Visualisation (IV)*, 242–246. https://doi.org/10.1109/IV.2016.70
184. Pranav Mistry and Pattie Maes. 2009. SixthSense: a wearable gestural interface. In *International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ASIA*. https://doi.org/DOI: 10.1145/1667146.1667160
185. Javier Monedero, Elizabeth J. Lyons, and Donal J. O’Gorman. 2015. Interactive Video Game Cycling Leads to Higher Energy Expenditure and Is More Enjoyable than Conventional Exercise in Adults. *PLOS ONE* 10, 3: e0118470. https://doi.org/10.1371/journal.pone.0118470
186. Diego Monteiro, Hai-Ning Liang, Andrew Abel, Nilufar Bahaei, and Rita de Cassia Monteiro. 2018. Evaluating Engagement of Virtual Reality Games Based on First and Third Person Perspective Using EEG and Subjective Metrics. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 53–60. https://doi.org/10.1109/AIVR.2018.00015
187. Diego Monteiro, Hai-Ning Liang, Wenge Xu, Marvin Brucker, Vijayakumar Nanjappan, and Yong Yue. 2018. Evaluating enjoyment, presence, and emulator sickness in VR games based on first- and third- person viewing perspectives. *Computer Animation and Virtual Worlds* 29, 3: 12. https://doi.org/10.1002/cav.1830
188. Diego Monteiro, Hai-Ning Liang, Yuxuan Zhao, and Andrew Abel. 2018. Comparing Event Related Arousal-Valence and Focus Among Different Viewing Perspectives in VR Gaming. In *Advances in Brain Inspired Cognitive Systems* (Lecture Notes in Computer Science), 770–779.
189. Diego Monteiro, Hai-Ning Liang, Yuxuan Zhao, and Andrew Abel. 2018. Comparing Event Related Arousal-Valence and Focus Among Different Viewing Perspectives in VR Gaming. In *Advances in Brain Inspired Cognitive Systems*, 770–779.
190. Ditte H Mortensen, Søren Bech, Durand R Begault, and Bernard D Adelstein. 2009. The Relative Importance of Visual, Auditory, and Haptic Information for the User’s Experience of Mechanical Switches. *Perception* 38, 10: 1560–1571. https://doi.org/10.1068/p5929

191. Martez E. Mott, Shane Williams, Jacob O. Wobbrock, and Meredith Ringel Morris. 2017. Improving Dwell-Based Gaze Typing with Dynamic, Cascading Dwell Times. 2558–2570. https://doi.org/10.1145/3025453.3025517
192. Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. 2019. Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. https://doi.org/10.1145/3290605.3300707
193. Lennart E. Nacke. 2010. Wiimote vs. controller: electroencephalographic measurement of affective gameplay interaction. In *Proceedings of the International Academic Conference on the Future of Game Design and Technology - Futureplay '10*, 159. https://doi.org/10.1145/1920778.1920801
194. Takashi Nakamura, Shin Takahashi, and Jiro Tanaka. 2008. Double-Crossing: A New Interaction Technique for Hand Gesture Interfaces. In *Computer-Human Interaction*, 292–300.
195. Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, 177. https://doi.org/10.1145/1978942.1978969
196. Juliet Norton, Chadwick A. Wingrave, and Joseph J. LaViola. 2010. Exploring strategies and guidelines for developing full body video game interfaces. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*, 155–162. https://doi.org/10.1145/1822348.1822369
197. Ju Young Oh, Jun Lee, Joong Ho Lee, and Ji Hyung Park. 2017. AnywhereTouch: Finger Tracking Method on Arbitrary Surface Using Nailed-Mounted IMU for Mobile HMD. In *International Conference on Human-Computer Interaction*, 185–191.
198. Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2013. Dynamic Text Management for See-through Wearable and Heads-up Display Systems. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces* (IUI '13), 363–370. https://doi.org/10.1145/2449396.2449443
199. Francisco Ortega, Katherine Tarre, Mathew Kress, Adam Williams, Armando Barreto, and Naphtali 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)*, 1723–1728. https://doi.org/10.1109/VR.2019.8798105
200. Yechiam Ostchega, Kathryn Porter, Jeffery Hughes, Charles F Dillon, and Tatiana Nwankwo. 2011. Resting pulse rate reference data for children, adolescents, and adults: United States, 1999-2008. *National health statistics reports* 41: 1–16.
201. Joseph A Paradiso, Kai-Yuh Hsiao, and Ari Benbasat. Interfacing to the Foot: Apparatus and Applications. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, 2. https://doi.org/10.1145/633292.633389
202. Hyung Min Park, Seok Han Lee, and Jong Soo Choi. 2008. Wearable augmented reality system using gaze interaction. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, 175–176. https://doi.org/10.1109/ISMAR.2008.4637353
203. Diogo Pedrosa, Maria Da Graça Pimentel, Amy Wright, and Khai N. Truong. 2015. Filteryedping: Design Challenges and User Performance of Dwell-Free

Eye Typing. *ACM Transactions on Accessible Computing* 6, 1: 1–37. https://doi.org/10.1145/2724728

204. Wei Peng, Jih-Hsuan Tammy Lin, and Julia Crouse Waddell. 2011. Is Playing Exergames Really Exercising? A Meta-Analysis of Energy Expenditure in Active Video Games. *Cyberpsychology, behavior and social networking* 14: 681–8. https://doi.org/10.1089/cyber.2010.0578
205. Sebastian Pick, Andrew S. Puika, and Torsten W. Kuhlen. 2016. SWIFTER: Design and evaluation of a speech-based text input metaphor for immersive virtual environments. In *3D User Interfaces (3DUI), 2016 IEEE Symposium on*, 109–112.
206. Thomas G. Plante, Arianna Aldridge, Denise Su, Ryan Bogdan, Martha Belo, and Kamran Kahn. 2003. Does Virtual Reality Enhance the Management of Stress When Paired With Exercise? An Exploratory Study. *International Journal of Stress Management* 10, 3: 203–216. https://doi.org/10.1037/1072-5245.10.3.203
207. Alessandra Pluchino, Sae Yong Lee, Shihab Asfour, Bernard A. Roos, and Joseph F. Signorile. 2012. Pilot Study Comparing Changes in Postural Control After Training Using a Video Game Balance Board Program and 2 Standard Activity-Based Balance Intervention Programs. *Archives of Physical Medicine and Rehabilitation* 93, 7: 1138–1146. https://doi.org/10.1016/j.apmr.2012.01.023
208. Alan T Pope, Edward H Bogart, and Debbie S Bartolome. 1995. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology* 40, 1–2: 187–195. https://doi.org/10.1016/0301-0511(95)05116-3
209. Marco Porta. 2015. A study on text entry methods based on eye gestures. *Journal of Assistive Technologies* 9, 1: 48–67. https://doi.org/10.1108/JAT-12-2013-0037
210. Marco Porta and Matteo Turina. 2008. Eye-S: a full-screen input modality for pure eye-based communication. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, 27–34.
211. Sara Price, Mona Sakr, and Carey Jewitt. 2016. Exploring Whole-Body Interaction and Design for Museums. *Interacting with Computers* 28, 5: 569–583. https://doi.org/10.1093/iwc/iwv032
212. Morten Proschowsky, Nette Schultz, and Niels Ebbe Jacobsen. 2006. An intuitive text input method for touch wheels. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 467–470.
213. Kun Qian, Chenshu Wu, Zimu Zhou, Yue Zheng, Zheng Yang, and Yunhao Liu. 2017. Inferring Motion Direction using Commodity Wi-Fi for Interactive Exergames. 1961–1972. https://doi.org/10.1145/3025453.3025678
214. Jagdish Lal Raheja, Radhey Shyam, Umesh Kumar, and Bhanu Prasad. 2010. Real-Time Robotic Hand Control Using Hand Gestures. In *2010 Second International Conference on Machine Learning and Computing*, 12–16. https://doi.org/10.1109/ICMLC.2010.12
215. Mahfuz Rahman, Sean Gustafson, Pourang Irani, and Sriram Subramanian. 2009. Tilt techniques: investigating the dexterity of wrist-based input. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 1943. https://doi.org/10.1145/1518701.1518997
216. Vijay Dandur Rajanna. 2016. Gaze and Foot Input: Toward a Rich and Assistive Interaction Modality. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces* (IUI ’16 Companion), 126–129. https://doi.org/10.1145/2876456.2876462

217. Niklas Ravaja. 2004. Contributions of Psychophysiology to Media Research: Review and Recommendations. *Media Psychology - MEDIA PSYCHOL* 6: 193–235. https://doi.org/10.1207/s1532785xmep0602_4
218. Lisa Rebenitsch and Charles Owen. 2016. Review on Cybersickness in Applications and Visual Displays. *Virtual Real.* 20, 2: 101–125. https://doi.org/10.1007/s10055-016-0285-9
219. Christian Reimann and Dirk Stichling. 2004. Kick-Up Menus. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1552–1552. https://doi.org/10.1145/985921.986128
220. Zhou Ren, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang. 2013. Robust Part-Based Hand Gesture Recognition Using Kinect Sensor. *IEEE transactions on multimedia* 15, 5: 1–11.
221. Zhou Ren, Junsong Yuan, and Zhengyou Zhang. 2011. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. *19th ACM international conference on Multimedia*, June 2014: 1093. https://doi.org/10.1145/2072298.2071946
222. Van Krevelen Rick and Ronald Poelman. 2010. A Survey of Augmented Reality Technologies, Applications and Limitations. *The International Journal of Virtual Reality* 9, 2: 1–20. https://doi.org/10.20870/IJVR.2010.9.2.2767
223. Albert Skip Rizzo, Belinda Lange, Evan A. Suma, and Mark Bolas. 2011. Virtual Reality and Interactive Digital Game Technology: New Tools to Address Obesity and Diabetes. *Journal of Diabetes Science and Technology* 5, 2: 256–264. https://doi.org/10.1177/193229681100500209
224. Dori Rosenberg, Colin A. Depp, Ipsit V. Vahia, Jennifer Reichstadt, Barton W. Palmer, Jacqueline Kerr, Greg Norman, and Dilip V. Jeste. 2010. Exergames for Subsyndromal Depression in Older Adults: A Pilot Study of a Novel Intervention. *The American Journal of Geriatric Psychiatry* 18, 3: 221–226. https://doi.org/10.1097/JGP.0b013e3181c534b5
225. Mikko Salminen and Niklas Ravaja. 2007. Oscillatory Brain Responses Evoked by Video Game Events: The Case of Super Monkey Ball 2. *CyberPsychology & Behavior* 10, 3: 330–338. https://doi.org/10.1089/cpb.2006.9947
226. Krystian Samp and Stefan Decker. 2010. Supporting Menu Design with Radial Layouts. In *Proceedings of the International Conference on Advanced Visual Interfaces* (AVI '10), 155–162. https://doi.org/10.1145/1842993.1843021
227. Krystian Samp and Stefan Decker. 2011. Visual Search in Radial Menus. In *Human-Computer Interaction – INTERACT 2011*, 248–255. https://doi.org/10.1007/978-3-642-23768-3_21
228. Mariann Sápi, Andrea Domján, Anna Fehérné Kiss, and Sándor Pintér. 2018. Is Kinect Training Superior to Conventional Balance Training for Healthy Older Adults to Improve Postural Control? *Games for Health Journal* 8, 1: 41–48. https://doi.org/10.1089/g4h.2018.0027
229. Sayan Sarcar, Prateek Panwar, and Tuhin Chakraborty. 2013. EyeK: an efficient dwell-free eye gaze-based text entry system. 215–220. https://doi.org/10.1145/2525194.2525288
230. Sayan Sarcar, Prateek Panwar, and Tuhin Chakraborty. 2013. EyeK: an efficient dwell-free eye gaze-based text entry system. In *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction - APCHI '13*, 215–220. https://doi.org/10.1145/2525194.2525288
231. Dimitrios Saredakis, Ancret Szpak, Brandon Birckhead, Hannah A. D. Keage, Albert Rizzo, and Tobias Loetscher. 2020. Factors Associated With Virtual

Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience* 14: 17. https://doi.org/10.3389/fnhum.2020.00096

232. William Saunders and Daniel Vogel. 2016. Tap-Kick-Click: Foot Interaction for a Standing Desk. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (DIS '16), 323–333. https://doi.org/10.1145/2901790.2901815
233. Christian Schönauer, Thomas Pintaric, and Hannes Kaufmann. 2011. Full Body Interaction for Serious Games in Motor Rehabilitation. In *Proceedings of the 2nd Augmented Human International Conference* (AH '11). https://doi.org/10.1145/1959826.1959830
234. Johannes Schöning, Florian Daiber, Antonio Krüger, and Michael Rohs. 2009. Using hands and feet to navigate and manipulate spatial data. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*, 4663. https://doi.org/10.1145/1520340.1520717
235. Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4: 40. https://doi.org/10.9781/ijimai.2017.445
236. Matt Schubert, Amy Clark, and Annie Rosa. 2018. The Polar OH1 Optical Heart Rate Sensor is Valid during Moderate-Vigorous Exercise. *Sports Medicine International Open* 2, 3: 67–70. https://doi.org/10.1055/a-0631-0920
237. Jeremy Scott, David Dearman, Koji Yatani, and Khai N. Truong. 2010. Sensing foot gestures from the pocket. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*, 199. https://doi.org/10.1145/1866029.1866063
238. María del Mar Seguí, Julio Cabrero-García, Ana Crespo, José Verdú, and Elena Ronda. 2015. A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. *Journal of Clinical Epidemiology* 68, 6: 662–673. https://doi.org/10.1016/j.jclinepi.2015.01.015
239. Sarah Sharples, Sue Cobb, Amanda Moody, and John R. Wilson. 2008. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 2: 58–69. https://doi.org/10.1016/j.displa.2007.09.005
240. Lindsay Alexander Shaw, Stefan Marks, Burkhard Claus Wunsche, Christof Lutteroth, Jude Buckley, and Paul Corballis. 2015. Development and Evaluation of an Exercycle Game Using Immersive Technologies. In *Australasian Workshop on Health Informatics and Knowledge Management (HIKM)*, 75–85.
241. Mike Sheinin and Carl Gutwin. 2014. Exertion in the small: improving differentiation and expressiveness in sports games with physical controls. 1845–1854. https://doi.org/10.1145/2556288.2557385
242. Jonathan Shepherd, Lewis Carter, Gert-Jan Pepping, and Leigh-Ellen Potter. 2018. Towards an Operational Framework for Designing Training Based Sports Virtual Reality Performance Simulators. *Proceedings* 2, 6: 214. https://doi.org/10.3390/proceedings2060214
243. Hiroaki Shigemasu, Toshiya Morita, Naoyuki Matsuzaki, Takao Sato, Masamitsu Harasawa, and Kiyoharu Aizawa. 2006. Effects of physical display size and amplitude of oscillation on visually induced motion sickness. In

Proceedings of the ACM symposium on Virtual reality software and technology, 372–375. https://doi.org/10.1145/1180495.1180571

244. Garth Shoemaker, Leah Findlater, Jessica Q. Dawson, and Kellogg S. Booth. 2009. Mid-air text input techniques for very large wall displays. In *Proceedings of Graphics interface 2009*, 231–238.
245. Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*, 281–288. https://doi.org/10.1145/332040.332445
246. Sanna E. Sihvonen, Sarianna Sipilä, and Pertti A. Era. 2004. Changes in Postural Balance in Frail Elderly Women during a 4-Week Visual Feedback Training: A Randomized Controlled Trial. *Gerontology* 50, 2: 87–95. https://doi.org/10.1159/000075559
247. Ricardo da Silva Alves, Denise Hollanda Iunes, Jovana Maria de Carvalho, Flávia da Silva Menezes, Andréia Maria Silva, Juliana Bassalobre Carvalho Borges, and Leonardo César Carvalho. 2018. Effects of Exergaming on Quality of Life in Cancer Patients. *Games for Health Journal* 7, 6: 385–392. https://doi.org/10.1089/g4h.2017.0174
248. Mara G. Silva and Doug A. Bowman. 2009. Body-Based Interaction for Desktop Games. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '09), 4249–4254. https://doi.org/10.1145/1520340.1520648
249. Adalberto Simeone, Eduardo Velloso, Jason Alexander, and Hans Gellersen. 2014. Feet movement in desktop 3D interaction. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, 71–74. https://doi.org/10.1109/3DUI.2014.6798845
250. R. William Soukoreff and I. Scott MacKenzie. 2003. Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 113–120.
251. Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–13. https://doi.org/10.1145/3173574.3174221
252. Srinath Sridhar, Anna Maria Feit, Christian Theobalt, and Antti Oulasvirta. 2015. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 3643–3652. https://doi.org/10.1145/2702123.2702136
253. Kay M. Stanney, Kelly S. Hale, Isabelina Nahmens, and Robert S. Kennedy. 2003. What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. *Human factors* 45, 3: 504–520. https://doi.org/10.1518/hfes.45.3.504.27254
254. Statista Research Department. 2020. Worldwide virtual and augmented reality headset shipments. *Global augmented and virtual reality headset shipment forecast 2019-2023*. Retrieved July 31, 2020 from https://www.statista.com/statistics/653390/worldwide-virtual-and-augmented-reality-headset-shipments/
255. Thomas Stoffregen, Elise Faugloire, Ken Yoshida, Moira Flanagan, and Omar Merhi. 2008. Motion Sickness and Postural Sway in Console Video Games. *Human factors* 50: 322–31. https://doi.org/10.1518/001872008X250755

256. Ivan E. Sutherland. 1968. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I* (AFIPS '68 (Fall, part I)), 757–764. https://doi.org/10.1145/1476589.1476686
257. Zsolt Szalavári and Michael Gervautz. 1997. The Personal Interaction Panel - a Two-Handed Interface for Augmented Reality. *Comput. Graph. Forum* 16: 335–346. https://doi.org/10.1111/1467-8659.16.3conferenceissue.35
258. Adrian Taylor, Nigel Cable, Guy Faulkner, Melvyn Hillsdon, Marco Narici, and Akke Van Der Bij. 2004. Physical activity and older adults: a review of health benefits and the effectiveness of interventions. *Journal of Sports Sciences* 22, 8: 703–725. https://doi.org/10.1080/02640410410001712421
259. Franco Tecchia, Giovanni Avveduto, Raffaello Brondi, Marcello Carrozzino, Massimo Bergamasco, and Leila Alem. 2014. I'm in VR!: using your own hands in a fully immersive MR system. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*, 73–76. https://doi.org/10.1145/2671015.2671123
260. Bruce Thomas and Wayne Piekarski. 2002. Glove Based User Interaction Techniques for Augmented Reality in an Outdoor Environment. *Virtual Reality* 6, 3: 167–180. https://doi.org/10.1007/s100550200017
261. Scott Thomas, Jeff Reading, and Roy Shephard. 1993. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). *Canadian journal of sport sciences* 17, 4: 338–45.
262. Cihan Topal, Burak Benligiray, and Cuneyt Akinlar. 2017. SliceType: Fast Gaze Typing with a Merging Keyboard. *arXiv preprint arXiv:1706.02499*.
263. Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. 1997. Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem." In *Human-Computer Interaction INTERACT '97*, Steve Howard, Judy Hammond and Gitte Lindgaard (eds.). Springer US, Boston, MA, 509–516. https://doi.org/10.1007/978-0-387-35175-9_77
264. Eduardo Velloso, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Interactions Under the Desk: A Characterisation of Foot Movements for Input in a Seated Position. In *Human-Computer Interaction – INTERACT 2015*, 384–401.
265. Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The Feet in Human–Computer Interaction: A Survey of Foot-Based Interaction. *ACM Computing surveys* 48, 2: 35. https://doi.org/10.1145/2816455
266. Paul Verhaeghen and Timothy Salthouse. 1997. Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological bulletin* 122, 3: 231–249. https://doi.org/10.1037//0033-2909.122.3.231
267. Ricardo Borges Viana, Claudia Lima Alves, Carlos Alexandre Vieira, Rodrigo Luiz Vancini, Mario Hebling Campos, Paulo Gentil, Marília Santos Andrade, and Claudio Andre Barbosa de Lira. 2017. Anxiolytic Effects of a Single Session of the Exergame Zumba® Fitness on Healthy Young Women. *Games for Health Journal* 6, 6: 365–370. https://doi.org/10.1089/g4h.2017.0085
268. Daniel Vogel and Patrick Baudisch. 2007. Shift: a technique for operating pen-based interfaces using touch. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, 657. https://doi.org/10.1145/1240624.1240727

269. Cong Wang, Zhong Liu, and Shing-Chow Chan. 2015. Superpixel-Based Hand Gesture Recognition With Kinect Depth Camera. *IEEE Transactions on Multimedia* 17, 1: 29–39. https://doi.org/10.1109/TMM.2014.2374357
270. Shiyao Wang, Michael Parsons, Jordan Stone-McLean, Peter Rogers, Sarah Boyd, Kristopher Hoover, Oscar Meruvia-Pastor, Minglun Gong, and Andrew Smith. 2017. Augmented reality as a telemedicine platform for remote procedural training. *Sensors* 17, 10: 2294. https://doi.org/10.3390/s17102294
271. Séamas Weech, Jae Moon, and Nikolaus F. Troje. 2018. Influence of bone-conducted vibration on simulator sickness in virtual reality. *PLOS ONE* 13, 3: e0194137. https://doi.org/10.1371/journal.pone.0194137
272. Gao-Xia Wei, Ting Xu, Feng-Mei Fan, Hao-Ming Dong, Li-Li Jiang, Hui-Jie Li, Zhi Yang, Jing Luo, and Xi-Nian Zuo. 2013. Can Taichi Reshape the Brain? A Brain Morphometry Study. *PLoS ONE* 8, 4: e61038. https://doi.org/10.1371/journal.pone.0061038
273. Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. 2013. Analysis of the Accuracy and Robustness of the Leap Motion Controller. *Sensors* 13, 5: 6380–6393. https://doi.org/10.3390/s130506380
274. David Matthew Whittinghill, Bradley Ziegler, T Case, and B Moore. 2015. Nasum virtualis: A simple technique for reducing simulator sickness. In *Proceedings of the Game Developers Conference (GDC)*. Retrieved from http://www.purdue.edu/newsroom/releases/2015/Q1/virtual-nosemay- reduce-simulator-sickness-in-video-games.html
275. Daniel Wigdor and Dennis Wixon. 2011. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann, Burlington, Mass.
276. Gareth Williams and Gema Fruhbeck. 2009. *Obesity: Science to Practice*. Wiley-Blackwell, West Sussex, UK. Retrieved May 5, 2019 from https://www.wiley.com/en-us/Obesity%3A+Science+to+Practice-p-9780470019115
277. Julius von Willich, Martin Schmitz, Florian Müller, Daniel Schmitt, and Max Mühlhäuser. 2020. Podoportation: Foot-Based Locomotion in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI ’20), 1–14. https://doi.org/10.1145/3313831.3376626
278. Andrew D. Wilson and Maneesh Agrawala. 2006. Text entry using a dual joystick game controller. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI ’06*, 475. https://doi.org/10.1145/1124772.1124844
279. Ian H. Witten, Eibe Frank, and Mark Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, Burlington, MA, USA. https://doi.org/10.1016/C2009-0-19715-5
280. Jacob O. Wobbrock, Duen Horng Chau, and Brad A. Myers. 2007. An alternative to push, press, and tap-tap-tap: gesturing on an isometric joystick for mobile phone text entry. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’07*, 667. https://doi.org/10.1145/1240624.1240728
281. Jacob O Wobbrock, James Rubinstein, Michael Sawyer, and Andrew T Duchowski. 2007. Not Typing but Writing: Eye-based Text Entry Using Letter-like Gestures. In *Proceedings of the Conference on Communications by Gaze Interaction (COGAIN)*, 61–64.
282. Jacob O. Wobbrock, James Rubinstein, Michael W. Sawyer, and Andrew T. Duchowski. 2008. Longitudinal evaluation of discrete consecutive gaze gestures

for text entry. In *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08*, 11. https://doi.org/10.1145/1344471.1344475

283. Steven L Wolf, Huiman X Barnhart, Gary L Ellison, Carol E Coogler, and Atlanta FICSIT Group. 1997. The Effect of Tai Chi Quan and Computerized Balance Training on Postural Stability in Older Subjects. *Physical Therapy* 77, 4: 371–381. https://doi.org/10.1093/ptj/77.4.371
284. Seline Wüest, Nunzio Alberto Borghese, Michele Pirovano, Renato Mainetti, Rolf van de Langenberg, and Eling D. de Bruin. 2014. Usability and Effects of an Exergame-Based Balance Training Program. *Games for Health Journal* 3, 2: 106–114. https://doi.org/10.1089/g4h.2013.0093
285. Wenge Xu, Hai-Ning Liang, Yifan Yu, Diego Monteiro, Khalad Hasan, and Charles Fleming. 2019. Assessing the Effects of a Full-body Motion-based Exergame in Virtual Reality. In *Proceedings of the Seventh International Symposium of Chinese CHI on - Chinese CHI 2019*. https://doi.org/10.1145/3332169.3333574
286. Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Difeng Yu, and Diego Monteiro. 2019. DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–14. https://doi.org/10.1145/3290605.3300674
287. Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Tianyu Zhang, Difeng Yu, and Diego Monteiro. 2019. RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. *IEEE transactions on visualization and computer graphics* 25, 5: 1991–2001. https://doi.org/10.1109/TVCG.2019.2898736
288. Tetsuya Yamamoto, Masahiko Tsukamoto, and Tomoki Yoshihisa. 2008. Foot-Step Input Method for Operating Information Devices While Jogging. In *2008 International Symposium on Applications and the Internet*, 173–176. https://doi.org/10.1109/SAINT.2008.97
289. Yukang Yan, Yingtian Shi, Chun Yu, and Yuanchun Shi. 2020. HeadCross: Exploring Head-Based Crossing Selection on Head-Mounted Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1: 22. https://doi.org/10.1145/3380983
290. Xin Yi, Chun Yu, Weijie Xu, Xiaojun Bi, and Yuanchun Shi. 2017. COMPASS: Rotational Keyboard on Non-Touch Smartwatches. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 705–715. https://doi.org/10.1145/3025453.3025454
291. Xin Yi, Chun Yu, Mingrui Zhang, Sida Gao, Ke Sun, and Yuanchun Shi. 2015. ATK: Enabling Ten-Finger Freehand Typing in Air Based on 3D Hand Tracking Data. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST '15*, 539–548. https://doi.org/10.1145/2807442.2807504
292. KangKang Yin and Dinesh K. Pai. 2003. FootSee: An Interactive Animation System. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (SCA '03), 329–338. https://doi.org/DOI:10.2312/SCA03/329-338
293. Birsen Yörük Açıkel, Uğur Turhan, and Yavuz Akbulut. 2017. Effect of Multitasking on Simulator Sickness and Performance in 3D Aerodrome Control Training. *Simulation & Gaming* 49, 1: 27–49. https://doi.org/10.1177/1046878117750417

294. Chun Yu, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. Tap, Dwell or Gesture?: Exploring Head-Based Text Entry Techniques for HMDs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4479–4488. https://doi.org/10.1145/3025453.3025964
295. Difeng Yu, Kaixuan Fan, Heng Zhang, Diego Monteiro, Wenge Xu, and Hai-Ning Liang. 2018. PizzaText: Text Entry for Virtual Reality Systems Using Dual Thumbsticks. *IEEE Transactions on Visualization and Computer Graphics* 24, 11: 2927–2935. https://doi.org/10.1109/TVCG.2018.2868581
296. Zhengyou Zhang. 2012. Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia* 19, 2: 4–10. https://doi.org/10.1109/MMUL.2012.24
297. Shengdong Zhao and Ravin Balakrishnan. 2004. Simple vs. compound mark hierarchical marking menus. In *Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04*, 33. https://doi.org/10.1145/1029632.1029639
298. The most frequent 10 000 words of English and Example Sentences - The list of most commonly used words. Retrieved April 1, 2018 from http://www.use-in-a-sentence.com/english-words/10000-words/the-most-frequent-10000-words-of-english.html
299. Design – Meta User Interface Guidelines | Optimize Your Workspace. *Meta*. Retrieved August 24, 2018 from https://devcenter.metavision.com/design/user-interface-guidelines-optimize-your-workspace
300. Magic Leap One: Creator Edition | Magic Leap. Retrieved September 14, 2018 from https://www.magicleap.com/magic-leap-one
301. VR Fitness Insider - Your Source for Virtual Reality Health and Fitness News. *VR Fitness Insider*. Retrieved January 7, 2019 from https://www.vrfitnessinsider.com/

Appendix 1 – NASA-TLX Questionnaire

1. Mental: How mentally demanding was the task (thinking, remembering, looking, searching, etc). Rate from 1 (low workload) to 21 (high workload).
2. Physical: How physically demanding was the task (e.g. turning, controlling, activating, etc.)? Rate from 1 (low workload) to 21 (high workload).
3. Temporal: How hurried or rushed was the pace of the task? Rate from 1 (low workload) to 21 (high workload).
4. Effort: How hard did you have to work to accomplish your level of performance? Rate from 1 (low workload) to 21 (high workload).
5. Frustration: How insecure, discourage, irritated, stressed, and annoyed were you? Rate from 1 (low workload) to 21 (high workload).
6. Performance: How successful were you in accomplishing what you were asked to do? Rate from 1 (superb) to 21 (Failure).
7. From these 2 which caused the greatest workload
   a. Mental
   b. Physical
8. From these 2 which caused the greatest workload
   a. Mental
   b. Temporal
9. From these 2 which caused the greatest workload
   a. Mental
   b. Performance
10. From these 2 which caused the greatest workload
    a. Mental
    b. Effort
11. From these 2 which caused the greatest workload
    a. Mental
    b. Frustration
12. From these 2 which caused the greatest workload
    a. Physical
    b. Temporal

13. From these 2 which caused the greatest workload
    a. Physical
    b. Performance
14. From these 2 which caused the greatest workload
    a. Physical
    b. Effort
15. From these 2 which caused the greatest workload
    a. Physical
    b. Frustration
16. From these 2 which caused the greatest workload
    a. Temporal
    b. Performance
17. From these 2 which caused the greatest workload
    a. Temporal
    b. Effort
18. From these 2 which caused the greatest workload
    a. Temporal
    b. Frustration
19. From these 2 which caused the greatest workload
    a. Performance
    b. Effort
20. From these 2 which caused the greatest workload
    a. Performance
    b. Frustration
21. From these 2 which caused the greatest workload
    a. Effort
    b. Frustration

Appendix 2 – Slater-Usoh-Steed Questionnaire

Please indicate how you felt during the condition for each of the items. Rate from 1 (not at all) to 7 (very much).

1. Please rate your sense of being in the virtual environment.
2. To what extent were there times during the experience when the virtual environment was the reality for you?
3. Do you think of the virtual environment more as images that you saw or more as somewhere that you visited?
4. Which was the strongest on the whole, your sense of being in the virtual environment or of being elsewhere?
5. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By 'structure of the memory' consider things like the extent to which you have a visual memory of the virtual environment, whether that memory is in color, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such structural elements.
6. Did you often think to yourself that you were actually in the virtual environment?

Appendix 3 – User Experience Questionnaire

Please indicate how you felt during the condition for each of the items.

1. Annoying/Enjoyable. Rate from 1 (annoying) to 7 (enjoyable).
2. Not understandable/Understandable. Rate from 1 (not understandable) to 7 (understandable).
3. Dull/Creative. Rate from 1 (dull) to 7 (creative).
4. Difficult to learn/Easy to learn. Rate from 1 (difficult to learn) to 7 (easy to learn).
5. Inferior/Valuable. Rate from 1 (inferior) to 7 (valuable).
6. Boring/Exciting. Rate from 1 (boring) to 7 (exciting).
7. Not interesting/Interesting. Rate from 1 (not interesting) to 7 (interesting).
8. unpredictable/Predictable. Rate from 1 (unpredictable) to 7 (predictable).
9. Slow/Fast. Rate from 1 (slow) to 7 (fast).
10. Conventional/Inventive. Rate from 1 (conventional) to 7 (inventive).
11. Obstructive/Supportive. Rate from 1 (obstructive) to 7 (supportive).
12. Bad/Good. Rate from 1 (bad) to 7 (good).
13. Complicated/Easy. Rate from 1 (complicated) to 7 (easy).
14. Unlikable/Pleasing. Rate from 1 (unlikable) to 7 (pleasing).
15. Usual/Leading edge. Rate from 1 (usual) to 7 (leading edge).
16. Unpleasant/Pleasant. Rate from 1 (unpleasant) to 7 (pleasant).
17. Not secure/Secure. Rate from 1 (not secure) to 7 (secure).
18. Motivating/Demotivating. Rate from 1 (motivating) to 7 (demotivating).
19. Does not meet expectations/Meet expectations. Rate from 1 (does not meet expectations) to 7 (meet expectations).
20. Inefficient/Efficient. Rate from 1 (inefficient) to 7 (efficient).
21. confusing/Clear. Rate from 1 (confusing) to 7 (clear).
22. Impractical/Practical. Rate from 1 (impractical) to 7 (practical).
23. Cluttered/Organized. Rate from 1 (cluttered) to 7 (organized).
24. Unattractive/Attractive. Rate from 1 (unattractive) to 7 (attractive).
25. Unfriendly/Friendly. Rate from 1 (unfriendly) to 7 (friendly).
26. Conservative/Innovative. Rate from 1 (conservative) to 7 (innovative).

Appendix 4 – Motion Sickness Assessment Questionnaire

Please indicate how you felt during the condition for each of the items. Rate from 1 (not at all) to 9 (severely).

1. I felt sick to my stomach.
2. I felt faint-like.
3. I felt annoyed/irritated.
4. I felt sweaty.
5. I felt queasy.
6. I felt lightheaded.
7. I felt drowsy.
8. I felt clammy/cold sweat.
9. I felt disoriented.
10. I felt tired/fatigued.
11. I felt nauseated.
12. I felt hot/warm.
13. I felt dizzy.
14. I felt like I was spinning.
15. I felt as if I may vomit.
16. I felt uneasy.

Appendix 5 – System Usability Scale

Please indicate how you felt during the condition for each of the items. Rate from 1 (strongly disagree) to 5 (strongly agree).

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Appendix 6 – Borg RPE CR10

1. Please select one option from the list that can describe your tiredness during the experiment:
    a. 0 – Nothing at all
    b. 0.5 – Very, very slight (just noticeable)
    c. 1 – Very slight
    d. 2 – Slight
    e. 3 – Moderate
    f. 4 – Somewhat severe
    g. 5 – Severe
    h. 6
    i. 7 – Very severe
    j. 8
    k. 9 – Very, very severe (almost maximal)
    l. 10 – Maximal

Appendix 7 – Computer Vision Syndrome Questionnaire

(0-3)

Frequency, intensity

1. Frequency of burning occurrence
   a. Never
   b. Occasionally
   c. Often
   d. Very often
2. Intensity of burning
   a. Moderate
   b. Intense
   c. Very intense
3. Frequency of itching occurrence
   a. Never
   b. Occasionally
   c. Often
   d. Very often
4. Intensity of itching
   a. Moderate
   b. Intense
   c. Very intense
5. Frequency of feeling of a foreign body occurrence
   a. Never
   b. Occasionally
   c. Often
   d. Very often
6. Intensity of feeling of a foreign body
   a. Moderate
   b. Intense
   c. Very intense
7. Frequency of tearing occurrence
   a. Never
   b. Occasionally

   c. Often
   d. Very often
8. Intensity of tearing
   a. Moderate
   b. Intense
   c. Very intense
9. Frequency of excessive blinking occurrence
   a. Never
   b. Occasionally
   c. Often
   d. Very often
10. Intensity of excessive blinking
   a. Moderate
   b. Intense
   c. Very intense
11. Frequency of eye redness occurrence
   a. Never
   b. Occasionally
   c. Often
   d. Very often
12. Intensity of eye redness
   a. Moderate
   b. Intense
   c. Very intense
13. Frequency of eye pain occurrence
   a. Never
   b. Occasionally
   c. Often
   d. Very often
14. Intensity of eye pain
   a. Moderate
   b. Intense
   c. Very intense
15. Frequency of heavy eyelids occurrence

  a. Never
  b. Occasionally
  c. Often
  d. Very often

16. Intensity of heavy eyelids
    a. Moderate
    b. Intense
    c. Very intense
17. Frequency of dryness occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
18. Intensity of dryness
    a. Moderate
    b. Intense
    c. Very intense
19. Frequency of blurred vision occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
20. Intensity of blurred vision
    a. Moderate
    b. Intense
    c. Very intense
21. Frequency of double vision occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
22. Intensity of double vision
    a. Moderate
    b. Intense

   c. Very intense

23. Frequency of difficulty focusing for near vision occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
24. Intensity of difficulty focusing for near vision
    a. Moderate
    b. Intense
    c. Very intense
25. Frequency of increased sensitivity to light occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
26. Intensity of increased sensitivity to light
    a. Moderate
    b. Intense
    c. Very intense
27. Frequency of colored halos around objects occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
28. Intensity of colored halos around objects
    a. Moderate
    b. Intense
    c. Very intense
29. Frequency of feeling that sight is worsening occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
30. Intensity of feeling that sight is worsening

a. Moderate
b. Intense
c. Very intense

31. Frequency of headache occurrence
    a. Never
    b. Occasionally
    c. Often
    d. Very often
32. Intensity of headache
    a. Moderate
    b. Intense
    c. Very intense

Appendix 8 – Simulator Sickness Questionnaire

Please indicate how you felt during the condition for each of the items. Rate from 0 (none) to 3 (severe).

1. General discomfort
2. Fatigue
3. Headache
4. Eye strain
5. Difficulty focusing
6. Salivation increasing
7. Sweating
8. Nausea
9. Difficulty concentrating
10. Fullness of the Head
11. Blurred vision
12. Dizziness with eyes open
13. Dizziness with eyes closed
14. Vertigo (vertigo is experienced as loss of orientation with respect to vertical upright)
15. Stomach awareness (stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea)
16. Burping

Appendix 9 – Game Experience Questionnaire

Please indicate how you felt while playing the game for each of the items. Rate from 1 (not at all) to 5 (extremely).

1. I felt content
2. I felt skilful
3. I was interested in the game's story
4. I thought it was fun
5. I was fully occupied with the game
6. I felt happy
7. It gave me a bad mood
8. I thought about other things
9. I found it tiresome
10. I felt competent
11. I thought it was hard
12. It was aesthetically pleasing
13. I forgot everything around me
14. I felt good
15. I was good at it
16. I felt bored
17. I felt successful
18. I felt imaginative
19. I felt that I could explore things
20. I enjoyed it
21. I was fast at reaching the game's targets
22. I felt annoyed
23. I felt pressured
24. I felt irritable
25. I lost track of time
26. I felt challenged
27. I found it impressive
28. I was deeply concentrated in the game
29. I felt frustrated
30. It felt like a rich experience

31. I lost connection with the outside world
32. I felt time pressure
33. I had to put a lot of effort into it

Appendix 10 – Borg RPE CR6-20

1. Please select one option from the list that can describe your tiredness during the experiment:
    a. 6 – Non Exertion (Little to no movement, very relaxed
    b. 7 – Extremely Light (Able to maintain pace)
    c. 8
    d. 9 – Very Light (Comfortable and breathing harder)
    e. 10
    f. 11 – Light (Minimal sweating, can talk easily)
    g. 12
    h. 13 – Somewhat Hard (Slight breathlessness, can talk)
    i. 14 – (Increased sweating, still able to hold conversation but with difficulty)
    j. 15 – Hard (Sweating, able to push and still maintain proper form)
    k. 16
    l. 17 – Very Hard (Can keep a fast pace for a short time period)
    m. 18
    n. 19 – Extremely Hard (Difficulty breathing, near muscle exhaustion)
    o. 20 – Maximally Hard (STOP exercising, total exhaustion)

Appendix 11 – Intrinsic Motivation Inventory

For each of the following statements, please indicate how true it is for you, rating from 1 (not at all) to 7 (very true):

1. I enjoyed doing this activity very much
2. This activity was fun to do
3. I thought this was a boring activity
4. This activity did not hold my attention at all
5. I would describe this activity as very interesting
6. I thought this activity was quite enjoyable
7. While I was doing this activity, I was thinking about how much I enjoyed it
8. I think I am pretty good at this activity
9. I think I did pretty well at this activity, compared to other students
10. After working at this activity for awhile, I felt pretty competent
11. I am satisfied with my performance at this task
12. I was pretty skilled at this activity
13. This was an activity that I couldn’t do very well
14. I did not feel nervous at all while doing this
15. I felt very tense while doing this activity
16. I was very relaxed in doing these
17. I was anxious while working on this task
18. I felt pressured while doing these
19. I believe this activity could be of some value to me
20. I think that doing this activity is useful for health
21. I think this is important to do because it can improve my health
22. I would be willing to do this again because it has some value to me
23. I think doing this activity could help me to build up my health
24. I believe doing this activity could be beneficial to me
25. I think this is an important activity

Appendix 12 – Physical Activity Readiness Questionnaire

1. Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?
2. Do you feel pain in your chest when you do physical activity?
3. In the past month, have you had chest pain when you were not doing physical activity?
4. Do you lose your balance because of dizziness or do you ever lose consciousness?
5. Do you have a bone or joint problem that could be made worse by a change in your physical activity?
6. Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?
7. Do you know of any other reason why you should not do physical activity?

Appendix 13 – Questionnaire Used in Chapter 3

Pre-experiment questionnaire:

1. Participant Number (given by researcher)
2. Your age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. How familiar are you with the QWERTY keyboard? Rating from 1 (no skill) to 5 (expert).
5. How good are you in remembering short English sentences? Rating from 1 (no skill) to 5 (expert).
6. How often do you type long texts? Rating from 1 (never) to 5 (always)
7. Have you ever experienced AR device (if yes in question 7)?
    a. Yes
    b. No
8. How often you use AR device?
    a. Daily
    b. Weekly
    c. Monthly
    d. Yearly

Post-condition questionnaire:

1. NASA-TLX questionnaire
2. Motion sickness assessment questionnaire
3. Slater usoh steed questionnaire
4. User experience questionnaire
5. Any comments for the technique you just tried? Feel free to write anything.

Post-experiment questionnaire:

1. Feel free to write anything.

Appendix 14 – Questionnaire Used in Chapter 5

Pre-experiment questionnaire:

1. Participant Number
2. Age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. Have you ever experienced AR device?
    a. Yes
    b. No
5. How often you use AR device (if yes in question 4)?
    a. Daily.
    b. Weekly
    c. Monthly
    d. Yearly
6. Have you ever experienced Magic Leap (if yes in question 4)?
    a. Yes
    b. No
7. Strong hand.
    a. Left-Handed
    b. Right-Handed

Post-condition questionnaire:

1. System usability scale
2. NASA-TLX questionnaire
3. User experience questionnaire
4. Borg CR10
5. Computer vision syndrome questionnaire
6. Any comments for the technique you just tried? Feel free to write anything.

Post-experiment questionnaire:

1. Please rank the 5 techniques (Static surface, dynamic surface, static coordinate line, dynamic coordinate line, benchmark). 1 for the most preferred option and 5 for the least preferred option.
2. Feel free to write anything.

Appendix 15 – Questionnaires Used in Chapter 6

Study 1

Pre-experiment questionnaire:

1. Participant Number (given by researcher)
2. Your age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. How familiar are you with the QWERTY keyboard? Rating from 1 (no skill) to 5 (expert).
5. How good are you in remembering short English sentences? Rating from 1 (no skill) to 5 (expert).
6. How often do you type long texts? Rating from 1 (never) to 5 (always)
7. Have you ever experienced VR device?
    a. Yes
    b. No
8. How often you use VR device (if yes in question 7)?
    a. Daily
    b. Weekly
    c. Monthly
    d. Yearly

Post-condition questionnaire:

1. Rate your experience of the technique you just experienced. From 1 (novice) to 5 (expert)
2. NASA-TLX questionnaire
3. Simulator sickness questionnaire

Post-experiment questionnaire:

1. Please rank the 3 alphabet starting position (top, left, right). 1 for the most preferred option and 3 for the least preferred option.

2. Please rank the 4 alphabet starting position (1 letter per region with small inner circle, 1 letter per region with large inner circle, 2 letters per region with small inner circle, 2 letters per region with large inner circle). 1 for the most preferred option and 4 for the least preferred option.
3. Feel free to write anything.

Study 2

Pre-experiment questionnaire:

1. Participant Number (given by researcher)
2. Your age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. How familiar are you with the QWERTY keyboard? Rating from 1 (no skill) to 5 (expert).
5. How good are you in remembering short English sentences? Rating from 1 (no skill) to 5 (expert).
6. How often do you type long texts? Rating from 1 (never) to 5 (always)
7. Have you ever experienced VR device?
    a. Yes
    b. No
8. How often you use VR device (if yes in question 7)?
    a. Daily
    b. Weekly
    c. Monthly
    d. Yearly

Post-condition questionnaire:

1. Any comments for the technique you just tried? Feel free to write anything.

Post-experiment questionnaire:

1. Feel free to write anything.

Study 3

Pre-experiment questionnaire:

1. Participant Number (given by researcher)
2. Your age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. How familiar are you with the QWERTY keyboard? Rating from 1 (no skill) to 5 (expert).
5. How good are you in remembering short English sentences? Rating from 1 (no skill) to 5 (expert).
6. How often do you type long texts? Rating from 1 (never) to 5 (always)
7. Have you ever experienced VR device?
    a. Yes
    b. No
8. How often you use VR device (if yes in question 7)?
    a. Daily
    b. Weekly
    c. Monthly
    d. Yearly

Appendix 16 – Questionnaire Used in Chapter 7

Study 1

Pre-experiment questionnaire:

1. Participant Number
2. Age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. Have you ever experienced AR device?
    a. Yes
    b. No
5. How often you use AR device (if yes in question 4)?
    a. Daily.
    b. Weekly
    c. Monthly
    d. Yearly
6. Strong hand.
    a. Left-Handed
    b. Right-Handed
7. Rate your balance skill in real-life. Rate from 1 (very bad) to 7 (strong)

Post-condition questionnaire:

Physical comfort for 8-direction model

1. Comfort score on North direction. Rate from 1 (extremely easy) to 5 (extremely hard).
2. Comfort score on North-East direction. Rate from 1 (extremely easy) to 5 (extremely hard).
3. Comfort score on East direction. Rate from 1 (extremely easy) to 5 (extremely hard).
4. Comfort score on South-East direction. Rate from 1 (extremely easy) to 5 (extremely hard).

5. Comfort score on South direction. Rate from 1 (extremely easy) to 5 (extremely hard).
6. Comfort score on South-West direction. Rate from 1 (extremely easy) to 5 (extremely hard).
7. Comfort score on West direction. Rate from 1 (extremely easy) to 5 (extremely hard).
8. Comfort score on North-West direction. Rate from 1 (extremely easy) to 5 (extremely hard).

Mental comfort for 8-direction model

1. Comfort score on North direction. Rate from 1 (extremely easy) to 5 (extremely hard).
2. Comfort score on North-East direction. Rate from 1 (extremely easy) to 5 (extremely hard).
3. Comfort score on East direction. Rate from 1 (extremely easy) to 5 (extremely hard).
4. Comfort score on South-East direction. Rate from 1 (extremely easy) to 5 (extremely hard).
5. Comfort score on South direction. Rate from 1 (extremely easy) to 5 (extremely hard).
6. Comfort score on South-West direction. Rate from 1 (extremely easy) to 5 (extremely hard).
7. Comfort score on West direction. Rate from 1 (extremely easy) to 5 (extremely hard).
8. Comfort score on North-West direction. Rate from 1 (extremely easy) to 5 (extremely hard).

Physical comfort 16-directional model

1. Comfort score on North direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
2. Comfort score on North direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
3. Comfort score on North-East direction close. Rate from 1 (extremely easy) to 5 (extremely hard).

4. Comfort score on North-East direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
5. Comfort score on East direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
6. Comfort score on East direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
7. Comfort score on South-East direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
8. Comfort score on South-East direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
9. Comfort score on South direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
10. Comfort score on South direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
11. Comfort score on South-West direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
12. Comfort score on South-West direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
13. Comfort score on West direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
14. Comfort score on West direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
15. Comfort score on North-West direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
16. Comfort score on North-West direction far. Rate from 1 (extremely easy) to 5 (extremely hard).

Mental comfort 16-directional model

1. Comfort score on North direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
2. Comfort score on North direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
3. Comfort score on North-East direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
4. Comfort score on North-East direction far. Rate from 1 (extremely easy) to 5 (extremely hard).

5. Comfort score on East direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
6. Comfort score on East direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
7. Comfort score on South-East direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
8. Comfort score on South-East direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
9. Comfort score on South direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
10. Comfort score on South direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
11. Comfort score on South-West direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
12. Comfort score on South-West direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
13. Comfort score on West direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
14. Comfort score on West direction far. Rate from 1 (extremely easy) to 5 (extremely hard).
15. Comfort score on North-West direction close. Rate from 1 (extremely easy) to 5 (extremely hard).
16. Comfort score on North-West direction far. Rate from 1 (extremely easy) to 5 (extremely hard).

Post-experiment questionnaire:

Social Acceptance

1. On a scale from 1 (I hated it, it felt terribly awkward) to 6 (I enjoyed it, it felt comfortable), what was your overall impression/emotion during the task.
2. Imagine that this motion direction gestures can be used to control a menu or to play dance game. Now, **in front of whom** do you think you would **feel comfortable** using such gestures? Select **one or more** items from the list below.
    a. I would not feel comfortable using them even when alone
    b. when alone
    c. in front of my partner

   d. in front of friends
   e. in front of family
   f. in front of colleagues
   g. in front of strangers

3. Now, **in which locations** do you think you would **feel comfortable** using such gestures? select **one or more** items from the list below.
   a. I would not feel comfortable using them no matter where I am
   b. at home
   c. on the sidewalk
   d. in a pub, cafe, or restaurant
   e. in a shop
   f. in a museum
   g. as a passenger on a bus or train
   h. at my workplace

Study 2

Pre-experiment questionnaire:

1. Participant Number
2. Age
3. Gender
   a. Female
   b. Male
   c. Prefer not to say
4. Have you ever experienced AR device?
   a. Yes
   b. No
5. How often you use AR device (if yes in question 4)?
   a. Daily.
   b. Weekly
   c. Monthly
   d. Yearly

Post-condition questionnaire:

1. NASA-TLX questionnaire
2. User experience questionnaire
3. Motion sickness assessment questionnaire
4. Any comments for the technique you just tried? Feel free to write anything.

Post-experiment questionnaire:

1. Feel free to write anything.

Appendix 17 – Questionnaire Used in Chapter 8

Pre-experiment questionnaire:

1. Participant Number (given by researcher)
2. Your age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. Have you ever experienced VR device?
    a. Yes
    b. No
5. How often you use VR device (if yes in question 4)?
    a. Daily
    b. Weekly
    c. Monthly
    d. Yearly

Post-condition questionnaire:

1. Simulator sickness questionnaire
2. Game experience questionnaire
3. Any comments for the technique you just tried? Feel free to write anything.

Post-experiment questionnaire:

1. Overall, what did you think about the game?
2. What did you like about the game?
3. What did not you like about the game?
4. Was there anything more difficult than you expected in the game?
5. Was there anything more confusing than you expected in the game?

Appendix 18 – Questionnaire Used in Chapter 9

Study 1

Pre-experiment questionnaire:

1. Participant Number (given by researcher)
2. Your age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. Have you ever experienced VR device (if the user belongs to VR group)?
    a. Yes
    b. No
5. How often you use VR device (if yes in question 4)?
    a. Daily
    b. Weekly
    c. Monthly
    d. Yearly

Post-experiment questionnaire:

1. NASA-TLX questionnaire
2. I like performing *Psi*. Rating from 1 (strongly disagree) to 7 (strongly agree).
3. I like performing *Squat*. Rating from 1 (strongly disagree) to 7 (strongly agree).
4. I like performing *Kick*. Rating from 1 (strongly disagree) to 7 (strongly agree).
5. I like performing *Walk*. Rating from 1 (strongly disagree) to 7 (strongly agree).
6. I like performing *Wheel*. Rating from 1 (strongly disagree) to 7 (strongly agree).
7. I like performing *Zoom*. Rating from 1 (strongly disagree) to 7 (strongly agree).
8. I like performing *Squat+Psi*. Rating from 1 (strongly disagree) to 7 (strongly agree).

9. I like performing *Squat+Wheel*. Rating from 1 (strongly disagree) to 7 (strongly agree).
10. I like performing *Kick+Zoom*. Rating from 1 (strongly disagree) to 7 (strongly agree).
11. I like performing *Kick+Wheel*. Rating from 1 (strongly disagree) to 7 (strongly agree).
12. I like performing *Walk+Psi*. Rating from 1 (strongly disagree) to 7 (strongly agree).
13. I like performing *Walk+Zoom*. Rating from 1 (strongly disagree) to 7 (strongly agree).

Study 2

Pre-experiment questionnaire:

1. Participant Number (given by researcher)
2. Your age
3. Gender
    a. Female
    b. Male
    c. Prefer not to say
4. Have you ever experienced VR device?
    a. Yes
    b. No
5. How often you use VR device (if yes in question 4)?
    a. Daily
    b. Weekly
    c. Monthly
    d. Yearly
6. Have you played videogames before?
    a. Yes
    b. No
7. How often you play videogames (if yes in question 6)?
    a. Daily
    b. Weekly

c. Monthly
d. Yearly

Post-condition questionnaire:

1. Fill in average heart rate (by experimenter)
2. Fill in max heart rate (by experimenter)
3. Fill in calories burned (by experimenter)
4. Simulator sickness questionnaire
5. Game experience questionnaire
6. Borg RPE CR6-20
7. Any comments for the technique you just tried? Feel free to write anything.

Post-experiment questionnaire:

1. Please rank the 4 versions (VR-1PP, VR-3PP, LD-1PP, LD-3PP). 1 for the most preferred option and 4 for the least preferred option.
2. Overall, what did you think about the game?
3. What did you like about the game?
4. What did not you like about the game?
5. Was there anything more difficult than you expected in the game?
6. Was there anything more confusing than you expected in the game?