Universal Usability of Virtual Reality

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of Doctor of Philosophy in
Computer Science and Engineering

by
Majed Al Zayer
Eelke Folmer, Ph.D., Dissertation Advisor
May 2019
We recommend that the dissertation prepared under our supervision by

MAJED AL ZAYER

entitled

Universal Usability of Virtual Reality

be accepted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Eelke Folmer, Ph.D. – Advisor

Sergiu Dascalu, Ph.D. – Committee Member

David Feil-Seifer, Ph.D. – Committee Member

Kostas Alexis, Ph.D. – Committee Member

Paul MacNeilage, Ph.D. – Graduate School Representative

David W. Zeh, Ph.D. – Dean, Graduate School

May 2019
ABSTRACT

Virtual Reality (VR) is envisioned to be a mainstream medium that would change how we work, connect, and play. For this vision to be realized, however, VR must be equally accessible and enjoyable for all users irrespective of their differences. This dissertation addresses two research themes that are concerned with the universal usability of VR.

In the first theme, the gap in user experience quality between high-end VR platforms (PC VR) and low-end VR platforms (mobile VR) is addressed through the development of two low-cost acoustic interaction techniques for mobile VR. The first technique, PAWdio, appropriates a regular earbud and the smartphone’s microphone to enable hand input for mobile VR. PAWdio was found to increase the immersiveness of the user experience compared to the traditional input of mobile VR. StereoTrack is another acoustic tracking technique that enables real walking in mobile VR using a pair of regular speakers and the smartphone’s microphone. The interaction fidelity of StereoTrack was evaluated in terms of accuracy, precision, and latency. The potential of StereoTrack to augment existing locomotion techniques for mobile VR was also demonstrated in the context of two VR games.

In the second theme, the effect of dynamic field-of-view (FOV) restriction as a technique to mitigate VR sickness on the sex gap in VR user experience was investigated through two studies that considered sex and FOV restriction as the variables under analysis. In the first study, the effect of dynamic FOV restriction on participants’ route knowledge was examined using a triangle completion task. In the second study, participants’ spatial learning as a response to changing the FOV was evaluated using a virtual version of the Morris Water Maze. Both studies aim to expand our understanding of the relationship between sex differences in spatial abilities, VR sickness, and FOV restriction in VR.
For my parents, Awatif and Rasheed, from whom I learned the value of being different. For Mona, my life partner and other half, who always believed in me and supported me unconditionally. And for my daughters, Noor and Maryam, who missed countless bedtime stories while I was in the lab.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Folmer, for empowering me all the way through my journey. I also would like to thank my dissertation committee for their time and support. I am grateful to all of my co-authors and peers who, directly or indirectly, helped me to become the researcher I am today. Special thanks to Sam Tregillus with whom I administered my first user study. I am also grateful to my fellow researchers who established the foundation for my research. I finally would like to thank Reviewer #2 whose feedback contributed to improving the quality of my submissions on the expense of high blood pressure. This work was partially supported by Google and Mozilla.
# TABLE OF CONTENTS

Abstract ......................................................... i
Dedication ....................................................... ii
Acknowledgements .............................................. iii
Table of Contents ............................................... iv
List of Tables .................................................... vii
List of Figures ................................................... viii

1 Introduction .............................. 1
  1.1 Publications ....................................... 7

2 A Survey of Virtual Locomotion Techniques .......... 9
  2.1 Introduction ....................................... 9
  2.2 VR Locomotion Techniques Overview ................ 10
    2.2.1 Walking-based .................................. 11
    2.2.2 Steering ......................................... 35
    2.2.3 Selection-Based .................................. 41
    2.2.4 Manipulation-Based ............................... 43
    2.2.5 Multiscale Virtual Locomotion .................. 44
  2.3 VR Locomotion Taxonomies ............................ 46
    2.3.1 General Taxonomies ............................... 46
    2.3.2 Specific Taxonomies ............................... 48
  2.4 Discussion ........................................... 48

3 PAWdio: Hand Input for Mobile VR using Acoustic Sensing ........ 57
  3.1 Introduction ....................................... 57
  3.2 Background .......................................... 59
  3.3 PAWdio Design ....................................... 60
  3.4 Evaluation .......................................... 63
    3.4.1 Instrumentation .................................. 63
    3.4.2 Participants ...................................... 64
    3.4.3 Procedure ........................................ 64
    3.4.4 Results .......................................... 66
  3.5 Discussion ........................................... 67
  3.6 Limitations and Future work ......................... 68

4 StereoTrack: 180-degree Low-cost Acoustic Positional Tracking for Mobile VR Platforms .......... 71
  4.1 Introduction ....................................... 71
  4.2 Background .......................................... 73
  4.3 Design of StereoTrack ............................... 74
  4.4 Evaluation .......................................... 77
    4.4.1 Instrumentation .................................. 78
4.4.2 Precision Analysis ........................................ 78
4.4.3 Accuracy Analysis ...................................... 79
4.4.4 Latency Analysis ........................................ 81
4.5 Example Applications ........................................ 82
4.5.1 Wave Shooter Game ...................................... 82
4.5.2 Endless Running Game .................................... 83
4.6 Discussion and Limitations .................................... 85

5 The Effect of Field-of-View Restriction on Sex Bias in VR Sickness and
Spatial Navigation Performance ................................. 87
5.1 Introduction .................................................. 87
5.2 Background .................................................. 90
5.2.1 VR Sickness restriction Mechanisms .................. 91
5.2.2 FOV Manipulation in Virtual Environments .......... 92
5.3 User Study .................................................. 92
5.3.1 Participants .............................................. 93
5.3.2 Experiment Design ....................................... 94
5.3.3 FOV Test Conditions .................................... 95
5.3.4 Instrumentation .......................................... 96
5.3.5 Virtual Environment ..................................... 96
5.3.6 Task .................................................... 98
5.3.7 Procedure ............................................... 100
5.3.8 Measurements .......................................... 102
5.4 Results ...................................................... 103
5.4.1 Spatial Navigation Performance ......................... 103
5.4.2 VR Sickness .............................................. 104
5.5 Discussion and Future Work ................................. 107

6 Effects of VR Field-of-View Restriction on Spatial Learning and VR Sickness
Among Men and Women Using a Virtual Morris Water Maze .......... 113
6.1 Introduction ................................................. 113
6.2 Background ................................................. 114
6.3 User Study .................................................. 117
6.3.1 Participants .............................................. 117
6.3.2 Tasks .................................................... 117
6.3.3 Experiment Design ....................................... 120
6.3.4 FOV Test Conditions .................................... 120
6.3.5 Instrumentation .......................................... 121
6.3.6 Virtual Environment ..................................... 122
6.3.7 Procedure ............................................... 123
6.3.8 Measurements .......................................... 124
6.4 Results ...................................................... 126
6.4.1 Virtual Water Maze Task ................................ 127
6.4.2 Object Placement Task ................................... 127
6.4.3 Simulator Sickness Questionnaire .......................... 129
6.5 Discussion and Future Work ................................. 129

7 Conclusion 133

Bibliography 135
LIST OF TABLES

2.1 Summary of redirection techniques implementations. Some implementations present more than one redirection technique. In those cases we use different symbols ■, ▲, or ♦ to distinguish different implementations and what methods they use. ........................................... 17

2.2 A summary of WIP implementations described in terms of the foot motion pattern, speed control, direction control, sensing requirements, and movements supported. Speed and direction controls are specified in terms of the body parts involved and the control mechanism used. ................................................................. 31

3.1 Average time and errors for all users. ............................................................. 66

3.2 Likert scores for each usability attribute per technique (SD). ................. 66

5.1 Summary of participants ratings of their frequency of using VR and their tendency of getting motion or VR sick on a scale of 1 (never) to 5 (very frequently). The results are reported in the form of percentage (count). ................................................................. 93

5.2 Quantitative measures of the Home-Position-Estimation error (HPE), discomfort scores, and relative Simulator Sickness Questionnaire in terms of mean (standard deviation). ................................. 103

6.1 Quantitative measures of virtual Morris Water Maze, Object placement, and simulator sickness questionnaire in terms of mean (standard deviation). ................................................................. 125
LIST OF FIGURES

3.1 PAWdio appropriates a pair of ordinary in-ear headphones to track the position of the hand. The user holds a single earbud in their hand that produces an inaudible tone. Doppler shifts are used to determine the velocity of the earbud from or towards the phone, which is then used to manipulate the Z-position of a virtual hand that is attached to the user’s gaze pointer. ........................................ 58
3.2 Ranking of input methods based on usability criteria. ......................... 66
3.3 Evaluation of immersion for both input techniques ................................. 68

4.1 StereoTrack appropriates an ordinary pair of speakers, for example, those available in a TV, to enable 180 degree positional tracking on mobile VR platforms using acoustic sensing. StereoTrack can be used to implement popular waveshooter games on mobile VR platforms. ................................................................. 72
4.2 Results of the accuracy evaluation of StereoTrack in terms of drift over time, measured as Root-Mean-Square (RMS) error in centimeters per second for the three tracking regions (top, bottom, all). . . . . 77
4.3 Accuracy comparison for three diamond-shaped trajectories using StereoTrack (blue) vs. ground truth (red). The crosshair indicates the starting position ................................................. 80
4.4 Endless running game that uses walking-in-place but also allows players to side step to collect coins using StereoTrack ............................. 83
4.5 Performance results of the feasibility user study to evaluate StereoTrack’s ability to enable effective maneuvering in two example applications. Left: Percentage of player hits in the space shooter game. Right: Percentage of side coins missed in the endless running game. 84

5.1 Popular VR apps like Google Earth reduce the field-of-view during locomotion (e.g., tunneling) to block peripheral motion perception as to mitigate visual-vestibular conflict and to reduce VR sickness. However, prior studies suggest that reducing the field-of-view can impede spatial navigation performance in women. ......................... 88
5.2 Left: the FOV restrictor we used in the RY condition at its maximum FOV restriction. Right: the waypoint used in the triangle completion task. .......................................................... 94
5.3 Top-down view of the virtual environment we used in the experiment sessions ................................................................. 97
5.4 The triangle completion task we use in this study. $S =$ starting position, $E =$ estimated position, $W_1 =$ first waypoint, $W_2 =$ second waypoint, $W_2S =$ the vector from the second waypoint to the starting position, and $S^E =$ the vector from the starting position to the estimated position ............................................. 98
5.5 Spatial navigation performance measured in HPE. Results are aggregated per sex and FOV condition. RN = No FOV restriction; RY = Dynamically changing FOV. Error bars represent the standard deviation.  

5.6 Average and ending levels of discomfort. Results are aggregated per sex and FOV condition. RN = No FOV restriction; RY = Dynamically changing FOV. Error bars represent the standard deviation.  

5.7 Relative Simulator Sickness Questionnaire results aggregated by sex and FOV condition. RN = No FOV restriction; RY = Dynamically changing FOV. Error bars represent the standard deviation.  

6.1 Popular VR apps like Google Earth reduce the field-of-view during locomotion (e.g., tunneling) to block peripheral motion perception and mitigate visual-vestibular conflict and thus reduce VR sickness. However, prior studies suggest that reducing the field-of-view can impede spatial navigation performance in women.  

6.2 A side-view of the virtual Morris Water Maze we used in this study  

6.3 Left: Learning rate shown in terms of the slope. Right: Distance traveled. Data in both plots are summarized by sex and FOV condition. RN = No FOV restriction; RY = Dynamic FOV restriction. Error bars represent the standard deviation.  

6.4 Left: Object placement accuracy in terms of placement error. Right: Object placement latency. Data in both plots are summarized by sex and FOV condition. RN = No FOV restriction; RY = Dynamic FOV restriction. Error bars represent the standard deviation.  

6.5 The relative total severity, nausea, oculomotor discomfort, and disorientation results summarized by sex and FOV condition. RN = No FOV restriction; RY = Dynamic FOV restriction. Error bars represent the standard deviation.
CHAPTER 1
INTRODUCTION

The ability to create and live in alternative realities has been a desire for some of us. In a different reality, we can possess a different set of capabilities, change the consequences of our actions, and materialize our imagination. Through Virtual Reality (VR), loosely defined as a computer simulation that renders artificial stimuli to the senses and sometimes responds to our tracked actions, this desire can be satisfied. VR has finally emerged from research labs into consumers’ hands. Recent advances in VR headset technology regarding tracking, latency, refresh rate, resolution and optics have allowed for the launch of major consumer VR platforms such as HTC Vive, Oculus Rift, and Sony VR. Since the mid-twentieth century, attempts have been made to create VR systems that are natural enough to replace reality by giving the sense the illusion that what we see, hear, touch, and hopefully smell and taste is real. Though VR technology has not yet reached the stage to completely "fool" the senses, it has effectively been used in training, planning, evaluation, communication, and entertainment [302] in the fields of education, psychology, sociology, business, tourism, and journalism [266], not to mention gaming.

Among other research disciplines, Human Computer Interaction (HCI) researchers have been working on improving the utility of VR through the development and evaluation of 3D user interfaces that enable selection, manipulation and travel in virtual environments (VEs) [174]. Making or breaking the design of these interfaces could be the difference between enjoying the feeling of being present in another reality and suffering with discomfort, disorientation, or even nausea.

Ensuring that these interfaces are usable is a key aspect in HCI research in VR.
According to the International Organization for Standardization (ISO), the term usability can be defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [21]. One could infer from this definition that the usability of VR interaction techniques can be decomposed to three main goals: effectiveness (e.g., accuracy and precision), efficiency (e.g., time on task and cognitive load), and satisfaction (e.g., VR sickness and presence). A more important inference to make, however, is that these usability goals do not live in a vacuum. In other words, the importance and magnitude of these usability goals are contingent on two factors: target users and context of use. Unlike the assumption often made by some mainstream designers, there can be substantial differences between target users in terms of their physical and cognitive capabilities that may affect the usability of a given interface. The context of use can also have direct implications on the user experience of VR. Contextual factors that are applicable to VR include the target VR platform (PC VR vs. mobile VR), the demands of the virtual task, and the physical space at which the user experiences VR. Due to its versatility, it is important to ensure that the usability of VR does not break due to differences in what users are capable of or what is offered by the context of use, an endeavor known in HCI research as universal usability [264]. This dissertation addresses two key universal usability issues relevant to the user experience of VR: the gap in interaction quality between low-end and high-end VR platforms, and the effect of sex differences on the quality of VR user experience.

What follows are summaries of the chapters presented in this dissertation.
Chapter 2: A Survey of Virtual Locomotion Techniques

Human interaction with the virtual world can take the form of selection, manipulation, and travel. Among these three forms of interaction, enabling users to effectively navigate virtual environments through virtual travel interfaces, also known as locomotion interfaces, has been one of the most challenging problems in VR interaction design [174]. Unlike selection and manipulation interfaces, the design of virtual locomotion techniques is surrounded by challenges related to both the context of use and the target user. Preference is often given to navigate virtual environments through real walking because this is how we naturally navigate physical spaces in the real world. However, it is a common scenario that the virtual environment is much larger than the physical space. There are also cases when the VR platform has limited tracking capabilities that make virtual locomotion using natural walking impossible to achieve. Due to this, alternative virtual locomotion techniques have been introduced over the past 25 years to allow users to navigate unlimited virtual spaces without necessarily requiring sophisticated tracking capabilities. The fundamental design strategy across all virtual locomotion techniques is to define a remapping between physical user gestures and changes in the virtual viewpoint’s position, orientation, and scale. The design of a given remapping has associated trade-offs with direct implications on the quality of the user experience of VR that vary with the capabilities of target users and their context of use. In light of this, VR interaction designers need to have sufficient input to make informed decisions while designing virtual locomotion techniques for a target user group.

This is an overdue survey of virtual locomotion techniques that have been developed over the past three decades. It provides VR interaction designers with a
categorization of a multitude of gestural interactions that allow users to navigate virtual environments beyond the confines of the available physical space. The survey is concluded with a discussion of the trade-offs associated with each virtual locomotion techniques category. The information and insights provided by this survey would hopefully guide the design of future locomotion techniques that suit the capabilities of target users and their context of use, which is at the core of universal usability of VR. The survey also provides a foundation for the research presented in Chapters 4, 5, and 6.

Chapter 3: PAWdio: Hand Input for Mobile VR using Acoustic Sensing

The ability to experience VR by simply sliding your smartphone to a VR headset made of inexpensive cardboard allowed millions of users to experience VR at a very low cost. Due to their limited input options, mobile VR platforms have been largely ideal for passive virtual experiences that do not require user interaction. Attempts have been made to improve the interactivity of mobile VR platforms. On one hand, some of the introduced interaction techniques are unnatural, which can hinder the immersiveness of the virtual experiences that adopt them. On the other hand, techniques that are more natural can be resource intensive, which can limit the reach of these techniques only to powerful devices that can support their requirements. As a result, it would be challenging for users who cannot afford expensive high-end VR platforms to have access to natural forms of VR interaction, which is against the vision of universal usability of VR. Consequently, further research was needed to investigate if natural VR interaction can be achieved in mobile VR without adding special instrumentation.

Using only a regular earbud and the smartphone’s microphone, PAWdio offers
a one degree-of-freedom hand input for mobile VR. PAWdio utilizes the change
in frequency of the ultrasonic signals emitted from the earbud held by the user’s
tracked hand to calculate the Doppler shift that is used to calculate the hand dis-
placement in the physical space. The displacement is then used to update the
position of the virtual hand in VR. A user study with 18 subjects evaluates PAW-
dio with button input that is commonly available on mobile VR adapters. Results
with a 3D target selection task found a similar accuracy and usability, a signifi-
cantly slower performance, but higher immersion for PAWdio.

Chapter 4: StereoTrack: 180-degree Low-cost Acoustic Positional Tracking for
Mobile VR Platforms

Besides the hand input capability provided by PAWdio, the introduction of a nat-
ural form of locomotion is needed to improve the user experience of mobile VR
platforms. As introduced in Chapter 2, real walking is the most desirable locomo-
tion technique when the goal is to offer the highest degree of immersion. In order
to offer real walking in mobile VR, positional tracking needs to be supported. To be
aligned with the vision of universal usability, the design of the positional tracking
technique should neither require devices with specific requirements nor demand
special instrumentation.

StereoTrack is a 180° positional tracking method that uses acoustic sensing that
only requires a pair of regular speakers and the smartphone’s microphone. A num-
ber of studies evaluate StereoTrack in terms of precision, accuracy, and effect on
overall latency. The feasibility of StereoTrack was tested in the context of two
games that demonstrate its potential to enhance existing mobile VR interaction
options.
Chapter 5: The Effect of Field-of-View Restriction on Sex Bias in VR Sickness and Spatial Navigation Performance

As VR is envisioned to be a universal medium, it must be designed in a way that encourages the adoption of all users regardless of their differences. A recent survey revealed that less than 5% of VR users were women [99], which also agrees with another anecdote which estimates that the percentage of female users of a VR social platform to be only 7% [14]. This is drastically different from the demographics of video games users where women constitute around 48% of the user base [12]. Among the possible reasons behind low VR adoption by female users, evaluating the effect of commonly used VR interaction designs on the quality of the user experience to females is important to ensure that these designs do not exclude half of the VR potential user base.

Recent studies show that women are more susceptible to visually-induced VR sickness, which might explain the low adoption rate of VR technology among women. Reducing field-of-view (FOV) during locomotion is already a widely used strategy to reduce VR sickness as it blocks peripheral optical flow perception and mitigates visual/vestibular conflict. Prior studies show that men are more adept at 3D spatial navigation than women, though this sex bias can be minimized by providing women with a larger FOV. Our study provides insight into the relationship between sex and FOV restriction with respect to VR sickness and spatial navigation performance which seem to conflict. In this chapter, the results of a study that examines the effect of dynamic FOV restriction on sex bias in spatial navigation performance and VR sickness are presented. The design of this study evaluates spatial navigation performance primarily in terms of the skill of spatial updating, which was achieved with a triangle completion task. Analysis of the
results found a positive effect of dynamic FOV restriction on VR sickness in both men and women without impeding their spatial updating ability.

Chapter 6: Effects of VR Field-of-View Restriction on Spatial Learning and VR Sickness Among Men and Women Using a Virtual Morris Water Maze

The study presented in Chapter 5 is primarily concerned with spatial updating as a measure of spatial navigation performance. Follow-up studies are, therefore, needed to examine the effect of FOV restriction on other navigation skills. In light of this, spatial learning skill was the focus of the study presented in this chapter.

In this study, the effect of FOV restriction on participants’ spatial cognitive learning was evaluated using two tasks. The first was a virtual version of the Morris Water Maze task that is one of the golden standards to evaluate spatial memory while the second was an object placement task to complement the potential limitations of the Morris Water Maze task. Our findings confirm a sex difference in spatial learning ability. While our analysis did not find an effect of FOV restriction on sex differences in spatial learning ability, it was found effective in mitigating VR sickness in both sexes.

1.1 Publications

The work presented in Chapter 2 has resulted in the following publication:

2018. 20 pages.

The work in Chapter 3 has resulted in the following publication:


The work in Chapter 4 has resulted in the following publication:


The work in Chapter 5 has resulted in the following publication:


The work in Chapter 6 is being prepared for submission to an upcoming conference.
CHAPTER 2
A SURVEY OF VIRTUAL LOCOMOTION TECHNIQUES

2.1 Introduction

Finding virtual travel interfaces suitable for all possible combinations of virtual experiences, user preferences, platform capabilities, and physical space available is an endless endeavor. While tracking technology has reached a decent level of maturity to offer real walking (RW) in VR, offering a natural walking interface is not sufficient. VR travel interfaces must accommodate usage scenarios where users need to (1) navigate VEs of a space that greatly exceeds that of the physical one, (2) inspect virtual architectures at different elevations and perspectives, (3) cover great virtual distances safely with less physical exertion, or (4) travel in a way that is consistent with the experienced virtual activity (e.g., flying or surfing).

These demands motivated the emergence of VR locomotion interfaces that map body movements, mediated by digital sensory input, to the control of the virtual viewpoint translation, orientation and scale [145, 174]. Such mapping can vary in its fidelity from being completely natural (e.g., walking) or, on the contrary, completely artificial (e.g., using a 3D mouse). The nature of the task at hand also varies. Users could search for a target, explore a virtual scene, or maneuver obstacles [174]. Due to the variability in both mapping fidelity and the nature of the virtual task at hand, understanding the effect of the design of a VLT on the efficiency of virtual navigation has been one of the primary focus areas in VR locomotion research. Efforts to reach such understanding have been in the form of evaluations whose basis is a myriad of quality attributes that include usability, performance, naturalism, spatial abilities and cognitive abilities [189]. Another primary focus area in
VR locomotion research has been concerned with the development of new VLTs or the optimization of existing ones given certain quality attributes or virtual tasks. In this paper, we aim to survey the state of the art in the field of VR locomotion and discuss the design implications of VLTs in terms of strengths, weaknesses, and applicability. Researchers can use the outcome of this effort as a bird-eye view of the VR locomotion problems that have been tackled so far. This could inspire the introduction of better solutions to problems that have been attempted already or trigger ideas of new VLTs interaction modalities. Application developers can also use this survey as a catalog of VLTs to find candidate techniques that would satisfy their application’s navigation requirements.

This paper is organized as follows. Section 2.1 is an introduction that gives a brief history of VR and a motivation to survey VR locomotion. Section 2.2 gives an overview of the existing VLTs and results of their evaluations. Section 2.3 presents the existing taxonomies of VLTs proposed at different scales of generality. Finally, Section 2.4 discusses the strengths, the weaknesses, and the applicability of each VLTs category.

2.2 VR Locomotion Techniques Overview

In this section, we survey various VLTs. Similar to [145] and [174], we organize the surveyed implementations into walking-based, steering-based, selection-based and manipulation-based techniques. We further generalize walking-based techniques, similar to [130] to include all VLTs that involve repetitive motor actions. We adopt this classification criteria to survey VLTs since it is easier to relate to for users [174]. An overview of multiscale locomotion techniques is also included for
completeness.

2.2.1 Walking-based

The distinctive features of such techniques are that they require exertion from the user and that they have a high level of interaction fidelity [191]. Such techniques are, therefore, the closest to natural ways of moving around in VR. While some researchers thought of this category as being exclusive to walking techniques [174], it is useful to generalize these techniques to any VLT that involves repetitive motor actions [130]. Such generality allows for the inclusion of other VLTs that borrow from motions that are by-products of human walking such as head-bobbing and arm swinging.

Since the vast majority of VLTs in this category aim to mimic walking, LaViola et al. [174] subdivided these according to their degree of resemblance of the human gait cycle that consists of the stance and swing phases. The stance phase starts when the foot strikes the ground and ends when the toe is lifted off. The swing phase, on the other hand, starts with the end of the stance phase and ends when the foot strikes the ground again after a swing. In the light of this, walking-based VLTs are classified as full gait, partial gait, and gait negation techniques.

Full gait techniques

Full gait techniques include both the stance and swing phases. Common techniques in this category include RW and redirection techniques.

Real walking: this is the most natural way to travel in VR due to its high biome-
mechanical symmetry with respect to how humans move. Chance et al. [73] studied the effect of locomotion mode on spatial orientation through a path integration task. Among RW, joystick control and gaze-directed steering (GDS), participants were found to have significantly greater spatial orientation with RW compared to joystick control.

Usoh et al. [310] compared between RW, walking-in-place (WIP), and flying where participants were asked to move an object between two locations separated by a virtual pit. RW was found be significantly easier to use than both WIP and flying. Participants who used RW reported a significantly greater presence than those who used flying, while real walkers’ sense of presence was significantly greater than those using WIP.

Zanbaka et al. [344] compared RW at room scale to three other VLTs: RW in a limited physical space complemented with joystick control; 3 degrees of freedom (DoF) head tracking with joystick positional control; and joystick control with a monitor. The goal of the study was to investigate the effect of these VLTs on cognition, categorized to knowledge, understanding and application, and high mental processes in the context of an exploration task. Information understanding and application was significantly greater in RW at room scale compared to joystick control with a monitor and limited RW with joystick control while it was notably better than joystick control with head-tracking. With respect to high mental processes, RW at room scale was significantly better than limited RW with joystick control. When participants were asked to sketch the VE they explored, those who used RW at room scale performed significantly better than those who used joystick control with a monitor. RW at room scale was also found to be significantly superior to joystick control with a monitor with respect to the sense of presence and to
all other VLTs in the study with respect to comfort.

Suma et al. [286] examined the effect of travel techniques on navigation and wayfinding abilities, where participants were asked to explore a 2-story 3D maze. Results showed that the RW group significantly outperformed those who used GDS and hand-directed steering (HDS) in object placement, task completion time and collision avoidance with walls of the maze.

Suma et al. [291] compared RW, natural walking in real world, and GDS to investigate their effect on VR sickness. Participants were asked to explore a complex maze using their assigned locomotion method for 5 minutes. Results showed that RW had a significantly increased overall simulator sickness score with a significantly greater disorientation score than the other two forms of locomotion, suggesting that a VLT would be less VR sickness inducing than a natural one when navigating complex VR environments.

Ruddle et al. [255], evaluated participants’ performance in terms of time, accuracy and speed in an experiment with 3 VLTs: joystick control with desktop display, GDS, and RW. Participants were asked to traverse a 24m-long route 10 times using the VLT they were assigned to. Results showed that subjects who used RW completed the task in significantly less time, with fewer collisions and less time being stationary.

Ruddle and Lessels [254] studied the effect of translational and rotational body-based information on users search abilities. Participants were evenly assigned to 3 groups: RW, GDS, and keyboard/mouse control with a monitor. Participants were asked to search for 8 targets in 16 boxes around the VE. When their performance was measured in terms of the number of times a box was rechecked (imperfect
search) or not (perfect search), participants who used RW had significantly more perfect searches than the other two groups. When perfect searches were made, real walkers were found to travel paths that are closer to the shortest path than the GDS and keyboard/mouse control groups (7% vs. 32% and 46% closer to the shortest path, respectively). Real walkers were also superior to the other two groups when imperfect searches were made in terms of the number of box rechecks. Participants who used RW also had significantly fewer collisions and missed targets. Similar results were obtained in a second experiment with a VE with less visual detail.

Nabiyouni et al. [205] compared the performance and usability of three locomotion interfaces that varied in their interaction fidelity: RW (natural interface), the Virtusphere [192] (semi-natural interface), and gamepad control (non-natural interface). Participants were asked to follow straight and right-angled paths to measure their completion time and amount of deviation from the path. The usability of each VLT was also measured subjectively. Both RW and game-pad VLTs were found to be significantly faster, more accurate, easier to learn and less fatiguing than the Virtusphere. RW was also found to be significantly more natural than the Virtusphere.

Attempts to enable RW in VR date back to late 1960s with Sutherland’s work on head tracking [295]. Several implementations followed since then and the most notable of which is the UNC Tracker Project [6] whose final product was the HiBall tracker [323] that was commercialized in late 2000s. Meyer et al. [195], Bhatnagar [37], Welch & Foxlin [324], and Baillot et al. [251] survey early implementations of VR positional tracking systems until early 2000s. The main motivation of early implementations was to improve the robustness of tracking while recent implementations were largely driven by different motivations such as reducing cost [79, 231]
and improving the scalability of the tracking solution [105, 183, 236]. With the emergence of consumer VR platforms such as the Oculus Rift and the HTC Vive, RW in VR has been enabled at a relatively high quality and low cost, making VR natural travel research less concerned about robustness and cost while being more focused on the problem of scale.

**Redirection Techniques**: achieving a VLT that is fully natural and unrestricted by physical space has been considered as one of the grand challenges of VR travel research [174]. An attempt to realize this vision is the implementation of redirection VLTs, whose underlying goal is to keep the user within the confines of the physical space while being able to travel in a larger VE. Razzaque et al. [244] exploited the dominance of the human visual system over the vestibular system to make users walk in an unlimited straight path in the VE while actually moving on a curve in the physical environment. Users' viewpoint was manipulated by injecting imperceptible rotational distortions when they were moving in a straight path towards a waypoint. To compensate for this effect, users unknowingly moved either in or against the direction of the induced rotations. Users head rotations were also distorted to reorient them away from the boundaries of the physical environment. Since the publication of their work, research in this domain has taken off several directions that include developments of more redirection techniques, studies of viewpoint manipulation detection thresholds, and the exploration of different redirection cues.

We use both classifications by Suma et al. [287] and Steinicke et al. [280] to guide our survey of redirection VLTs. Suma et al. [287] characterize such techniques in terms of their redirection tactic, continuity, and subtlety. The used redirection tactics often depend on either controlling the user's physical orientation or
translation, both having the goal of keeping the user within the limits of the tracked space. Manipulation of orientation and translation can either be applied at a certain rate (continuous) or just once (discrete). These manipulations can be either imperceptible (subtle) or otherwise noticeable by the user (overt). Steinicke et al. [280] focuses on the kinds of orientation and translation control rates, known as gains, that are applied to what they refer to as the locomotion triple that is composed of three vectors: the strafe vector $s$, the up vector $u$ and the direction of walking vector $w$. Three gains were also identified. The translation gain scales the virtual translations with respect to the physical ones, preferably to vector $w$ [280, 137]. Rotation gains scale rotations made by the user, where roll, pitch, and yaw rotation gains are applied to the $w$, $s$, and $u$ vectors, respectively. Both translation and rotation gains are used to multiply the user’s actual translations or rotations. Curvature gains, on the other hand, are added rotational offsets to the virtual viewpoint that make users physically walk on a curve while they are virtually walking straight. A less common form of curvature gains are added translational offsets to the virtual viewpoint while users turn their head, forcing them to compensate for these offsets by walking to the opposite direction [280]. Langbehn et al. [165] recently introduced a fourth gain, the bending gain, to include applied gains to curved virtual paths.

Nitzsche et al. [218] developed a subtle continuous redirected walking (RDW) technique, dubbed as Motion Compression, that is more generalized than Razaque et al.’s [244] with respect to the reliance on predefined paths. Instead, Motion Compression induces curvature gains depending on the user’s predicted path in the VE. Motion Compression computes curvature gains using an optimization function that minimizes the amount of deviation from the physical path for manipulations to be as imperceptible as possible. Engel et al. [96] similarly aimed
to minimize the amount of dynamic rotation gains with an optimization function that considers users’ discomfort and probability of collision with the walls of the tracked space. Goldfeather and Interrante [113] used subtle curvature and translation gains to minimize the amount of deviation from the physical path with the least possible collisions with the boundaries of the physical space. Zhang and Kuhl [347] proposed another redirection VLT that uses translation and rotation gains, where the latter is guided by heuristics that are based on prior knowledge about the physical environment. Bruder et al. [64] developed a redirection technique for 3D architectural model exploration named ArchExplore, where users can travel beyond the tracked space using rotation, curvature and translation gains. The implementation also uses portals for users to select a specific exploration space. Steinicke et al. [277] applied curvature gains to allow users to play a geocaching game in a virtual city larger than the tracked space. A few implementations capi-
talized only on translation gains. An example is the Seven-League-Boot [138] that scales the user’s physical translations in the direction of their walking.

In an attempt to liberate subtle reorientation techniques from assumptions associated with the VE and the user’s task, Razzaque [243] proposed a set of generalized steering algorithms that steers the user towards the center (steer-to-center), in circles (steer-onto-orbit) or towards given targets (steer-to-alternating-targets) by primarily applying imperceptible curvature gains. Later implementations aimed to build on these algorithms to offer RDW that works with fewer assumptions about the task and VE [128] and works in more constrained VEs [351, 208, 30].

Instead of manipulating the viewpoint rotations and translations, some research efforts considered the manipulation of the environment’s geometry. Suma et al. [289] proposed an approach that changes the doorways’ location at the VE using the change blindness illusion to keep the user within the boundaries of the tracking space. Impossible spaces [293] is a redirection technique that allows users to imperceptibly navigate a virtual environment larger than the physical tracking space using self-overlapping virtual architectures. Vasylevska et al. [314] later developed flexible spaces, an algorithm that utilizes both change blindness and self-overlapping architectures to automate redirection.

Another adopted redirection strategy is to force the user to make excess head rotations by using visual distractors [225, 75]. Generally, rotation gains are applied when users respond to a distractor by turning their head to either direction.

The techniques we surveyed so far use visual manipulations to achieve redirected locomotion. Matsumoto et al. [186] explored the feasibility of using visual-haptic feedback through the development of the unlimited corridor, where users
walk in a straight path in the VE while physically walking around a physical wall. This is achieved by synchronizing haptic feedback from touching the physical wall with the virtual hand that is seen touching a corresponding virtual wall.

Due to the limitations of the current redirection techniques, chances of their failure to redirect the user away from the boundaries of the physical environment are inevitable. Users have to be instructed in such cases to "reset" their orientation and/or position before they could continue navigating the VE. To this end, Williams et al. [332] proposed three resetting techniques. Freeze-Backup stops the manipulation of the user’s viewpoint until they make several steps backwards after which the viewpoint is unfrozen. Freeze-Turn has a similar behavior except that the user is instructed to yaw by 180°. The 2:1-Turn resetting technique instructs the user to make a 180° physical turn while the viewpoint is turned by 360°. Further improvements were proposed [82] to make Freeze-Backup and 2:1-Turn more user-friendly and flexible in constrained physical spaces. Cirio et al. [83] proposed two resetting techniques for 3-walled CAVE environments. The goal of these techniques is to reset the user’s position when they reach the boundaries of the physical space and to reset the user’s orientation when they are nearly facing the CAVE’s missing wall (to avoid breaking the presence). In the first technique, position reset is signalled using a no-way sign when the user approaches a wall while turn sign is used to instruct the user to move away from the missing wall. In the second technique, a virtual bird is used to deter the user from colliding against a wall by angrily flapping its wings against the user’s face until they reset their position. The same technique is used when the user faces the missing wall until they reset their orientation. Having a similar goal in mind, LaViola et al. [173] proposed applying rotation gains to the VE rendered on a three-walled CAVE in a direction opposite to the user’s rotation depending on the angle between their
waist and the VE’s forward vector as well as their distance from the CAVE’s back wall. Similarly, Razzaque et al. [245] proposed redirected WIP, a subtle continuous reorientation technique that applies imperceptible rotation gains to move the user away from the CAVE’s missing wall towards the front wall as they walk in place depending on their virtual speed, head orientation, and head angular velocity. Freitag et al. [106] achieved overt discrete reorientation by using portals that move the user to their desired destination in the VE while facing away from the physical environment boundaries. Users select their target location by creating a target portal. A start portal then automatically appears in a physical location that keeps users away from the CAVE walls. Bookshelf and Bird [342] are teleportation techniques that are designed to blend with the narrative of the presented VR experience. Users freely walk in the VE until they decide to move to a virtual location beyond the tracked space. Users are then teleported to the desired destination through a metaphor (e.g., bookshelf or a bird) that conforms with the narrative of the VR experience. Lubos et al. [181] proposed Safe and Round, an overt continuous reorientation technique that aims to enable RDW at room-scale. The technique works by applying overt curvature gains to the user’s viewpoint when they exit a defined safe region until their re-entry. Sargunam et al. [257] proposed a continuous reorientation technique for the user’s head in seated VR experiences. As rotation gains are used to allow for 360° viewpoint control without the need to physically rotate as much, the proposed technique resets the user’s head to its physical forward direction by applying gradual rotation gains during travel. A summary of the surveyed redirection techniques is shown in Table 2.1.

Several studies focused on determining the detection thresholds of the induced translation, rotation, curvature and bending gains. Steinicke et al. [278] aimed to determine the limits of translation, rotation, and curvature gains in two-
alternative-forced-choice tasks. Results showed that translations could be imperceptibly manipulated by 22% more or less than their actual translation rate while rotation gains were considered imperceptible when they were 10% less or 68% more than their perceived virtual rotation. It was also shown that curvature gains with radius of at least 24m made subjects perceive that they were walking straight. Another study by the same authors [279] reported translation gains thresholds to be 14% less or 26% more, rotation gain thresholds as 20% less or 49% more, and curvature gain radius of 22m. The former rotation gain thresholds were also confirmed in a follow-up study [65] that also found that the subjects were less sensitive to rotation gains as their rotation angle is larger, and vice versa. Neth et al. [209] studied the effect of walking velocity on the sensitivity to curvature gains and found that subjects were significantly less sensitive to curvature gains when they walked slower. Zhang and Kuhl [348] examined the difference between abrupt and gradual rotation gains. Participants were asked to make a 360°-turn while varying the rotation gains during the turn and no difference was found between abrupt and gradual rotation gains. Grechkin et al. [118] aimed to analyze the effect of adding translation gains on curvature gains threshold and to revisit the estimation of the minimum detection threshold of curvature gains. It was found that translation gains do not cause an increase on curvature gains detection threshold and that the estimation of curvature gains threshold can be as low as 6.4m when a different threshold estimation method was used. Paludan et al. [221] examined the effect of the visual scene’s density on rotation gain threshold and no difference was found between 0, 4 and 16 objects in the scene. In a recent study, Langbehn et al. [165] attempted to determine the imperceptibility threshold of the bending gain and found that a bending gain of 4.35 times the real bending radius would go unnoticed by subjects. Serfain et al. [263] aimed to estimate the detection thresholds
of rotation gains when acoustic redirection cues were used. It was found that subjects can be rotated 12\% more or 20\% less than their perceived virtual rotation. A later study by Nilsson et al. [217] found no effect of adding sound cues on rotation gain detection thresholds. Schmitz et al. [259] proposed the threshold of limited immersion to establish a relationship between the amount of rotation gain and the point at which self-reported presence breaks. Rotation gains were continuously increased and decreased while participants took part in targets collection task and were asked to report breaks in presence. While the lower limit of the threshold of limited immersion was comparable to the reported detection thresholds (0.58 and 0.67, respectively) [279], a significant difference was found for the upper limits (1.85 and 1.24, respectively) [279]. It was also found that rotation gains greater than 1 had less effect on the rate of breaks in presence compared to gains less than 1.

Other studies focused on examining the effect of gains on users performance. Williams et al. [331] studied the effect of translation gains on the subjects’ ability to orient themselves, timely and accurately, to previously seen targets in two experiments. The results showed that the amount of translation gain had no significant effect on subject’s latency in both experiments with contradicting results on the effect of varying gains on the subjects’ accuracy among the two experiments. Xie et al. [338] studied the effect of combining translation gain and the resetting methods implemented by Williams et al. [332] in two experiments similar to the former one by Williams et al. [331]. It was found that amount of resets had a significant effect on subjects’ spatial accuracy. Bruder et al. [63] found that adding curvature gain larger than $\frac{1}{10}$, which corresponds to an arc with a diameter of 10m, had a significant effect on walking behavior by asking participants to follow a virtual sign in a 7m long straight path. Such gain was also found to have a significant
effect on both spatial and verbal memories when participants were asked to perform two-back cognitive tasks. Similarly, Hodgson et al. [128] conducted a study to examine the effect on their subtle RDW technique with dynamic curvature gains on participants spatial memory and no significant effect was found. Kopper et al. [163] studied the effect of rotation gains on participant’s performance in terms of visual scanning and counting abilities during head rotations. While the amount of rotation gain had no notable effect on visual scanning performance, participants counting performance significantly worsened as more rotation gain was applied. Ragan et al. [241] examined the effect of rotation gains in both head-mounted display (HMD) and CAVE platforms on performance, spatial orientation, and VR sickness in a naive search task conducted over a practice session that involved different levels of rotation amplification and an assessment session with no amplification. Rotation gains had no significant effect on search performance during practice. However, participants who practiced with the highest levels of rotation amplification found significantly more targets in the assessment session than those who practiced with lower amplification. It was also found that high rotation gains experienced with the HMD were linked to greater spatial disorientation, greater VR sickness, and less usability. Freitag et al. [108] studied the effect of a range of rotation gains (0.8 to 1.18) on spatial knowledge, VR sickness, presence, subjective cognitive load, and measured cognitive performance in a CAVE VE. Results showed that the rotation gains under analysis had no significant effects on any of the measures except for a degrading spatial knowledge in first-time CAVE users who experienced stronger rotation gains.

Some research efforts were concerned with comparing the effectiveness of different redirection techniques with one another. Peck et al. [224] compared the effectiveness of 4 redirection techniques: motion compression [218], 2:1-Turn [332], the
amplified rotations described in the traditional RDW by Razzaque [243], and the authors’ reorientation with distractors technique. It was found that both the amplified rotations and reorientation with distractor techniques were ranked better on the scales of sensed presence, user preference, and naturalness. Suma et al. [292] compared the effects of Peck et al.’s [225] reorientation with distractors and reorientation with change blindness [288] on spatial orientation both in virtual and real world. Change blindness reorientation caused more disorientation than reorientation with distractors. The latter technique was also found to have very similar spatial orientation performance to the control condition that involved no reorientation. It was also found that reorientation with distractors had a strong influence on participant’s spatial updating both in virtual and real worlds. Langbehn et al. [166] compared between RDW and teleportation along with joystick-control locomotion in pointing and spatial arrangement tasks. RDW outperformed the other two techniques in the spatial arrangement task while RDW and teleportation were comparable in terms of preference while they were both preferred over joystick-control. Except for joystick-control locomotion, neither RDW nor teleportation caused a significant increase in VR sickness and none of the locomotion techniques had a significant effect on presence. Hodgson and Bachmann [126] compared the performance of four generalized RDW algorithms. Steer-to-Center was superior to Steer-to-Orbit, Steer-to-Multiple-Targets and Steer-to-Multiple+Center in all performance measures, namely: ability to contain users in terms of mean and maximum distance traveled to the center, maximum physical distance traveled in terms of the number of wall contacts, and mean redirection rate. Steer-to-Orbit was found to have better performance than Steer-to-Center for long and straight paths. In a later study, Hodgson et al. [127] compared between Steer-to-Center and Steer-to-Orbit in a constrained VE (virtual grocery store) and the latter was found
to have significantly fewer wall contacts, marginally less mean physical distance covered, and marginally better task completion time.

**Partial Gait Techniques**

The majority of partial gait techniques take advantage of the stance phase, where users step in place without making any physical translations. Slater et al. [267] proposed the first WIP implementation with the virtual treadmill. When users walk in place, their resulting head oscillations are fed to a neural network that, if a walking action was detected, translates the users’ viewpoint forward in the direction they are looking. The locomotion interface was later extended to support climbing steps and ladders [269]. The virtual treadmill had four key limitations: its underlying neural net has to be trained for each user, only forward travel is supported, the direction of travel is coupled with head direction, and a proxy measure to speed (head oscillations frequency) is used. Research efforts that followed aimed to overcome some of these limitations. Templeman et al. [302] proposed the Gaiter WIP VLT that uses force and inertial sensing to transform steps in place to translation in the VE. Displacement and angle of the knee determined speed and direction of travel, respectively. The Gaiter offers forward, backward, and lateral (strafing) movements and does not require prior neural net training like the virtual treadmill. Aside from the traditional WIP gestures that resemble marching, Nilsson et al. [212] proposed two other gestures: wiping and tapping in place. With wiping in place, users move their feet alternately backwards against the floor while bending the knee of the moving leg. With tapping in place, on the other hand, users alternate the movement of each heel while having the front of their feet in contact with the ground. Similarly, Guy et al. [120] proposed two partial gait locomotion
techniques. In the first technique, users step with one foot forward or backward to translate in the corresponding direction; while they bend their knees alternately for translation in the second. Rotation is enabled via upper-body gestures. Zielasko et al. [349] proposed two WIP VLTs for seated virtual experiences. In adapted WIP, users move their thigh that is attached to a smartphone upwards and downwards to translate forward in the direction of their head with a speed that corresponds to their thigh movement frequency. In the accelerator pedal technique, a heel tapping gesture of the foot attached to a smartphone is transformed to translation, where lowering the heel below a predefined zone moves the virtual viewpoint forward in the direction of the head with a speed that is derived from the distance of the heel from the predefined zone. Lifting the heel above the predefined zone translates the viewpoint backwards.

Other implementations utilized dedicated stepping platforms to offer WIP. Bouguila and Sato [48] developed a WIP platform that utilizes a turntable equipped with pressure sensors to detect stepping gestures. Turning in place was enabled via tracked infrared (IR) markers that can be fitted on the user’s waist or head. A successor implementation enabled jumps and squats in addition to the simulation of walking over uneven terrains [47]. The walking pad [46] also uses pressure sensing, but uses switch sensors to eliminate the need for an IR camera to detect rotations in place. A similar implementation was also proposed by Lala and Nishida [164]. Zielinski et al. [350] implemented a WIP VLT, dubbed as Shadow Walking, that captures the shadow of the user’s feet using an under-floor camera. Both forward and strafe movements are supported. Williams et al. [329] used a Wii balance board through which users could walk in place in the direction of their gaze. In a later implementation, a platform-less WIP technique using a Microsoft Kinect was proposed [330].
Several efforts were focused towards making WIP more natural by primarily providing *speed profiles* that are better approximations of natural walking. Fujita [110] proposed a responsive WIP technique that attempts to reduce the travel speed estimation delay. Locomotion speed is estimated as a function of hip joint difference angle, stepping frequency, stepping amplitude, and leg length. Yan and Allison [340] used data collected from the back of subjects’ calf while physically walking to calibrate their WIP implementation for it to offer more realistic speeds. Kim et al. [157] used a sensor fusion step detection approach that uses accelerometer and magnetometer data to offer WIP that accommodates dynamic stepping frequencies ranging from .75Hz to 2.8Hz. Feasel et al. [98] implemented the Low-Latency, Continuous-Motion WIP (LLCM-WIP) that aims to reduce the latency, smooth the virtual viewpoint transition between steps, and give more granular speed control of WIP locomotion by processing heel-tracking data at each movement frame. Improvements to the LLCM-WIP, such as the GUD-WIP [327] and SAS-WIP [66] were later proposed. A similar recent effort was made by Bruno et al. [67], where a WIP speed estimation model based on the lower limbs skeletal data was proposed. The range of the perceptually natural walking speeds of WIP was studied by Nilsson et al. [214, 215]. Results of these studies showed a positive trend in the relationship between stepping frequency and the speed gains perceived as natural [215]. The results also showed that perceived natural speeds were underestimated by a factor of 2 and such underestimation is inversely related to the field of view [214].

Some implementations *substitute* the use of legs with other body parts to offer WIP locomotion. Kim et al. [154] implemented the Finger Walking In Place (FWIP) technique. Users slide two of their fingers alternately to move forwards or backwards while radial sliding gestures enable virtual turning. Speed is controlled
via the length of the sliding gesture. A similar implementation was proposed for multi-touch enabled smartphones [155]. McCullough et al. [188] used a wearable armband to implement an arm-swinging WIP VLT, where users swing both of their arms to travel forward in the direction of their head. Arm-swinging frequency is used to control the speed of travel. Sarupuri et al. [258] implemented Trigger Walking, a technique that mimics bipedal locomotion by having users alternate between the trigger buttons of the right and left controllers. Speed is controlled by changing the controller’s angle with the frontal plane. Three methods are proposed to control travel direction: using the heading angle of one controller with respect to the user’s heading, using the average bearing of both controllers, or using the user’s heading. Terziman et al. [303] offered a WIP implementation using head gestures. Lateral head motions were transformed to virtual forward translations while upward and downward vertical head motions were transformed to jumping and crawling, respectively. Users roll their head to either side to change the direction of travel. Speed of horizontal and vertical travel is controlled by the lateral and vertical head oscillations amplitudes, respectively.

With the recent trend of mobile VR, some WIP implementations for such platform were proposed. Tregillus and Folmer [306] implemented VR-Step, a WIP implementation for mobile VR that is enabled via inertial sensing. VR-Step moves users forward in the direction of their head in a speed that is approximated by their stepping frequency. The technique has also the ability to detect vertical translation. Pfeiffer et al. [229] proposed a similar implementation with the addition of a limited ability to look around while moving forward and the ability to move backwards. A summary of the surveyed WIP implementations is shown in Table 2.2.
Several evaluations of WIP techniques have been concerned with their naturalness as proxies of natural walking. In a study by Slater et al. [270], WIP scored more subjective presence compared to flying. This finding was consistent with a later study by Usoh et al. [310], where WIP was found to score significantly greater presence than flying while being comparable with RW. WIP scored significantly less than RW in terms of ease according to that study, however. Tregillus and Folmer [306] compared WIP to auto-walking and subjects rated WIP significantly greater in terms of subjective presence and intuitiveness. In another study by Muhammad et al. [298], subjects felt more present while navigating 360° videos compared to the traditional travel method. Nilsson et al. [212] compared the traditional marching WIP with two other WIP gestures, wiping and tapping, in terms of naturalness, presence and unintentional positional drift. Tapping in place was found to be significantly better than wiping with respect to naturalness. Tapping also scored significantly less than wiping and traditional WIP in terms of required effort and amount of unintentional positional drift.

Other efforts have examined the effect of WIP on spatial abilities and understanding. Williams et al. [329] found WIP to be comparable to RW with respect to spatial orientation. In a virtual maze navigation task, Peck et al. [226] found WIP to be significantly worse than RDW with distractors in terms of navigation and wayfinding abilities. Terziman et al. [304] conducted a spatial and temporal analyses of the trajectories made by WIP compared to those by joystick control. Results showed that subjects had greater control over their speed with WIP, but had more difficulty making natural turns while traveling compared to the joystick. Xu et al. [339] compared between WIP, joystick control, and teleportation in a spatial knowledge acquisition task and found all three methods comparable in performance.
Some studies examined the effect of using different body parts, e.g., fingers or arms, to implement WIP on spatial understanding and naturalness. Kim et al. [156] found that subjects had better route knowledge acquisition abilities when they walked in place with their fingers compared to flying. McCullough et al. [188] found WIP with arm swinging to be comparable to RW with respect to spatial orientation, which contradicts with a later study by Wilson et al. [333] who reported that subjects scored less than both RW and traditional (leg-based) WIP when arm swinging was used in a spatial orientation task. Nilsson et al. [213] studied the difference between traditional WIP, arm swinging, and hip movement (by swinging hips while being in place) with respect to naturalness, presence, and unintentional positional drift. Both traditional WIP and arm swinging were rated significantly more natural than hip movement. Traditional WIP, however, was found to be a significantly closer proxy to natural walking in terms of physical expenditure than arm swinging. Traditional WIP scored a comparable level of sensed presence to that of both arm swinging and hip movement while participants felt significantly more present when using arm swinging compared to hip movement. Arm swinging and hip movement were comparable in their measured unintentional positional drift, which was significantly lower than that of traditional WIP. Zielasko et al. [349] compared two seated WIP techniques (marching vs. tapping), Terziman et al.’s [303] head shaking technique, LDS and gamepad control in a virtual graph analysis task. Head shaking was found slower than both gamepad control and WIP with tapping while no significant differences were found in error rates.
<table>
<thead>
<tr>
<th>Impl.</th>
<th>Pattern</th>
<th>Features</th>
<th>Part</th>
<th>Control</th>
<th>Part</th>
<th>Control</th>
<th>Sensing</th>
<th>Movements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slater et al. [267]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head</td>
<td>yawning</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Jeppiepenman et al. [302]</td>
<td>marching</td>
<td>feet &amp; knees</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward, backward, strafe</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump, squat</td>
</tr>
<tr>
<td>Bouguila et al. [46]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [46]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump, squat</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump, squat</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump, squat</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump, squat</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
<tr>
<td>Slater et al. [269]</td>
<td>marching</td>
<td>head</td>
<td>head</td>
<td>frequency</td>
<td>head (walk) &amp; hands (climb or descend)</td>
<td>yawning (walk) &amp; hand position (climb or descend)</td>
<td>magnetic</td>
<td>forward, climb, descend</td>
</tr>
<tr>
<td>Bouguila &amp; Sato [48]</td>
<td>marching</td>
<td>feet</td>
<td>head</td>
<td>frequency</td>
<td>head or torso</td>
<td>yawning</td>
<td>infrared &amp; pressure</td>
<td>forward,</td>
</tr>
<tr>
<td>Bouguila et al. [47]</td>
<td>marching</td>
<td>feet</td>
<td>feet</td>
<td>frequency</td>
<td>feet</td>
<td>yawning</td>
<td>pressure</td>
<td>forward, jump</td>
</tr>
<tr>
<td>Templeman et al. [302]</td>
<td>marching</td>
<td>feet</td>
<td>knees</td>
<td>frequency</td>
<td>knees</td>
<td>yawning</td>
<td>inertial &amp; force</td>
<td>forward,</td>
</tr>
</tbody>
</table>

**Table 2.2**: A summary of WIP implementations described in terms of the foot motion pattern, speed control, direction control, sensing requirements, and movements supported. Speed and direction controls are specified in terms of the body parts involved and the control mechanism used.
Gait Negation Techniques

These techniques aim to provide the full gait cycle while keeping the user stationary. They, therefore, overcome the room-scale issue with RW and provide a greater level of fidelity than the partial gait techniques. All of these methods depend heavily on dedicated mechanical devices which include treadmills, step-based devices, and low-friction surfaces [174].

Treadmills were first proposed by Brooks et al. to offer more natural means of virtual architectural walkthroughs [62, 60], where a custom-made treadmill utilized a bicycle handlebar for steering. Darken and Carmein [89] proposed a different implementation with their Omni-Directional Treadmill (ODT), at which two perpendicular treadmills were used, one inside the other. A qualitative evaluation with one subject compared between walking naturally and using the ODT while performing representative locomotion moves such as resting, walking, jogging, acceleration, change in direction, and maneuvering. Effects of tracking quality and user re-centering criteria were found to limit the subject’s ability to maneuver freely and achieve responsive locomotion experience. Further improvements to the ODT were proposed in [86]. The Torus omnidirectional treadmill [139] similarly employs two groups of treadmills to move the user along the X and Y directions. A distance perception study with 18 participants showed that the Torus resulted in significantly lower distance estimation error than joystick-control and locomotion using a motorized chair. The ATLAS [199] is another omnidirectional treadmill that dynamically adjusts its running speed according to that of the user. The Sarcos Treadport [132, 129] is a treadmill that simulates contact with physical constraints (e.g., walls), slopes, and inertial forces by manipulating the pushing and pulling forces of a tether around the user’s torso. Users can walk forwards
or backwards at a speed that is proportional to the user’s distance from the center. The rate of turning is proportional to either the angle between head direction and torso or the amount of a sidestep to either side. With a set of ball bearings arranged around a disc, Huang et al. [135] implemented the gait sensing disc that offers omnidirectional locomotion while being stationary. This is similar to the Ball Array Treadmill [206] and the Cybercarpet [261]. A more recent implementation of an omnidirectional treadmill is the CyberWalk [273] that distinguishes itself from previous implementations by its ability to handle abrupt changes in the users’ walking speed to keep them stable and close to the center of the treadmill. CyberWalk was compared to RW in a study that evaluated participants’ walking behavior and spatial updating performance while following a moving target on an arc. No notable differences were found between CyberWalk and RW in both measures. Several research works on the development of VR treadmills focused on addressing more locomotion scenarios other than walking on flat surfaces by simulating slopes [309, 131], uneven terrains [284, 219], and stairs [123].

**Step-based devices** offer another alternative to provide full gait navigation while being stationary. Roston and Peurach [253] proposed a virtual locomotion device that allows users to navigate VEs by stepping over two motion platforms. Wang et al. [319] implemented a step-based locomotion device that consists of two foot boards that follow the motion of the user’s foot while physically walking to move them back to their initial position. Iwata et al. [143] similarly implemented the Gait Master, a step-based locomotion device that is able to simulate uneven virtual terrains. Boian et al. [39] used a pair of robot manipulators to implement a step-based locomotion device that offers a more realistic walking experience with respect to terrains with different shapes and physical properties. A similar work was also proposed by Yoon and Ryu [341].
The Wizdish [296] is a concave low-friction surface at which users slide their feet in opposite direction to travel in VR, which is similar to the implementation proposed by Grant et al. [117]. The platform was made slippery for users to slide their feet back and forth similar to the wiping WIP gesture. Speed was estimated in accordance with the user’s wiping magnitude. Iwata and Fuji [141] implemented the virtual perambulator that offers omnidirectional locomotion using a pair of specialized sandals with a low friction film at the middle of their soles. The naturalism of the virtual perambulator was evaluated in terms of participants’ ability to perform rhythmic feet alternation and smooth change of direction. Out of the 235 participants, six participants failed to perform rhythmical walking while seven failed to turn smoothly.

Aside from treadmills, step-based, and low-friction devices, several gait negation locomotion devices were implemented. The Cybersphere [101] is a spherical projection and locomotion system that offers omnidirectional virtual travel as a result of physical walking of the user inside a sphere while being stationary, which is similar to the Virtusphere [192]. Iwata and Noma [142] developed the CirculaFloor that consists of a set of tiles that detect the user’s physical walking velocity and move accordingly to pull the user back, creating an infinite walking surface. The String Walker [144] is another gait negation locomotion device that ties the user’s feet with strings attached to a turntable. The pulling force of the strings is used to control a motor-pulley that is transformed to virtual omnidirectional locomotion.
2.2.2 Steering

The key feature associated with steering VLTs is the continuous control of direction [174]. Speed control can be part of a given steering technique though it’s not the primary focus of its design. Steering VLTs can be categorized as spatial steering techniques (i.e., controlled using parts of the user’s body) or physical steering techniques (i.e., controlled using vehicular props) [174].

Spatial Steering techniques

Steering in these techniques is controlled using body gestures that are mapped to control of virtual direction. In the light of this, steering can be controlled by gaze, hand, body leaning, or torso.

In gaze-directed techniques, the direction of virtual travel follows that of where the user is looking. While such techniques are commonly known as gaze-directed [196], few implementations track the movement of the user’s eyes [281], but rather depend on the tracked orientation of the user’s head. A few studies compared between GDS and RW. Ruddle and Lessels [254] compared between RW and GDS to investigate the relationship between the amount of body-based information and navigation performance in search tasks. Participants who used GDS performed significantly worse than real walkers. Suma et al. [291] studied the effect of the VLT (RW vs. GDS) on the reported VR sickness in complex virtual environments. Participants who used GDS reported significantly less VR sickness after the study compared to those who used RW. The common criticism of GDS techniques is that they limit the user’s ability to look around while navigating in a certain direction [196]. This limitation is overcome by using HDS, torso-directed steering (TDS),
and lean-directed steering (LDS).

Several **hand-directed steering** techniques were implemented using metaphors such as flying or skiing. Robinett [250] proposed an early implementation of VR locomotion that mimics flying, where hand input determined the direction of travel. Fairchild et al. [97] proposed flying among the possible VLTs to be used in their "Heaven and Earth" VR system, where the orientation and relative position to the user’s head controls the virtual direction and velocity, respectively. A similar technique was also implemented by Boudoin et al. [45] with their FlyOver 3D HDS navigation model. Bowman et al. [50] developed a two-handed flying VLT using a pair of tracked data gloves, where the direction and speed of travel are determined based on the direction and length vector between the hands, respectively. Haydar et al. [122] implemented a HDS gesture that utilizes how tanks are steered to control the virtual direction where, for instance, turning left is achieved by moving the right hand forward and the left one backwards. Speed control, on the other hand, borrows from skiing where the angle between the hands determines the rate of translation. Cabral et al. [69] defined a set of two-handed gestures to control translation, rotation, and scaling of the virtual viewpoint as well as to switch between navigation, manipulation and visualization modes. Cirio et al. [82] implemented a HDS technique for users to be able to navigate the VE when they reach the boundaries of the physical tracking space, represented as a barrier tape in the VE. When users "push" through the barrier tape, the direction and amount of tape penetration are used to control the direction and speed of travel, respectively. Cirio et al. [83] later proposed another HDS technique through which users control the virtual viewpoint by manipulating virtual reins attached to a virtual bird. Activation, deactivation, translation rate control, and steering are achieved via 2-handed gestures as follows. Moving the arms up and down activates locomotion while
crossing the arms deactivates it. Acceleration and deceleration are achieved by moving the arms forwards and backwards, respectively. Turning to the left or to the right is achieved by moving away the corresponding arm to a distance that is mapped to the desired turning rate. NuNav3D [222] is a HDS VLT that uses the offset between two recognized body poses to calculate the amount of rotation and translation updates based on offset vectors of the right and left hands, respectively. LMTravel [70] uses recognized hand gestures to determine viewpoint control activation, translation, and rotation. Viewpoint control is activated or deactivated by opening or closing both hands, respectively. Translation rate is controlled by the number of fingers stretched while rotation is controlled by the tilt angle of the right hand. Zhang et al. [346] similarly developed a set of two-handed gestures for VR locomotion. Steering left or right is done by the orienting the right thumb to the left or right, respectively. Translation, on the other hand, is achieved by pointing the left palm upwards or downwards to move forward or backward, respectively. Ferracani et al. [102] proposed two one-handed steering techniques: one that controls the direction of travel according to that the user’s index finger, and another that enables steering with their fist.

Several evaluations compared between HDS and GDS. Bowman et al. [54] conducted two experiments to assess participants’ performance in terms of time to reach the target given the VLT (HDS or GDS). No significant difference among the two techniques was found when participants were asked to move directly towards an object while HDS was found superior when they were asked to move relative to an object. The latter result, according to the authors, was due to the coupling of the user’s gaze with steering control. In a target following task, Suma et al. [290] compared between GDS and HDS techniques among other VLTs when task complexity and type (single navigation or divided task) were varied. The two techniques were
found comparable in terms of performance and VR sickness. Suma et al. [285] also compared between the two techniques in a multi-level 3D maze exploration task to measure participants’ performance, cognitive abilities and VR sickness. While the two techniques were comparable with respect to cognitive abilities and VR sickness, GDS was superior to HDS in terms of completion time and collisions. Contrary results were reported by Christou et al. [80] when participants were asked to perform a wayfinding task, where HDS had significantly better performance than gaze in terms of task completion time and errors made. In another wayfinding task by Christou et al. [81], the two techniques were found comparable in terms of VR sickness, disorientation, and completion time while HDS resulted in significantly less success rates than GDS.

**Lean-directed** steering techniques varied in terms of the body parts involved and the choice of sensing platform. Among the earliest LDS techniques were the ones proposed by Fairchild et al. [97], where two LDS techniques were proposed: one that transforms head translations from a defined central point (e.g., waist) to scaled virtual translations; and another that moves the virtual viewpoint in the direction of body leaning in a speed that corresponds to the amount of leaning. The latter technique was refined by LaViola et al. [173] to reduce the chance of classifying resting postures as leaning. DeHaan et al. [91] proposed a LDS VLT using the Wii Balance Board. Leaning forward or backward causes a virtual forward or backward translation, respectively, while leaning on either side causes strafing. To turn, users press with the toe of one foot and the heel of the other. The rate of translation corresponds to the amount of leaning. A similar implementation is that of the Human Transporter [311], except that strafing is not offered and turning is activated by leaning on one of the Wii Balance Board sides. Wang and Lindemann [317] implemented a leaning interface that resembles the surfing metaphor using a
Wii Balance Board for steering control and an accelerometer on one arm to control speed of travel. A later implementation [316] offered vertical travel using multi-touch gestures with a touchpad on the user’s leg. Carrozzino et al. [71] developed a pressure mat to offer LDS VLT whose leaning gestures are mapped similarly to those of the Human Transporter [311]. Leaning for translation and rotation was among the locomotion techniques implemented by Guy et al. [120] using depth sensing, where leaning to the right or left rotates the virtual viewpoint clockwise or anticlockwise, respectively. Translation was enabled either by body leaning or a partial gait gesture. Zielasko et al. [349] designed a LDS technique for seated virtual experiences, where leaning forwards or backwards translates the virtual viewpoint in the corresponding direction. The distance of the user’s head from a predefined zone is mapped to translation speed while steering is determined based on the yaw angle of the HMD. LDS was also implemented using rotating chairs. The ChairIO [34] and NaviChair [159] appropriate a swivel stool chair to offer a leaning interface that supports moving forward, backward, upwards, downwards, sideways as well as turning. Three different chair-based leaning interfaces were compared to GDS and joystick control in a recent study that involved a search task [158]. All techniques were comparable in terms of sense of vection, spatial perception, enjoyment, engagement, presence, and physical exertion. Joystick control was found better than the rest of techniques in terms of spatial orientation, accuracy, ease-of-use, controllability, and comfort. LDS interfaces were also offered on mobile VR. Tregillus et al. [305] developed an omnidirectional leaning VLT for mobile VR via head tilt. With this interface, users could control travel direction and speed according to the direction and degree of their head tilt, respectively.

Very few research efforts have explored *torso-directed steering*, probably because it requires more sensors, compared to HDS and GDS techniques that utilize
the already available sensors at most VR tracking systems [174]. TDS through tracking shoulder or hip rotation are among the recent TDS implementations by Guy et al. [120]. Bowman and Hodges [55] compared between TDS, HDS, and GDS techniques to examine their effect on cognitive load in an information gathering task while navigating a maze. All steering techniques were similar with respect to time and collisions. Similar results were obtained by Suma et al. [290]. Bowman et al. [53] compared TDS with two other steering techniques (HDS and GDS), two manipulation-based techniques (HOMER [53] and Go-Go [234]), two target selection techniques (map dragging [53] and teleportation). Participants were asked to take part in naive and primed search tasks to find targets distributed over an open space. Steering techniques were generally better than all other techniques while TDS was generally worse than the other two steering techniques in terms of both thinking and travel times. Thinking time was the time elapsed from the start of the task until the beginning of movement, while travel time was the time elapsed from the beginning of movement until the reaching the target. In a path following task [120] that involved a secondary action (e.g., holding a mug), TDS caused less physical exertion compared to LDS, but was more disruptive to secondary actions.

**Physical Steering techniques**

These techniques utilize physical props to offer steering. The advent of consumer VR technology has made physical steering devices limited to specific scenarios such as the need to mimic the experience of a real-world vehicle. Examples of physical steering devices include the bike locomotion platforms made for the Olympic bicycle race and the Border Collie virtual environments [59], the Sarcos Uniport [130], as well as VR simulators for aircrafts, merchant ships, cars, boats,
and spaceships [61]. For further discussion on physical steering techniques, readers may refer to these sources: [174, 130, 61].

2.2.3 Selection-Based

Also called automated VLTs [145, 189], these techniques liberate the user’s mind from thinking about how to get somewhere and allow users focus on where to get to. Two known techniques in this category are target selection and route planning [174].

In target selection, the user selects a target destination in the VE after which the virtual viewpoint is moved to the target, which can be either a selected position [54] or scene object [53]. A widely known target selection technique is teleportation [260], which either moves the virtual viewpoint to the destination instantly [58]; or gradually at a variable [182] or fixed [35] speed. Map dragging [53] is another target selection technique through which users drag an icon that represents their current location to a new location on a miniature 2D map of the VE. While the majority of target selection techniques are activated with a controller, some implementations offer target selection control using body gestures. The Step WIM [173] is a target selection technique that enables users to move to a new destination by stepping over its corresponding location in a World-In-Miniature (WIM) projected on the floor of a CAVE. Jumper [41] allows the user to be teleported to a destination by jumping in place while looking towards the intended destination.

Several studies were conducted to examine the effectiveness of teleportation in different virtual tasks. Bowman [54] studied the effect of teleportation speed on the participants’ spatial understanding. Four teleportation speed profiles were used:
slow constant speed, fast constant speed, slow-in/slow-out, and instant teleportation. Participants were familiarized with the VE that consisted of colored boxes labeled with a given letter. Participants then moved to multiple locations in the VE and were asked after each move to locate a certain box and indicate its label. Instant teleportation was found to have the worst performance among the evaluated techniques in terms of task completion time, suggesting that instant teleportation was the most disorienting technique to the participants. Bakker et al. [32] compared between automated continuous locomotion and teleportation in a spatial memory task. Participants were periodically asked to point to previously seen objects while freely exploring the VE using one of the VLTs. Significantly more pointing delays were incurred when teleportation was used. Christou et al. [81] compared between teleportation, GDS, and HDS in a primed search task. Participants were asked to find their way to a previously seen target within a time limit and make frequent stops on the side of the way to collect tokens. Teleportation was found to be significantly the least VR sickness inducing in terms of nausea, oculomotor discomfort and disorientation. It also resulted in significantly greater task completion success rates than HDS, but comparable to that of GDS. Participants were able to complete the task in significantly less time when teleportation was used. Teleportation, however, resulted in the smallest number of collected tokens, indicating its tendency to miss spatial details.

**Route-planning** VLTs offer more granular control over virtual travel than target selection techniques by providing the user with means to select the route to be navigated from the source to the destination. Few attempts have been made to implement this technique, one of which is proposed by Bowman et al. [51].
2.2.4 Manipulation-Based

Manipulation-based VLTs generally work by manipulating the user’s position, orientation, or scale using gestures that either control the virtual viewpoint or the virtual world [174]. Ware and Osborne [322] proposed two manipulation-based VR locomotion metaphors. With the Eyeball-in-hand interaction metaphor, the position and orientation of the virtual viewpoint are updated according to changes in the position and orientation of the user’s hand; while Scene-in-hand offers control of the virtual world through the manipulation of a physical prop. A qualitative study compared between the Eyeball-in-hand, the Scene-in-hand, and flying metaphors in the context of an exploration and movie making tasks. No notable differences were found between the three metaphors in terms of ease of control, ease of movie making, and ease of exploration. As suggested in [174], existing manipulation techniques can be used to control the virtual viewpoint by "grabbing the air" hand gestures. Examples are the Go-Go manipulation technique [234] that enables manipulation of remote objects by extending the user’s virtual arm to unrealistic distances, and the HOMER manipulation technique [52] that similarly allows for the manipulation of remote objects when selected by ray casting. A study by Bowman et al. [53] compared between Go-Go and HOMER among other locomotion techniques. Go-Go was found significantly faster than HOMER in a naive search task while the two techniques yielded comparable search times when search was primed. The same study reported that Go-Go caused dizziness, nausea and arm-strain in some users. Manipulation-based techniques also performed generally worse than steering techniques in that study.

Several manipulation-based VLTs also utilize the WIM metaphor, where the virtual viewpoint is updated as a result of manipulating a representation of the
user’s virtual avatar in a 3D map of the VE [283, 252, 311, 335]. Stoakley et al. [283] conducted a qualitative evaluation of using a WIM metaphor to model a virtual office space that involved using the WIM for locomotion and object manipulation. It was observed that updating the virtual viewpoint using the WIM caused disorientation. Valkov et al. [311] explored the value of using a WIM metaphor in augmenting the capabilities of their leaning interface. Manipulation of the WIM had no effect on the primary viewpoint and were only used to assist subjects in self-orientation and wayfinding. In the study, participants were asked to explore a virtual city with and without and WIM. Using the WIM resulted in significantly better results in terms of ease of self-orientation and wayfinding. Using the WIM did not show a notable value, however, in terms of ease of navigation, locomotion speed, precision, intuitiveness, learnability, and fatigue. Wingrave et al. [335] studied the effect of the ability to change the scale of the WIM on ease of use and spatial performance. The improved WIM design (Scaled Scrolling WIM) was compared against a fixed-scale WIM design (standard WIM) in a spatial task that involved looking for and traveling to a blue sphere. No significant difference was found between the two techniques in terms of ease of use and trial completion time. The Scaled Scrolling technique, however, resulted in significantly higher accuracy than the standard WIM.

2.2.5 Multiscale Virtual Locomotion

Aside from grounded virtual navigation, there are use cases for which vertical navigation (e.g., flying) or multiscale exploration of the VE is required. Multiscale VLTs are designed for these purposes. Depending on how the scale is controlled, these techniques can be classified as either active or automatic scaling techniques.
In active scaling techniques, scale is manipulated using controllers [68, 335], hand [185, 197] or foot [173] gestures. Automatic scaling techniques aim to relieve the user from the task of controlling the scale to focus on the task at hand. Arge-laguet et al. [25] classifies automatic scaling techniques according to the granularity of scaling to either discrete or continuous.

In discrete scaling techniques, the user moves into and out of hierarchical levels of scale. Kopper et al. [162] proposed a discrete auto scaling VLT for multiscale VEs. Users can explore different levels of scale using a virtual magnifying glass and move into or out of a new level of scale either instantly using teleportation or gradually by flying. Bacim et al. [31] later improved the aforementioned technique by adding wayfinding aids such as marked maps, WIMs, and hierarchical representation of the multiscale environment.

Scale is manipulated gradually in continuous scaling techniques. To ensure usability and comfort during continuous scaling, several research efforts have been made to automatically adjust other parameters such as travel speed and stereo parameters [25]. Most of the speed adjustment implementations modulate the speed depending on the distance between the virtual viewpoint and the virtual surroundings [321, 187, 308]. Other implementations also consider other criteria such as optical flow [24, 25] as well as informativeness of [107] and degree of interest in [198] the current viewpoint. Depth-based information has also been used to adjust stereo parameters (e.g., inter-pupillary distance) [72] and resolve stereo fusion issues that result from scale adjustment [78].
2.3 VR Locomotion Taxonomies

The majority of research work that aim to examine VR locomotion at a high level has been in the form of developing meaningful taxonomies. Taxonomies that decompose VLTs to a set of design components, similar to that of Bowman et al. [54], can serve as design palettes from which developers and interaction designers could assemble new techniques. Taxonomies that assign VLTs to clusters of techniques that share common characteristics, similar to that of Suma et al. [287], can be helpful to researchers in at least two ways. First, experiment designs of techniques that share the same category can be reused to evaluate a new VLT that belongs to the same category. This not only saves time and effort in coming up with a new study design, but also helps by making results comparable. Second, categorization of VLTs allows for "macro" comparisons between categories of VLTs. When such comparison is made at this high level, common strengths and weaknesses among categories of VLTs can be found. Existing taxonomies are either targeted towards classifying all VLTs in general or just one family of techniques in particular.

2.3.1 General Taxonomies

Mine [196] decomposed VLTs according to their two fundamental components: direction and speed. Direction control was subdivided into: HDS, GDS, physical control, virtual control, object-driven, and goal-driven. Speed control, on the other hand, was categorized to: constant speed, constant acceleration, hand-controlled, physically-controlled, and virtually-controlled. Bowman et al. [54] proposed a similar decomposition criteria. Along with the direction and speed controls, they added input conditions as a third decomposition component to describe how in-
put to start and stop virtual travel could be supplied. Arns [26] similarly proposed a taxonomy that decomposes VLTs according to direction and speed control while also incorporating system factors such as the type of VR display and interaction device as additional decomposition components. Bowman et al. [51] proposed another taxonomy of travel techniques based on the amount of control that the user has over starting/stopping travel, controlling position, and controlling orientation. Aside from decomposing VLTs based on their design components, Nilsson et al. [216] proposed a classification that assumes the dimensions of metaphor plausibility, virtual movement source, and user mobility. Adopted from Slater and Usoh’s VR interaction techniques classification [268], metaphor plausibility classifies travel techniques into mundane (i.e., uses a realistic metaphor) or magical (i.e., uses unrealistic metaphor). The second dimension, virtual movement source, can either be body-centric or vehicular. Similar to Wendt’s classification of walking techniques [326], the dimension of user mobility classifies travel techniques into ones that require physical translation (mobile) or ones that make the user stationary. Virtual travel techniques were also classified into two broad categories: active and passive. Some researchers [130, 171] consider a travel interface as active if it closely resembles how humans walk by exerting a repetitive limb motion either using their legs or arms while others [174] classify a technique as active when it requires a body-driven motion irrespective of its pattern. Another common taxonomy [174, 145] is the one we used to organize the overview of VLTs in Section 2.2 which classifies VLTs as physical, steering, selection-based, and manipulation-based.
2.3.2 Specific Taxonomies

Another set of taxonomies were targeted to classify the VLTs that belong to a certain virtual locomotion family. Steinicke et al. [280] introduced a categorization of RDW techniques according to the type of gain applied to the virtual viewpoint in order to keep the user confined within the boundaries of the physical tracking space. Three types of gains were listed: translation, rotation, and curvature. A more generalized taxonomy for redirection techniques was proposed by Suma et al. [287] in which redirection techniques were classified into three dimensions: technique used (reorientation vs. repositioning), continuity (discrete vs. continuous), and imperceptibility (overt vs. covert). Wendt [326] proposed a taxonomy for walking-based VLTs, represented as a decision tree. Techniques were distinguished according to design decisions such as whether or not mobility is required, body parts used to activate virtual locomotion, and the control mechanism used to implement the technique. A different specific taxonomy [204] aimed to decompose walking-based VLTs into six components: movement range, walking surface, transfer function, user support, walking movement style, and input properties sensed.

2.4 Discussion

The majority of the surveyed studies demonstrated the superiority of real walking over virtual locomotion techniques in search [254], exploration [344], and path traversal tasks [255, 205] in terms of spatial understanding [73, 344], navigation performance [254, 255, 205], VR sickness [73], presence [310, 344], and usability [310, 205]. Such superiority is mostly due to the translational and rotational body
inputs that RW provides [73, 254]. Generally, RW was found mostly applicable in tasks that require problem solving, spatial understanding [344], or obstacles avoidance in complex VEs [254] where relatively little training is required [255, 344]. RW is also desirable in virtual experiences that call for a high degree of realism such as training [302]. RW, however, is not the silver bullet of VR locomotion. For it be enabled, RW requires a tracking space that is void of obstacles as well as a positional tracking system. These are costs whose benefits are not always justified [286]. While RW has high interaction fidelity [205], it is still not suitable to support some virtual scenarios such as flying or even running.

**Redirection techniques**’ ultimate goal is to enable unlimited real walking in virtual environments in order to both reap the benefits of RW and improve the safety of virtual navigation. Redirecting users imperceptibly has been the most attractive alternative as it aims to make users walk infinitely in VR without the need to break their experience when reaching the boundaries of the physical space. With a required tracking space that varies between 12 meters [118] and 44 meters [277] in width, such techniques have been difficult to evaluate with live user testing [128] and to use in a typical home setting. It is possible to achieve redirection with perceptible gains, but such gains have showed a negative effect on walking patterns and cognitive performance [63]. RDW with distractors has also been shown to be an effective alternative for imperceptible redirection [225]. Users, however, can choose to ignore the distractor, causing the redirection to fail [75]. The distractor should also be well-blended with the narrative of the virtual experience for it be perceived as natural [224]. Imperceptible redirection through the manipulation of the VE’s structure was also shown to be feasible [289, 293, 314] though it is limited to structured VEs. While redirection using this technique enables exploration of an infinite number of virtual spaces, it does not allow walking in a single vir-
tual space that is larger than the tracking space. Resetting techniques have been proposed to mitigate the potential failure of redirection techniques. A common criticism of these techniques is the interruption of the virtual experience that may break presence. Such limitation motivated the design of active resetting techniques such as portals [106] and cell-based redirection [342].

Partial gait techniques such as WIP offer another alternative for infinite bipedal locomotion in VR especially when tracking space [329, 226] or computing power [306] (e.g., on mobile VR) are limited. Although it only covers parts of the human gait cycle, WIP has shown its effectiveness from a multitude of perspectives. Due to the proprioceptive and vestibular cues provided by the stepping gestures [270], WIP was as effective as RW in assisting participants to maintain their orientation while navigating VEs [329, 188]. These gestures also helped participants to have better sense of presence compared to virtual locomotion techniques [303, 229, 202] while some studies reported presence scores that are comparable to RW [310]. WIP was also rated as more usable than virtual techniques [306]. This can be due to the fact that WIP is hands-free, requiring no switching between interaction modes, compared to other hand-based VLTs [267]; and has better resemblance of how we walk, even in terms of the level of body exertion [213]. Despite these benefits, the design and implementation of effective WIP gestures faces a number of challenges. Marching WIP gestures often have to be exaggerated [302], making them perceived as strenuous [212]. This also has made it difficult for participants to differentiate between the physical states of walking and running [327]. Such issue motivated the design of less strenuous WIP gestures such as tapping in place [212], ones that capitalize on the by-products of walking [303], or ones that use entirely different body parts to resemble WIP [154]. While WIP, in concept, requires very limited physical space, the issue of unintentional positional
drift, especially associated with marching WIP gestures [213], still requires a sufficient space that is void of obstacles. Otherwise, alternative WIP gestures such as tapping or wiping [212] should be used instead. Platform-based implementations of WIP were proposed to satisfy a special target experience such as uneven terrain simulation [47]. Aside from the cost and instrumentation that such solutions require [267], platform-based implementations pose the risk of falling off the platform during locomotion [329]. Recent WIP implementations often capitalize on the sensors already available at the HMD, but they are criticized because they couple steering to head direction [306]. Aside from solving this issue with the addition of extra sensors [302, 212], partial decoupling solutions were proposed without any extra sensors [229]. Other solutions also proposed to complement WIP with LDS interfaces [305] to achieve decoupling. While such coupling did hinder participants’ navigation performance in some studies [54], other studies found that WIP with head steering resulted in better spatial orientation compared to that with the torso [330]. Control precision of current WIP implementations is one of the most contemporary challenges to the applicability of WIP. This issue likely stems from limitations of the step detection algorithms such as false positives and negatives [267, 310], which usually manifest as latency in starting and stopping of locomotion [98]. Due to this, the use of WIP in narrow spaces (e.g., mazes) has been challenging [226]. The control limitations of WIP has also made it difficult to navigate curved paths [306, 304].

**Gait negation** techniques keep users within the confines of the physical space by canceling their displacement [144] while experiencing the full gait cycle. Such techniques depend on dedicated platforms that often offer rich locomotion experiences that go beyond walking such as the simulation of inertial forces [131, 273], uneven terrains [143] and surface textures [39]. Some of these platforms also sup-
port decoupling of head direction from steering [89], allowing for more realistic locomotion. Their dependency on a platform, however, has introduced a number of challenges that can limit their adoption and applicability. Users can lose balance and fall off these platforms, especially when adjusting their orientation [199]. A harness can be used to mitigate this risk [136], but it can restrict users’ movement and affect the naturalism of locomotion [140, 273]. The range of speeds and accelerations supported by most of these platforms is limited [139, 261], especially for step-based devices [142, 39], making it impossible to support virtual scenarios that require running. Some of these platforms are very unfamiliar, which has affected their performance [205], and required more time for training [296] and balance adaptation [192]. Finally, acquiring such platforms is costly, affecting their chances of adoption by the masses.

Spatial steering techniques offer continuous steering body input provided by the head, hands, torso or leaning. Their implementation often depends on either the already available sensors of the HMD or cheap sensing platforms, making them affordable for the masses. Because users are stationary while using these techniques, unlimited virtual distances can be covered with limited physical space and low exertion [346]. Spatial steering techniques yielded the shortest completion time in naive and primed search tasks compared to selection and manipulation-based techniques, making them good candidates when this measure is the most important [53]. Aside from a few implementations [317, 91], many spatial steering techniques require upper-body input, making them candidates for seated virtual experiences and people with motor impairments. Unlike walking-based techniques, spatial steering techniques, especially HDS, make it possible to offer vertical locomotion, usually designed using the flying metaphor. Although such techniques are usually labeled as virtual [285], they have shown their potential to
mimic ecological forms of locomotion other than natural walking such as skiing [122] and surfing [317]. The suitability of these techniques, however, is affected by their design characteristics. GDS couples the viewpoint with steering direction, making users unable to look around while moving in a certain direction. Such coupling can make GDS easier to learn and effective to some degree in basic navigation tasks such as path integration [254], but it can also impede effective information gathering during exploration [54, 55]. GDS uses the head to provide steering input. This may have the benefits of better steering control, which could be why GDS had better accuracy than HDS [54]. It may also provide synchronized vestibular input with yaw optical flow, which may reduce the incidence of VR sickness [171]. This overloading of the head, however, may result in excess head rotations, leading to discomfort [53]. On the contrary, HDS decouples steering from viewing by delegating steering to the hands. While decoupling was useful in search tasks [53], the use of the hands for steering has introduced a number of challenges such as fatigue resulting from prolonged use [70] and hand overloading with both locomotion and object manipulation, which requires learning how to switch between the two modes [97]. TDS offers a hands-free spatial steering interface that also decouples viewing from steering at the expense of adding an extra sensor [290]. LDS also has similar benefits to TDS with the caveats that leaning interfaces were not found optimal for precise locomotion scenarios [318] and that chair interfaces had usability issues that were linked to their unfamiliar design [159, 158].

**Physical steering** techniques help users steer with a physical prop that often aims to mimic the steering experience of a real-world vehicle. The haptic feedback as well as the natural mapping of the physical prop to its real-world counterpart makes such techniques suitable in cases when realistic vehicular locomotion is desired. A challenge to these techniques is the potential mismatch between physical
realism of the steering prop and the realism of feedback forces that it provides, which may negatively affect the user experience [174]. Such a mismatch can make the performance of a technique worse than both virtual and natural locomotion techniques, placing it at the uncanny valley of interaction performance [190].

**Selection-based** techniques offer minimal involvement of the user in the process of locomotion. Discrete target selection techniques, such as teleportation, allow users to cover great virtual distances quickly with minimal physical effort [32]. The selection nature of these techniques poses a challenge in virtual scenarios when precise locomotion is required [53]. Due to their discrete nature, these techniques are most suitable for primed search tasks at which users have a particular target in mind [53] and have prior knowledge about the destination [32]. Their discreteness also makes them inefficient in naive search and exploration tasks [53]. This is in part due to the potential loss of information along the path between the source and the destination [81]. The lack of any optical flow is one of the key reasons that made teleportation popular in the VR industry as it helps reducing the incidence of VR sickness [81]. This instant jumping, however, is the cause of disorientation, which has been the most critical issue of teleportation [54, 41, 32]. The effects of disorientation can be reduced by introducing an accelerated transition between the source and destination [182, 41, 35], but speed of the transition has to be slow enough to enable spatial awareness and fast enough so that the duration of visual-vestibular conflict is reduced [174]. Disorientation can also be reduced by spatial familiarization of the destination beforehand [32]. As a low-exertion locomotion technique, teleportation may discourage users from using real walking, if available, which may lead to low presence [36]. Techniques that enable teleportation with body input [41] are, therefore, encouraged if more body engagement is desired. Most of the current target selection techniques require an input device,
which may affect the overall performance of users in scenarios with demanding object manipulation. To mitigate this limitation, hands-free target selection techniques can be used instead [173, 41]. In tasks that require intensive cognitive load, target selection techniques that depend on maps should be avoided as they tend to have a negative effect on cognitive load [53, 51].

*Manipulation-based* techniques offer virtual locomotion either by manipulating the virtual viewpoint or by manipulating the virtual world. This has been realized either with hand gestures that are often borrowed from existing object manipulation techniques [234, 53] or with virtual avatar manipulation techniques on a WIM [283]. Hand gesture techniques are suitable in scenarios that demand heavy object manipulation, at which users can reuse the same technique for both object manipulation and locomotion. This, however, is at the expense of having to switch regularly between the two modes [174]. Due to the great physical effort that they often demand, such techniques are not suitable to travel long virtual distances [53]. Unlike grounded locomotion that mimics how we navigate the real world, locomotion using WIMs offers a larger spatial context of VE that can be viewed from multiple perspectives and at various scales [283]. Similar to the issue with teleportation, moving the virtual viewpoint instantly when the user’s avatar is placed at a new location in the WIM can be disorienting. This can be mitigated by smoothly moving the viewpoint as it is transitioned from the source to the destination [283]. When the VE has structured paths (e.g., a virtual city), the viewpoint can be transitioned over a path that is constrained by the structure of the VE to provide better spatial awareness [252]. Varying the scale of the WIM is another issue that needs to be dealt with to improve the effectiveness of such technique. This can be addressed either by introducing WIM scrolling interfaces [335] or by using a dedicated scaling gesture [173].
Multiscale VLTs are suitable for VEs with details that cannot be shown all at once either due to their complexity or due to their hierarchical nature [162]. Techniques that move the virtual viewpoint through discrete levels of scale carry the concern associated with target selection techniques in that they should keep the user spatially oriented throughout the locomotion experience, which has been addressed by adding wayfinding aids [31]. Discrete techniques should also make objects with nested scales discoverable using appropriate metaphors (e.g., a magnifying glass [162]). Along with the automatic adjustment of scale, auto scaling techniques should also modulate the navigation speed and the stereo visual parameters to avoid inducing VR sickness [24] or eye discomfort resulting from vergence-accommodation conflict [321]. To improve the navigation performance, auto scaling techniques should also consider collision avoidance as a factor when the aforementioned parameters are automatically modulated [187].
CHAPTER 3
PAWDIO: HAND INPUT FOR MOBILE VR USING ACOUSTIC SENSING

3.1 Introduction

Smartphone VR adapters, such as Google Cardboard, have a potential to bring VR to the masses as they can turn the now-ubiquitous smartphone into a head-mounted VR display at a low cost. A criticism of current mobile VR apps is that they only deliver simple “look-and-see” or “rollercoaster”-like experiences [33] as the input options of mobile VR adapters are limited [168, 306]. Google Cardboard requires users to hold the adapter with both hands to limit the head rotation speed to the torso, which helps mitigate simulation sickness, [7], but limits the use of a controller. Google Cardboard features a single button that is either activated using magnetic sensing [271] or through an internal mechanism that generates a touchscreen input. Due to these constraints, mobile VR apps largely rely on gaze input -facilitated using the smartphone’s inertial sensors. This leads to using GUI elements, such as gaze or stare buttons, whose usage has been found to be detrimental to immersion [115] when compared to conventional input.

The hand is the most natural input device for direct manipulation in VR [56], and is considered more immersive than using a controller. Hand input, however, has been difficult to implement on mobile VR platforms. Marker-less hand tracking is feasible on a smartphone [2] but its accuracy is subject to background and illumination changes, something that occurs frequently as VR users often turn their head to look around. Smartphone cameras have a small field of view; limiting hand-tracking to a small area, which may be detrimental to immersion. Robust hand-tracking usually requires an external depth sensor (i.e., Leap Motion) and
Figure 3.1: PAWdio appropriates a pair of ordinary in-ear headphones to track the position of the hand. The user holds a single earbud in their hand that produces an inaudible tone. Doppler shifts are used to determine the velocity of the earbud from or towards the phone, which is then used to manipulate the Z-position of a virtual hand that is attached to the user’s gaze pointer. Because this sensor is powered by the smartphone it reduces battery life significantly. Mobile computer vision apps are generally computationally intensive [85], which may create lag or degrade the frame rate of the VR simulation that runs on the same smartphone. This is undesirable, as a high frame rate is required to maintain immersion and minimize simulation sickness [193].

This paper addresses a need for low-cost, immersive forms of input on mobile VR by presenting PAWdio; a one DOF hand tracking technique that uses acoustic sensing. PAWdio is low cost as it can be implemented using an ordinary pair of in-ear headphones. It doesn’t require any training or calibration while having a low computational overhead that assures a high frame rate. PAWdio was found to be just as efficient and accurate as the standard single button input for selecting targets but was found considerably more immersive. PAWdio enables immersive implementations of various game related actions, such as; grabbing, punching, pushing or thrusting and throwing.
3.2 Background

Acoustic sensing uses properties of sound such as the Doppler effect to provide user input [242]. Compared with vision based sensing it can be implemented with little computational overhead [121]. Speakers and microphones on current smartphones are capable of producing and recording audio up to 24kHz [176]. The human hearing range lies between 20Hz to 20kHz, which degrades as the result of aging; and the highest frequency a typical adult can hear is 15kHz [282]. Acoustic sensing methods using the [15-21kHz] spectrum can be implemented on any mobile device and are non-obtrusive to adults. Acoustic sensing methods can be distinguished into passive (only recording) and active (emitting and recording). We survey approaches pertaining to gesture detection and localization.

Acoustic gesture sensing uses Doppler shifts, i.e., changes in frequency of soundwaves that occur due to a positional change between an emitter and a receiver. One-handed gestures in 3D space can be recognized using three low-cost stationary ultrasonic receivers and a hand held transmitter [147]. This approach isn’t mobile and relies on custom hardware. SoundWave [119] uses ordinary laptop speakers and microphone to detect various in air hand gestures using observed Doppler shifts and its properties, such as velocity and direction. Dolphin [237] implements this approach on a smartphone. AAMouse [343] localizes a smartphone between two TV speakers that emit different inaudible tones. Doppler shifts and triangulation allows AAMouse to achieve a localization accuracy of 1.4cm using an initial calibration. Phone-to-phone localization with reasonable accuracy (≈10cm) can be implemented using time-difference of-arrival (TDoA) [228] or time of flight (ToF) [238], but these approaches require two mobile devices. Sweepsense [169] detects whether an earbud of a headset has fallen out through the detection of
a non-audible tone that is played through each earbud. Acoustic Ruler [1] is an iPhone app that lets users measure the distance from their smartphone to an earbud of their headset. It uses ToF with an audible tone and with an initial calibration claims a resolution of 1mm.

### 3.3 PAWdio Design

An ideal implementation of hand input for Mobile VR offers 3D input, requires no specialized hardware or prior calibration and offers a low computational overhead to assure a high frame rate with no lag in tracking to maintain a high immersion [193]. PAWdio was designed within these constraints but a compromise had to be made regarding the available degrees of freedom (DOF) of the hand input.

Three dimensional acoustic localization approaches either require a stationary setup [147] or using two devices [238], which is not practical or cost-effective in mobile VR contexts. Instead PAWdio uses more practical and low-cost hardware setup -as introduced by Acoustic Ruler [1]- by appropriating an ordinary set of in-ear headphones to sense the distance of an earbud that the user holds in their hand (see Figure 3.1 for the hardware setup). This setup has two limitations: (1) because we only acquire a distance to the earbud, this only allows 1 DOF hand input, e.g., a virtual hand can only be moved forwards and backward; and (2) with the headset used for hand tracking, users cannot listen to sounds/music which may be detrimental to the VR experience.

Another constraint is that the microphone and speaker need to be directed at each other for best performance. This is challenging for mobile VR as the smartphone is inside the VR adapter and corrugated cardboard has excellent sound
absorption properties. Typically, smartphones feature a pair of microphones for noise cancellation, with a primary microphone located near the mouth and secondary near the ear. When the smartphone is in the adapter, the primary microphone is occluded and pointed towards the user. The location of the second microphone seems to vary (e.g., side, top), but recent smartphones, like the Nexus smartphones, have the second microphone on the back close to the location of the camera. Because most smartphone VR adapters feature a cutout for the camera for AR applications, this conveniently leaves the second microphone unobstructed and aimed towards the hand.

Acoustic localization can be achieved using: TDoA [228], ToF[238] or Doppler shifts [343]. ToF measures the exact time it takes for a sound wave to travel from the emitter to the receiver and then—using the known speed of sound in the air—calculates the absolute distance between them. The use of ToF in a non-real-time operating system, such as Android or iOS, results in large estimation errors [228] as their non-determinism makes it difficult to exactly measure the time when a sound is actually sent or received. Peng et al [228] identified for a smartphone a lower-bound delay of 2.5 ms, which yields an estimation error of $\approx 85$ cm. Because PAWdio requires estimation distances up to an arm length ($\approx 65$ cm) using ToF is not feasible. Acoustic Ruler [1] also uses ToF but circumvents timing limitations using a multi-step calibration process, which we deem too cumbersome for PAWdio, as users may want to quickly engage in VR. TDoA does not require absolute timing, but it requires at least two non-coplanar microphones to achieve one-dimensional ranging, which would require a second device. Doppler shifts are the most feasible method for PAWdio as it doesn’t rely on precise timing or calibration.
PAWdio emits a pure sine wave with a base frequency \( f_0 \) in the inaudible spectrum [15-21kHz] over the earbud used for hand tracking. Besides being non-obtrusive to adults, this frequency doesn’t seem to be susceptible to environmental interference given the ubiquitous application of ultrasonic sensors, such as automatic door openers [169]. The generated wave is recorded with the smartphone’s secondary microphone and for analysis in real-time it is divided into chunks of size \( B \). The size of \( B \) determines the resolution of the frequency estimation and it dictates the latency of the hand tracking. A larger value for \( B \) allows for tracking finer motions of the hand but also increases latency. After an audio chunk is captured, a fast-Fourier transform (FFT) is used to convert it from the time domain to the frequency domain. We then extract a frequency spectrum subdomain around the base frequency \( f_0 \) of size \( W \). Within \( W \), we find the frequency component \( f_{\text{peak}} \) with the highest amplitude. Comparison of \( f_{\text{peak}} \) and \( f_0 \) provides an estimate of the amount of Doppler shift, which allows for estimating the velocity of the hand \( (v) \) as follows:

\[
v = \frac{c(f_0 - f_{\text{peak}})}{f_{\text{peak}}}
\]

(3.1)

where \( c \) is the speed of sound in air at room temperature. By calculating the time difference between two consecutive sound chunks, \( t_s \), along with the calculated velocity \( (v) \), we can estimate the relative displacement \( (D_p) \) of the earbud from or towards the phone using: \( D_p = vt_s \). Because the physical and virtual worlds are at different scales, we translate the physical displacement \( D_p \) to a virtual displacement \( D_v \) using a scaling factor \( \alpha \). The virtual hand’s X and Y locations are coupled to the user’s gaze pointer and its Z location is bounded by a maximum/minimum distance. Sampled changes in \( v \) can be associated with scaled displacement values \( D_v \) that determine the virtual hand’s Z location at each frame.
3.4 Evaluation

Comparing PAWdio to vision based methods is difficult as they offer 3D input; they require additional sensors; and they don’t offer acceptable frame rates on mobile devices. We therefore compare PAWdio to single button input that is available on mobile VR platforms, as this is most similar in terms of available hardware. In our study, subjects perform a 3D target selection task by combining PAWdio with gaze input. Fitts’ law analysis [272] has been performed for related 3D selection studies [161, 301]. Because PAWdio uses gaze & hand selection, a detailed Fitts’ analysis does not generate significant results.

3.4.1 Instrumentation

We used a Cardboard V2 adapter (I AM Cardboard), whose button generates touch input that is considered more reliable than magnetic input [271]. We used a Motorola X Pure Edition smartphone with a Qualcomm Snapdragon 808 CPU with an Adreno 418 GPU which can render 3D simulations at a high frame rate. We used Apple’s EarPods for our in-ear headphones. Our game was implemented using Unity 5 and the Google Cardboard for Unity SDK. PAWdio was implemented as a Unity plugin using the TarsosDSP [5] digital sound processing library. Specific to PAWdio’s implementation, $18KH\text{z}$ was selected for $f_0$ and $340m/s$ for $c$. We experimentally found 8,192 to work best for $B$ to accurately track the moderate motions required to grab an apple with no lag. We found that it is difficult for PAWdio to pick up very fast or very slow motions of the hand as this is difficult to detect using Doppler shifts. For the game, we determined a scaling factor of $\alpha = 5.8$, which moved the hand most realistically. In Unity we did not observe a differ-
ence in frame rate between both input options, which demonstrates PAWdio’s low computational overhead.

### 3.4.2 Participants

We recruited 18 participants (9 females, average age 25.2, SD=4.4, 1 left handed) to participate in our user study. None of the subjects self-reported any non-correctable impairments in visual/audio perception or limitations in mobility. Individuals who self-reported to have previously experienced simulator sickness were excluded from participation, as they were at a higher risk of not completing the study. The user study was approved by an IRB. 13 participants had prior experience with VR and 14 had experience with 3D games.

### 3.4.3 Procedure

To evaluate the performance, accuracy and usability of PAWdio with button input, we conducted a within-subject study using a 3D target acquisition task that was implemented as a game. The use of a game is motivated by that this is a major application area for VR and a game context often allows for eliciting optimal performance. The game consists of a virtual hand whose X, Y position is tied to the gaze pointer and its Z position is controlled by the button or PAWdio. A button click moves the hand forward and another click retracts it and we selected a velocity for the hand that felt most realistic. Using PAWdio, the user’s hand motions are directly mapped to forward/backward motions of the virtual hand. Targets, i.e., apples with a fixed size, are generated at a random location on an imaginary
sphere around the participant where the virtual hand’s movement is bounded by the sphere’s boundaries. A gaze pointer is rendered to aid with the X, Y target selection task. The hand grabs the apple when they collide and when the hand is fully retracted, the apple disappears and a new apple is generated at least 60 degrees away from the previous apple. Participants played two sessions of the game each using one of the input methods with a 2-minute break between them. The order in which input method was used first was counterbalanced among participants. To mitigate the effect of the visual search task on overall task performance, we used the same sequence of apples for both sessions. For each session, participants had to collect 10 apples in the training session and 25 in the experiments.

Each participant received a brief tutorial on how the game works with each input method. When using PAWdio, participants were asked to reach out for the apple using their dominant hand like they would in the real world. To avoid picking up unintended motions, we asked users to retract their hand when looking for apples. We also asked users not to extend their arm extremely fast, as such motions were difficult to detect by PAWdio. We measured total task time from the first to the last apple and errors. For the error count, we define an error as any forward movement of the virtual hand that exceeds a given threshold and does not result in grabbing an apple. Both game sessions were video-recorded for analysis. After the trials, basic demographic and qualitative feedback was collected using a questionnaire. The entire session, including training and questionnaire, took about 30 minutes.
3.4.4 Results

A Grubb’s test found no outliers and table 3.1 lists the average results for all users. A paired T-test found that the PAWdio input technique took significantly longer to perform ($t_{17} = -2.64, p < .05$) but no statistically significant difference was found for error rate ($t_{17} = 1.88, p = .076$).

<table>
<thead>
<tr>
<th>Input method</th>
<th>Time (s)</th>
<th>SD</th>
<th>Errors</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Button</td>
<td>130.82</td>
<td>41.6</td>
<td>1.83</td>
<td>2.4</td>
</tr>
<tr>
<td>PAWdio</td>
<td>147.71</td>
<td>55.4</td>
<td>0.83</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Table 3.1: Average time and errors for all users.

<table>
<thead>
<tr>
<th>Input</th>
<th>Efficiency</th>
<th>Errors</th>
<th>Learnability</th>
<th>Likeability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Button</td>
<td>4.50 (1.0)</td>
<td>3.61 (1.5)</td>
<td>4.28 (1.1)</td>
<td>4.11 (1.1)</td>
</tr>
<tr>
<td>PAWdio</td>
<td>4.28 (.6)</td>
<td>3.17 (1.3)</td>
<td>4.28 (.6)</td>
<td>4.0 (.8)</td>
</tr>
</tbody>
</table>

Table 3.2: Likert scores for each usability attribute per technique (SD).

![Figure 3.2: Ranking of input methods based on usability criteria.](image)

Qualitative results. After the trial, questionnaires were used to collect qualitative feedback. First, using a 5-point Likert scale, users evaluated each input
method on 4 usability criteria: efficiency, accuracy, learnability and likability. Table 3.2 lists the results and a paired T-test found no statistical significant difference between input methods for any attribute. To assess VR immersion, we selected 3 questions from the Witmer and Singer presence questionnaire [336] who were most closely related to input and which are evaluated on a scale of 1 to 5. Figure 3.3 lists the questions and results. PAWdio received significantly higher scores ($t_{17} = -2.17(Q1), -5.68(Q2), -2.95(Q3)$, $p < .05(Q1, Q2, Q3)$). We then let subjects rank each input type on the four usability criteria and VR immersion, where subjects could select “No preference” as a 3rd option. Figure 3.2 lists the results, and a $\chi^2$ test found the rankings for efficiency ($p = .027$), and overall immersion ($p = .001$) to be statistically significantly different. Finally, participants (# who said this) provided feedback about each input technique. For button input, participants liked its simplicity (6) and that it was easy to learn (3). Button input felt efficient (5) and was easy on the arms (2). Several participants complained that the button input was too sensitive (4) with users unintentionally activating it when they merely rested their finger on the button. PAWdio felt natural and realistic (11) and it allowed for making corrections half-way (2) something that is not possible with button input. Some users expressed that PAWdio was somewhat strenuous to perform (7) and that it was susceptible to some inaccuracy (5), e.g., the hand moving unintentionally.

### 3.5 Discussion

Button input is faster as the virtual hand moves with a fixed speed when the button is pressed, where PAWdio follows the actual position of the user’s hand, which takes more time. Surprisingly, users felt the button input was more accurate, but re-
Figure 3.3: Evaluation of immersion for both input techniques

Results showed the error rate for button input to be twice that of PAWdio (not significant for $\alpha = 0.05$ but significant for $0.1$). This error was caused by the sensitivity of the button input of the Cardboard adapter we used. Occasionally PAWdio would pick up false positives when the user would turn quickly without moving their arm; which would cause a Doppler shift. Hand input was considered more strenuous than button input, but this is an inherent characteristic of hand input. Though the usability of both input techniques was similar, PAWdio was found to be significantly more immersive, which is important for VR.

3.6 Limitations and Future work

Playing audio is limited when using PAWdio, which is considered essential to immersion. Some earphones may allow for one earbud to be in the ear while the other one can be used for hand tracking; but audio will then be limited to mono. Android allows for playing different audio streams over the earphones or speakers at the same time [4], and this allows for PAWdio to work while playing audio over the speakers. PAWdio uses a frequency that is hearable by children and pets, and $f_0$ should be set to the highest frequency, e.g., 22KHz. Because we did not eval-
uate PAWdio robustness to noise, our results are preliminary. It may be difficult for PAWdio users to activate the button on a VR adapter as they hold an earbud in their hand. We were able to create an alternative input by having users cap the earbud with their thumb. This blocks the ultrasonic tone and its absence/presence can be detected and used for input, i.e., similar to Sweepsense [169]. PAWdio’s one DOF hand input is constrained when compared to 3D hand input; PAWdio only tracks a single hand, and very fast or slow motions of the hand are difficult to detect using Doppler shift. However, using different values of $B$, either faster or slower motions may be detected at the cost of performance. False positives may occur when the user turns quickly, but this is less likely to occur with most VR apps as they typically don’t contain a large visual search task, like the game we used for the evaluation. The $X$, $Y$ position of the virtual hand is tied to the gaze, which may be considered less immersive than 3D hand input. However, Cardboard already requires users to hold the adapter with both hands [7], so requiring the user’s hand to remain in the field of view, is not a major constraint.

Despite these limitations, we demonstrated that PAWdio can offer basic hand input without relying on expensive sensors. PAWdio offers a slightly lower performance, but higher immersion than current VR input options. A number of VR demos were developed that demonstrate how PAWdio can be used to enable various “linear” game actions, e.g., punching, throwing object, thrusting a spear or pool cue, opening a door, and pushing or pulling objects.

For future work, we will research how to enhance PAWdio to enable 3D hand input, which is more immersive than 1D hand input. Many in-ear headsets often feature a microphone. When attaching this microphone to an earbud, three distinct speaker-microphone pairs could be defined and it might be possible to triangulate
the location of the hand using ToF [238].
CHAPTER 4
STEREOTRACK: 180-DEGREE LOW-COST ACOUSTIC POSITIONAL TRACKING FOR MOBILE VR PLATFORMS

4.1 Introduction

With nearly 2 billion smartphones in the world, mobile virtual reality (VR) platforms, such as Google Cardboard or Gear VR are considered to have the largest potential for bringing VR to the masses [92]. VR is considered to offer the highest presence when it closely mimics interaction with the real world [294]. A problem with current mobile VR platforms is that their input options are rather limited [306] which limits developers in their ability to create highly immersive experiences, similar to those available on PC VR platforms (e.g., the Rift or the Vive).

On PC VR platforms, realistic interaction can be implemented using positional tracking –which tracks the absolute position of the headset and controllers within a 3D space. This provides high accuracy and low latency to assure high presence. Positional tracking typically requires installing external cameras or infrared beacons in an environment, which adds to cost and confines the available tracking space. Some approaches [3, 16, 18] implement positional tracking using a regular smartphone camera. Aside from techniques that require installation of fiduciary markers or mapping the space, performing computer vision on a mobile device is generally computationally intensive [85]. This does not only decrease battery life but also may affect the frame rate of the already CPU-intensive VR simulation that also runs on the phone. This is undesirable, as a high frame rate is required to maintain presence and minimize the occurrence of VR sickness [193]. The robustness of such techniques is also affected by challenging illumination conditions...
Figure 4.1: StereoTrack appropriates an ordinary pair of speakers, for example, those available in a TV, to enable 180 degree positional tracking on mobile VR platforms using acoustic sensing. StereoTrack can be used to implement popular waveshooter games on mobile VR platforms.

This paper explores the feasibility of positional tracking on mobile VR platforms using a technique called StereoTrack; an acoustic sensing based positional tracking technique. Acoustic sensing exploits properties of sound to enable user input [242] and offers a significant benefit over computer vision based sensing in that it can be implemented with little computational overhead [121], often using existing hardware. Smartphone speakers and microphones can produce and record audio up to 24kHz [170]. Humans can hear sound frequencies up to 20 kHz [232] and the hearing threshold increases abruptly after 16 kHz [27], leaving a short frequency band that can be appropriated for acoustic sensing which is non-obtrusive to adults and does not require specialized hardware.
4.2 Background

A number of projects attempted to enable positional tracking on mobile VR platforms [16, 10, 3, 18]. These developments indicate that bringing positional tracking to mobile VR platforms has a lot of potential. They, however, require the acquisition of a custom-made hardware; use a computer-vision-based localization approach that is sensitive to illumination conditions and quickly drains the limited smartphone’s resources; or demand a time-consuming setup that is difficult to port.

The use of acoustics has been another attractive alternative to achieve mobile positional tracking. BeepBeep is an active phone-to-phone acoustic localization technique that uses time-difference of-arrival (TDoA) [228]. It uses chirps in the [2-6kHz] range and provides a distance between phones with an accuracy of \(\approx 10\) cm. Qiu et al. [238] extends this approach using time-of-flight (ToF), power level and inertial sensing with a Kalman filter to allow for real-time 3D phone-to-phone localization. AAMouse [343] turns a smartphone into a mouse for a TV. The TV speakers emit tones in the 17-21kHz range. Doppler shifts are used to determine the smartphone’s distance from each speaker and its \((X, Y)\) position is then determined using trilateration. AAMouse achieves a tracking accuracy of 1.4 cm. CAT [184] builds on the work done at AAMouse to provide an acoustic localization technique that combines Doppler shifts, ToF, and inertial sensing to achieve a mm-level 3D position estimation accuracy. EchoTrack [77] uses a combination of ToF and Doppler shifts to estimate the location of the user’s hand with a tracking accuracy of 2 to 3 cm.

A few approaches use acoustic sensing to increase the limited input options
of mobile VR platforms. PAWdio [345] is a 1D hand-input technique for mobile VR. The user holds a single earbud in their hand that generates a sound wave in the inaudible range. Doppler shifts are used to determine the velocity and displacement of the hand and this is used to update the Z-position of a virtual hand that is attached to the user’s gaze pointer. Constellation [104] is a 3D positional tracking system built to support AR/VR applications. The solution assumes a head-mounted-display (HMD) that is equipped with an inertial measurement unit (IMU) and an array of at least three microphones. Ultrasonic signals emitted by beacons installed at a known constellation are captured by the microphone array and ToF is used to measure the distance from the source. Kalman filtering is used to fuse the measured distances with the estimated position from the IMU to determine the user’s 3D location at an accuracy that ranges between 1 to 8 mm. The Intersense I-900 is a 6 DoF positional tracking system designed for VR applications [20]. With combined acoustic and inertial sensing, it reports to have a static accuracy from 2 to 3 mm.

4.3 Design of StereoTrack

A positional tracking system for mobile VR would have the following desirable features. It should offer 3D tracking to track user’s horizontal as well as vertical movements. It should also be able to track the user in a 360° range, irrespective of the direction the user is facing. The user should also be tracked with a low latency to maintain high presence and to minimize VR sickness. With these features in mind, an implementation for mobile VR should be low-cost, CPU-friendly, and require little to no calibration. A determining factor in cost is instrumentation and to keep this low preferably all computation and sensing should take place on
the smartphone (inside-out tracking). StereoTrack strives to satisfy most of these requirements, with the exception of some trade-offs that we discuss in this section.

StereoTrack appropriates a pair of speakers, such as those found in most TVs and personal computers, to provide positional tracking, making the setup both cost effective and ubiquitous. While desirable, such horizontally coplanar setup of common speakers implies that StereoTrack would not support vertical positional tracking. This setup also would force the user to always be facing the speakers for their motions to be tracked. In other words, the microphone has to be no more than 90° away from the speakers for StereoTrack to function properly, which is what we mean by 180° tracking. This type of tracking, i.e., 180°, is also available on the popular Oculus PC VR platform [15], with many games using it.

Positional Tracking using acoustic sensing can be primarily achieved using one of three approaches: Time of Flight (ToF) [238], Time Difference of Arrival (TDoA) [228] or Doppler shifts [343]. ToF measures the absolute distance between an emitter and a receiver by measuring the time it takes for the emitter signal to travel from the emitter to the receiver. Providing a quality estimation using ToF is challenging, especially when implemented on non-realtime operating systems [228] due to the inability to provide precise time synchronization between the emitter and receiver, and a perfect estimation of signal emission and receipt times. TDoA relies instead on the time difference of signal arrival from different emitters situated at known locations, but requires at least three emitters to achieve 2D localization. Because Doppler shifts enable 2D positional tracking with only two emitters [343], we adopted it as the acoustic tracking approach to implement StereoTrack.

Similar to previous work [343, 76], StereoTrack offers positional tracking as follows. A sweep of inaudible tones is emitted by two speakers that are coplanar
and distanced from each other by a known distance $D$. The emitted tones are recorded by the smartphone’s microphone at regular sampling periods of length $B$. The localization algorithm transforms the audio samples from time domain to frequency domain using Fast Fourier Transform (FFT) to estimate the Doppler shifts from each speaker by searching for frequency peaks found in a spectral window of width $W$, based on which the smartphone’s distance from each speaker is updated, assuming an initial distance, as follows [343]:

$$D_{i,k} = D_{i-1,k} + \frac{f_{s,k} - f_{0,k}}{c} t_s$$ (4.1)

where $D_{i,k}$ is the smartphone’s distance from speaker $k$ at time $i$ in m/s, $f_{s,k}$ is the estimated Doppler shift of speaker $k$ in Hz, $f_{0,k}$ is the base frequency of speaker $k$ in Hz, $c$ is the speed of sound in the air in m/s, and $t_s$ is the length of the time step in seconds. Distances from both speakers are then used to update the smartphone’s position on a 2D plane similar to [343, 76]. The estimated position is then passed together with the estimated speed, calculated from the Euclidean distance between the previous and current positions over $t_s$, to the VR engine to move the user’s viewpoint in the virtual environment at the given speed [345]. We choose to record from the secondary microphone as it has a better chance of being directly facing the emitting speakers, which could lead to a better recording quality. Mobile VR devices are required to have a refresh rate of at least 60Hz [17]. At a sampling rate of 44.1kHz, the size of the audio sample buffer $B$, therefore, should be at least 735, which implies a frequency resolution of 30Hz. Several trials showed that this buffer size has an estimation quality that is too poor to be used for VR positional tracking due to noticeable drift and jitter. Empirically, we found a buffer size of 1764 to be satisfactory, which is also similar to the chosen value of $B$ in [343]. This buffer size has a refresh rate of 25Hz and a frequency resolution of 12.5Hz. A position update rate of 25Hz implies that StereoTrack’s position updates are generated
Figure 4.2: Results of the accuracy evaluation of StereoTrack in terms of drift over time, measured as Root-Mean-Square (RMS) error in centimeters per second for the three tracking regions (top, bottom, all)

at a rate that is 35Hz less than that of the mobile VR display refresh rate, resulting in abrupt changes in position as the user translates. We reduce this effect by linearly interpolating between the current and new positions by a certain rate $r_I$ as follows:

$$r_I = \frac{\sqrt{(x_{\text{new}} - x_{\text{current}})^2 + (z_{\text{new}} - z_{\text{current}})^2}}{t_s}$$

(4.2)

where $(x_{\text{current}}, z_{\text{current}})$ and $(x_{\text{new}}, z_{\text{new}})$ are the current and new positions, respectively, while $t_s$ is the length of the time step in seconds. A frequency resolution of 12.5Hz would still cause drift and jitter. We overcome this by applying a weighted average of 5 tones per speaker, with the weights being the inverse of the noise floor, and a Kalman Filter to smooth the signal as proposed in [343]. We also apply a low-pass filter that filters out shifts with a magnitude below 3Hz. For our solution to support walking in VR, the size of the spectral window $W$ has to be at least 138Hz for the average preferred human walking speed of 1.4 m/s [200] to be detectable.

4.4 Evaluation

We evaluate StereoTrack in terms of precision, accuracy, and latency which are among the key factors that characterize the suitability of an interaction technique.
4.4.1 Instrumentation

In our evaluations, we used a Google Pixel smartphone with Qualcomm Snapdragon 821 processor whose built-in camcorder’s microphone was used to record the ultrasonic tones. A total of 10 pure sinusoid sound tones were played from the two speakers of a Samsung K450 Bluetooth sound bar (distance between speakers \( D \approx .76m \)) with 5 signals per speaker. Signals emitted from the right speaker ranged from 17kHz to 17.8kHz while the ones emitted from the left speaker were from 18kHz to 18.8kHz, where each signal was 200Hz apart from the next one. The tones were played at sampling rate of 44.1kHz. An Android program that implements our design was developed that measures the Doppler shifts and the outputs a new position every 40ms. The audio samples buffer size \( B \) was 1764 and the spectral window \( W \) was 175Hz. We used TarsosDSP [265] as the digital signal processing library. We customized the Android app as per the requirements of each trial as we elaborate below.

4.4.2 Precision Analysis

Poor positional tracking in VR manifests as jitter, which was shown to hamper performance in 3D object manipulation tasks [239], seen to have a negative impact on interaction fidelity [191], induce VR sickness and cause break-in presence [145]. We assert the precision of StereoTrack by testing its ability to measure repeatable Doppler shifts in static conditions. We conducted the study in a 3m \( \times \) 4.5m in-
door space, forming an area that exceeds the maximum tracking areas of major high-end VR platforms [134, 22]. We divided the area to 9 grid cells (1m \times 1.5m in size) and performed 3 trials at the middle of each cell. We see such sampling of test locations representative of the tracking area with respect to the quality of the received sound signals. At each trial, the smartphone was placed on a tripod (to ensure the smartphone’s stability) at the middle of the grid cell. The lab door was closed throughout the evaluation to avoid any sudden changes in air pressure and a line of sight was always maintained between the smartphone and the speakers to avoid any interference with sound waves propagation. The Android app was turned on to collect 1,500 Doppler shifts after which it was automatically turned off. A total of 9 locations \times 3 trials \times 1,500 data points/trial resulted in 40,500 data points, where each data point contains two estimates, one for each speaker. The mean Doppler shifts were -.0070Hz (SD = .003) and .0070Hz (SD = .004) for the left and right speakers, respectively. At a room temperature of 25° and a base frequency of 17kHz, these Doppler shifts translate to \approx \pm 0.006mm, which is negligible, indicating our solution’s stability while being stationary.

4.4.3 Accuracy Analysis

StereoTrack’s implementation is based on dead reckoning that is known to incur drift. We evaluate the significance of drift through measuring the rate of accuracy degradation in terms of the Root-Mean-Square (RMS) error over time. We used HTC Vive’s Lighthouse tracking system, with a reported accuracy of \approx 2mm [8] as the ground truth. A Vive controller was used to track the ground truth whose coordinate system was mapped to that of StereoTrack. A PC program running on Unity collected the controller’s X and Z coordinates using SteamVR plugin while
an Android app running StereoTrack was used to record the coordinates of the smartphone. System times of the PC and the smartphone were manually aligned. A server program was hosted at the same PC to allow the Android app to synchronize starting and stopping of data collection among the two systems.

People have different walking profiles and to incorporate the effect of such differences on the tracking quality we conducted this study with 3 participants, where each participant was asked to walk along three defined paths. Diamond-shaped paths were used in this study: a large one that covers the whole tracking space, and two small ones that cover the upper and lower parts of the tracking space, respectively. A diamond shape, as opposed to a rectangular one, forces participants to walk diagonally, making the task more realistic than following right-angled paths. Before each trial, participants were asked to stand on the starting point marked on the ground (see Figure 4.3). Participants held the Vive controller with the smartphone mounted on the controller against their chest. When the participant is ready...
to walk the path, the experimenter triggered data collection on both systems from the Android app after which participants were asked to start following the path. A trial finishes when the participant returns to the starting point, where the experimenter stops data collection on both systems using the Android app. All participants walked along the diamond paths counterclockwise in a walking speed of their preference. Each participant performed 5 trials for each of the 3 paths. Due to the difference between StereoTrack’s and SteamVR’s refresh rates (25Hz vs. 90Hz), we compare pairs of trajectory points (i.e., one point from each tracking system) that belong to the same percentile. For each of the 1,000 pairs of coordinates that we sampled, we calculate the Euclidean distance from which we compute the RMS error. The mean RMS error for each of the three paths is then normalized by its corresponding mean duration. The results are shown in Figure 4.2 and a sample trajectory for each path is shown in Figure 4.3.

4.4.4 Latency Analysis

Low frame rate in VR can cause break-in-presence [193] and induce VR sickness [43]. We investigate the effect of StereoTrack on the end-to-end latency of a mobile VR experience by measuring the frame rate with and without StereoTrack. We developed a Unity app for Google Cardboard that reads position updates from an Android plugin that implements StereoTrack. We conducted two trials: one with the position updates turned on and another with the position updates turned off. At each trial, the Unity application counts the number of frames per seconds (fps) for a total of 1 minute and saves the results to a file for later analysis. Mean frame rates were very similar (with StereoTrack: mean = 59.71, SD = 3.00; without StereoTrack: mean = 59.63, SD = 0.36) and a t-test showed no significant difference
4.5 Example Applications

Two applications were developed to show how StereoTrack could enable simple VR experiences on mobile platforms and overcome some limitations of existing mobile VR locomotion techniques.

4.5.1 Wave Shooter Game

A key interaction element of wave shooter games is the player’s ability to dodge enemies’ lasers by making short and fast moves, preferably in a natural and responsive manner. We developed a simple space shooter game that utilizes StereoTrack to demonstrate its ability to enable the maneuverability required by such experience. The game consists of a spaceship that continuously follows the player as s/he moves in the virtual environment and shoots laser beams at the player at random intervals (see Figure 4.1). The player side-steps to the right or to the left in order to avoid being hit by the laser. The speed of the laser beam was adjusted to give users a reaction time of within one second from the time the laser beam is shot to the time it hits the player. This was made intentionally short to demonstrate the ability of StereoTrack to allow for quick responses. To keep the task restricted to maneuvering, we did not let players fire back. We asked 5 participants to try the game. Each participant played 4 sessions, one of which was for training. A total of 10 laser beams were shot at the training session while 30 laser beams were shot at each of the three remaining sessions. As shown in Figure 4.5-left, all players ended
the last session with less than 20% error rate, defined as the number of times being hit by the 30 laser beams shot in a session.

4.5.2 Endless Running Game

![Endless running game](image)

**Figure 4.4:** Endless running game that uses walking-in-place but also allows players to side step to collect coins using StereoTrack

Walking-in-place (WIP) is a hands-free VR locomotion technique that closely mimics walking where users make step-like motions while remaining stationary [302]. A common limitation of WIP implementations on mobile VR [306, 229] is the lack of the ability to move sideways in a natural manner [305]. We integrated StereoTrack with an existing WIP implementation [9] to demonstrate its ability to effectively augment existing VR locomotion techniques through, in the case of WIP, providing responsive side-stepping. We developed a simple mobile VR coin collector game, similar to popular endless runner games like TempleRun. The game
Figure 4.5: Performance results of the feasibility user study to evaluate Stereo-Track’s ability to enable effective maneuvering in two example applications. Left: Percentage of player hits in the space shooter game. Right: Percentage of side coins missed in the endless running game.

consists of a long path of coins that has 90° turns as shown in Figure 4.4. To collect the coins, users either use WIP to move forward or side-step using StereoTrack. Side-stepping is required to collect side coins, which are ones positioned exactly to the right or to the left of a given coin. We disabled WIP’s head-directed steering to force players to side-step using StereoTrack. Forward locomotion and rate control were enabled by WIP while StereoTrack’s position updates on the x-axis enabled side-stepping. Five participants tried the game, where they were asked to move on a pace of their preference and collect as many coins as they could. Participants played 4 game sessions, one of which was for training. Each session consisted of a total of 50 coins, 11 of which were side coins. We measured the participants performance in terms of the number of side coins missed. By the third session, 4 out of the 5 players had less than 20% error rate, defined as the number of side coins missed out of the 11 side coins in a session, as shown in Figure 4.5-right.
4.6 Discussion and Limitations

We presented the design of StereoTrack, conducted three system suitability evaluations and showcased two example mobile VR applications. Precision evaluation showed that StereoTrack has a precision within 0.006mm at static conditions. This came at the expense of less motion sensitivity. Due to the strict low-pass filter, StereoTrack only considers frequencies above 3Hz. This implies that StereoTrack won’t detect velocities less than 0.06 m/s, making subtle body oscillations go undetected. StereoTrack has a mean RMS error over time of 1.3 cm/s to 1.6 cm/s, indicating that StereoTrack’s estimates would be $\approx 1$ m away from the ground truth after it is used continuously for 1 minute. This makes StereoTrack suitable for VR experiences that are either very short or don’t require a lot of translations (e.g., maneuvering). VR experiences that are longer and more active may be enabled by StereoTrack, but with frequent calibrations.

StereoTrack doesn’t require expensive instrumentation and has a low computational overhead that maintains a high frame rate. However, it is subject to a few limitations. Vertical positional tracking is not supported and it could be offered either by adding more speakers [184] or by gestures (e.g. jumps) enabled using inertial sensing [229]. Users can’t look more than 90° away from the speakers while translating for StereoTrack to function properly, limiting users freedom to freely explore the environment around them. Oculus PC VR platform widely uses 180° tracking, where notifications are used to redirect the user to face the sensor in order to overcome such limitation. Another candidate solution to overcome this limitation is the use of amplified head rotations [240]. StereoTrack only uses one sensor, the microphone, to achieve positional tracking. Sensor fusion with inertial sensing can be used as it was shown that it could improve the accuracy of acoustic
localization by 3% to 6% [184]. StereoTrack can be implemented with audio frequencies that are above the human hearing threshold of 19kHz to 20kHz [232], but the smartphones’ maximum supported frequency of 24kHz makes StereoTrack obtrusive to pets such as dogs whose hearing threshold is as high as 45kHz [176]. While developing StereoTrack, we noticed that sudden changes in room air pressure (e.g. when the lab door was opened) or occlusions to the line of sight (e.g. when someone passes between the smartphone and the speakers) causes momentarily noticeable jitter. Formal evaluations are needed to validate the significance of this effect and response measures should be implemented to, for instance, stop the tracking when its quality reaches a certain threshold.

The results of precision and accuracy evaluations were promising, but were achieved under optimized conditions such avoiding changes in air pressure and maintaining a line of sight. The accuracy evaluation was conducted with predetermined paths, which may not reflect realistic usage scenarios. Measuring the tracking delay was outside of the scope of our evaluation, though it is considered as key contributor to the end-to-end delay of a VR experience [145]. Our pilot studies were conducted with a limited number of participants, making it challenging to draw any rigorous statistically-driven conclusions. The example applications hinted at the potential of StereoTrack for mobile VR positional tracking. They were, however, limited to maneuvering tasks, leaving exploration and search tasks [174] unexamined.
5.1 Introduction

Virtual Reality (VR) holds a significant promise to transform and define the way we interact with computers [297], but mass market adoption of VR is threatened by VR sickness [207]. VR sickness (also known as cybersickness or simulator sickness) may involve a suite of symptoms including nausea, pallor, sweating, stomach awareness, increased heart rate, drowsiness, disorientation, and general discomfort [150]. While up to 67% of adults may experience mild to severe symptoms [74], there is substantial evidence that women are more likely to experience VR sickness than men [116, 203, 223, 111, 103]. Games have been a significant driver of innovation in VR and though gaming demographics have shown a near gender parity for a decade (48% of gamers were women in 2014 [12]), a recent survey [13] of 2,500 HTC Vive owners (a popular consumer VR headset) revealed that fewer than 5% of them were women. This low adoption rate of VR technology among women suggests that VR is currently less accessible to women than it is to men [57], and sex differences in the incidence of VR sickness are likely contributing to this problem.

Though there are various theories that aim to explain VR sickness, vection [175], i.e., the visually-induced illusion of self-motion, is currently considered the most likely trigger of VR sickness [42]. Self-motion typically involves inputs from the visual and vestibular systems and usually these inputs are in agreement. From a perceptual standpoint, natural walking in VR with the viewpoint updated by
Figure 5.1: Popular VR apps like Google Earth reduce the field-of-view during locomotion (e.g., tunneling) to block peripheral motion perception as to mitigate visual-vestibular conflict and to reduce VR sickness. However, prior studies suggest that reducing the field-of-view can impede spatial navigation performance in women.

Positional tracking should not generate VR sickness [178] because it generates vestibular and proprioceptive afferents (i.e., sensory signals) that match the perceived optical flow (i.e., full-field visual motion).

Unfortunately, the use of natural walking in VR is bounded by available positional tracking space, which in most consumer home environments is limited [99]. To navigate beyond the confines of limited available tracking space, users must switch from walking to using a controller-based artificial locomotion technique (ALT). Popular ALTs include target selection techniques, the most common of which is teleportation; or steering-based techniques (e.g., using a controller). The optical flow generated using steering-based locomotion generatesvection, which in the absence of any vestibular/proprioceptive afferents confuses the senses and
may lead to VR sickness [42]. Teleportation circumvents sensory conflict because it instantly translates the virtual viewpoint which avoids any optical flow generation. Though it is a standard ALT in many VR experiences, there are significant concerns with using teleportation such as low presence and disorientation [49, 146]. The discontinuous locomotion offered by teleportation is also a challenge for multiplayer games.

To reduce vection-induced VR sickness, various solutions have been proposed (see [90] for an overview). Motion from optical flow is primarily detected by the rods on the periphery of the retina [312]. Blocking the perception of peripheral motion by reducing the user’s field-of-view (FOV) during locomotion [262, 177] is therefore considered an effective strategy to reduce VR sickness. FOV restriction -also known as “tunneling” -is already widely used in popular VR experiences like Google Earth VR and is recommended by both Google’s [11] and Oculus’ [15] VR design guidelines as a feasible strategy for reducing VR sickness.

However, these design guidelines seem to conflict directly with results from prior studies [87, 299] that found that women benefit from spatial navigation using a larger FOV. Sex differences in spatial cognition have been well documented [315]. Women navigate predominantly using landmarks where men rely mostly on geometric information, such as distance and vestibular cues [256]. These differences affect spatial navigation performance in real [88] and virtual 3D environments [93]. Studies have shown that men are more adept at 3D spatial navigation [88], but this sex bias can be reduced by providing women with a larger FOV –which improves their ability to perceive landmarks [87, 299]. Though prior studies used desktop environments, their findings seem relevant to VR given that the FOV of consumer VR headsets (up to 110°) is still well below the human binocular FOV (up to 190°).
Existing studies on the effectiveness of an FOV restrictor to reduce VR sickness [100, 262, 177] have not explored sex as a variable (i.e., not enrolled an equal number of women/men) nor evaluated the effect of this strategy on spatial navigation performance in women. Given that women were found to have significantly higher thresholds for motion perception [112] and fewer rods in their retinas to detect peripheral motion [28], one can question whether FOV restriction is an effective strategy for reducing VR sickness in women, given that they likely impede their spatial navigation performance [87, 299]. This paper makes the following contributions: (1) we investigate if FOV restriction impedes spatial navigation performance in women; and (2) we investigate whether an FOV restrictor is equally effective in reducing VR sickness symptoms in both sexes. Both contributions provide insight into how to make VR more accessible to women.

5.2 Background

Motion sickness (MS) is experienced as a result of motion patterns of an organism that result in symptoms that include dizziness, cold sweating, headache, increased salivation, and nausea [151]. Visually induced motion sickness (VIMS) is a related phenomena that has induced symptoms similar to those of MS without being subject to physical motion [160]. VIMS is a common adverse effect that results from the exposure to computer simulations in general and VR experiences in particular. Several terms have been given to VIMS in the literature [151], the most common of which are: simulator sickness, cybersickness, and VR sickness. We choose to use the term "VR sickness" to refer to VIMS from this point forward. Several theories
attempted to explain the cause of VR sickness. Some theories attribute VR sickness symptoms to a sensory conflict [246] while others believe that it is a result of a failure to maintain postural stability while being immersed in the virtual environment [248]. Other less prominent theories include the eye movement theory [95] and the poison theory [307]. None of these theories is complete, though the sensory conflict theory is the most accepted [151, 160]. Another body of research on VR sickness aimed to show the influence of individual differences on the susceptibility to VR sickness [160, 274, 172, 114]. Sex was among these investigated differences, with women being more susceptible to VR sickness [246, 274, 203]. Theories that attempt to explain such sex difference include the under-reporting of VR sickness symptoms by men [38], hormonal differences [84], evolutionary differences [114], and wider field of view of women [160, 172].

5.2.1 VR Sickness restriction Mechanisms

Several medical and behavioral countermeasures to VR sickness have been proposed [151, 172]. Medical interventions suffer from adverse effects that limit their applicability [114]. Behavioral countermeasures, on the other hand, focus on mutating the course of interaction between the user and the virtual stimuli. Some of these countermeasures aim to modify the behavior of the user (e.g., by reducing head movements or by regulating breathing) to minimize the incidence of VR sickness while others manipulate the visual or vestibular stimuli without counting on the user’s involvement for the countermeasure to work [151]. Relevant examples of the latter to this research are ones that involve manipulations of the visual stimulus such as the restriction of field of view [44, 152, 177, 100, 40], the use of independent virtual backgrounds and fixed reference frames [235, 94, 328],
the dynamic control of travel velocity [300, 109], freezing head rotations [148], and non-salient objects blurring [211].

5.2.2 FOV Manipulation in Virtual Environments

Manipulation of FOV is the most relevant intervention to our study. Several studies investigated the effects of manipulation of both the horizontal display field of view, defined as the angle that extends from the eye to the left and right edges of the display, and the horizontal geometric field of view, defined as the angle that extends from the virtual camera to the left and right edges of the viewpoint frustum [87, 44]. Many of these studies revealed positive effects of FOV manipulation on VR sickness [262, 177, 44, 152, 100] and presence [262, 177, 167]. Other studies, on the other hand, showed the negative effect of FOV manipulation on task performance on virtual environments [325, 210, 233] and the magnitude of such effect was shown to be more significant on women [87, 299].

5.3 User Study

The goal of this study is to examine the effect of dynamic FOV restriction on sex differences in VR sickness and spatial navigation performance in the context of a triangle completion task.
Table 5.1: Summary of participants ratings of their frequency of using VR and their tendency of getting motion or VR sick on a scale of 1 (never) to 5 (very frequently). The results are reported in the form of percentage (count).

5.3.1 Participants

We recruited 30 participants, two women of whom exited during the first session due to severe discomfort\(^1\), leaving us with 28 participants (14 women) whose data is used in this study. Participants age ranged from 18 to 33 years (average = 23.04, SD = 3.59). Participants were recruited by flyers and word-of-mouth and they were affiliated with the local higher education institutions. Participants were asked to rate their frequency of using VR and their tendency to get motion or VR sick on a scale of 1 (never) to 5 (very frequently). The results are summarized in Table 5.1.

All participants were compensated with a $15 Amazon gift card. The user study was approved by an IRB.

\(^1\)One of these women exited while experiencing the full FOV condition while the other was experiencing the dynamically changing FOV condition.
5.3.2 Experiment Design

Our study is a $2 \times 2$ mixed factorial design with sex as the between-subject factor and the FOV condition as the within-subject factor. The latter factor has two levels: no FOV restriction (RN) and dynamically changing FOV (RY). We inspect the effect of these factors on seven dependent variables: (1) the home position estimation error (HPE), (2) the Simulator Sickness Questionnaire [150] (SSQ) total severity score (TS), (3) the SSQ-Nausea score (N), (4) the SSQ-Oculomotor discomfort score (O), (5) the SSQ-Disorientation score (D), (6) the average discomfort score (ADS) [100], and (7) the ending discomfort score (EDS) [100]. To account for order effects, half of the participants started with the RN condition (Group A) while the remaining half started with the RY condition (Group B). To ensure that each group contained an equal number of men and women, we alternated the assignment of men and women across the two groups.
5.3.3 FOV Test Conditions

In the RN condition, no FOV restriction was applied and participants were consequently exposed to the full visual field provided by the HMD’s FOV. In the RY test condition, on the other hand, we followed the strategy of Bolas et al. [40] and Fernandes and Feiner [100] who proposed to manipulate the FOV as a function of the user’s current state. We manipulated the FOV according to changes of the participant’s linear and angular velocities [100]. In other words, the FOV was decreased as the participant’s virtual speed (linear or angular) increases. To restrict the FOV, we used a black texture with a transparent circular cut-off, shown in Figure 5.2-left, whose radius \( R_{FOV,r} \) is controlled according to following equation:\(^2\)

\[
FOV_{r,t} = FOV_{r,t-1} \times \left[ 1 - (RF_{max} \times max(\frac{v_r}{v_{max}}, \frac{\omega_r}{\omega_{max}})) \right]
\] (5.1)

\( FOV_{r,t-1} \) is the radius of the circular cut-off at time \( t - 1 \). \( RF_{max} \) is the amount of restriction applied to \( FOV_{r,t-1} \) at the maximum virtual speed. \( v_r \) and \( \omega_r \) are the virtual linear and angular virtual speeds, respectively, at time \( t \). \( v_{max} \) and \( \omega_{max} \) are linear and angular virtual speeds, respectively, at which the maximum FOV restriction is applied. We set \( RF_{max} \) to 0.75. This is equivalent to a minimum FOV of \( 50^\circ \) on the HTC Vive with a horizontal FOV of around \( 100^\circ \), which we empirically found close to the max FOV restriction applied by popular VR experiences such as Google Earth VR [19]. The value of \( v_{max} \) was set to 1.4 m/s, a value that matches the average preferred walking speed of humans [200]. We empirically found \( 180^\circ /\text{sec} \) worked best as a maximum angular speed to ensure a frequent FOV restriction as a response to the dynamics of head movement expected from our task. The FOV restriction was applied gradually over time and the edges of the circular cut-off were

\(^2\)https://github.com/SixWays/UnityVrTunnelling
feathered as these factors were found to reduce participants’ distraction [100].

5.3.4 Instrumentation

The artificial stimuli, both visual and aural, were delivered via the HTC Vive HMD with a diagonal FOV of 110°, refresh rate of 90Hz, a combined resolution of 2160×1200 pixels, six degrees of freedom (DoF) for position and orientation tracking, and adjustable interpupillary (IPD) and focal distances. The headset was powered with a 2.8GHz Intel Core i7 processor with 16GB of memory and NVIDIA GeForce GTX 1070 graphics card running Windows 10. Participants provided input using an XBox controller that we preferred over the Vive’s motion sensing controller because participants were likely to be more familiar with this controller and the profile of the thumbstick used for navigation provides better tactile feedback than the Vive’s touchpad. We used Unity3D engine and the SteamVR plugin to develop the artificial stimuli. We used the tunneling effect implementation of SixWays³ to dynamically manipulate the FOV as per the specifications mentioned in Section 6.3.4. Participants’ IPD was measured using the PD Meter app⁴ that runs on Android.

5.3.5 Virtual Environment

For both experiment sessions, we adapted the Rocky Hills Environment - Light Pack asset⁵ from the Unity Asset Store. We mapped the environment’s measure-

³https://github.com/SixWays/UnityVrTunnelling
Figure 5.3: Top-down view of the virtual environment we used in the experiment sessions

ment system from Unity units to Metric units such that three Unity units are equivalent to one meter. Such mapping was important for design decisions that involved knowledge about distance such as target travel distance for a given task and appropriate travel speed. The environment (Figure 5.3) is a $200 \times 200$ meter forest-like space that consists of trees, rocks and hills that can be used as subtle spatial cues. Three forest cabins were distributed over the hills of the environment to be used as salient landmarks during the task. Some parts of the terrain were made uneven in order to expose participants to optic flow at the vertical axis. For the training session, we used the Mecanim Example Scene\(^6\) from the Unity Asset Store. Five-meter high red posts (Figure 5.2-right) were used as waypoints, each representing a triangle vertex that participants navigate to, one after the other. Unity’s UI panels, text, and sliders were used to communicate the task instructions and to collect the discomfort score from participants during trials. Participants used the controller’s thumbstick to control the rate of travel at a speed that varied between 0 and 1.4 m/s. The same thumbstick was used for steering in a direction relative the head’s

\(^6\)https://assetstore.unity.com/packages/essentials/tutorial-projects/mecanim-example-scenes-5328
forward vector.

**Figure 5.4:** The triangle completion task we use in this study. $S$ = starting position, $E$ = estimated position, $W_1$ = first waypoint, $W_2$ = second waypoint, $W_2S$ = the vector from the second waypoint to the starting position, and $SE$ = the vector from the starting position to the estimated position

### 5.3.6 Task

Humans use several fundamental skills to achieve effective navigation such as spatial updating, spatial cognitive mapping, and constrained route planning [249]. In this study, we examine the effect of FOV restriction on participants’ spatial updating abilities.

Spatial updating can be achieved through (1) path integration, where users update their current position based on an estimate of the direction and distance travelled obtained from visual, vestibular and proprioceptive senses [312]; and (2) landmark navigation, where users update their current position when a known landmark is identified [179]. Humans use the information collected from path integration and landmark navigation to form a survey representation that captures the distances and directions of the traversed trajectory to help them plan future navigation tasks [180]. To examine the quality of participants’ survey knowledge, we used a "triangle completion" task that was described in early work on spatial navigation [180] and that has been used for assessing spatial navigation performance.
in VR [146]. This task required participants to travel from a starting position to two consecutive waypoints (Figure 5.4) shown one after the other, which are non-collinear with the starting position and which form two adjacent legs of a triangle. After arriving at the second waypoint, participants were then asked to navigate back to the starting position and confirm their selection using the controller’s (A) button. Being able to navigate back to each starting location relies on a combination of path integration and landmark-based navigation. Participants’ ability to navigate back to the starting position requires them to compute a new trajectory; an ability that is contingent on the quality of their survey representation [180]. As the focus of our study is to examine the effect FOV restriction (a visual manipulation intervention) on forming an effective survey representation, we aimed to limit the spatial updating cues to optical flow and minimize the interference from proprioceptive and vestibular cues. We achieved this by having participants navigate the virtual environment using joystick-control locomotion.

Similar to Loomis et al. [180], a total of 27 triangles were produced as a result of varying the distance of the first two triangle legs (A and B, respectively) and the turning angle $\alpha$ that corresponds to $180^\circ$ minus the angle between the two legs. Leg A was one of three values (10, 15, or 20m) as well as Leg B (8, 12, 18m) and $\alpha$ (60, 90 or 120$^\circ$). These distances and angles were selected such that the total exposure time after completing the 27 triangles is 25 minutes per session. To minimize learning effect, we varied the starting position of participants. The triangles were distributed over three selected zones and the starting vertex was varied across triangles that belong to the same zone. The order of the produced triangles was randomized to minimize the chance of having two consecutive triangles that have the same zone and starting vertex. Two sequences of triangles were produced for each experiment session.
5.3.7 Procedure

The experiment was conducted in a room that is free of noises and physical obstacles. Participants were greeted and seated to be given a short presentation at which the experimenter explained the goal the study, the sequence of the experiment, the risks involved, the collected data, and the details of the training and experiment sessions. Participants were then asked to fill the first SSQ [150] to provide a baseline input of their relevant symptoms. The participants’ IPD was then measured and the measurement was used to set the IPD of the VR headset accordingly. If a participant’s measured IPD was lower than the headset’s minimum (60.8mm for the Vive), the headset’s IPD was set to its minimum. Participants were then asked to stand at a marked position in the tracking space and were assisted to wear the VR headset and hold the controller so that they could start the training session. We chose to have participants standing in order to rotate with their body. Standing and rotating in place do generate some vestibular and proprioceptive cues. While the magnitude of these cues while standing and rotating in place is significantly lower than that of walking, these cues might interfere with the goal of our study that aims to limit spatial updating cues to optical flow. However, we made this choice since viewpoint rotation using a controller was found to cause unnecessary discomfort [100], aside from being uncommon to rotate with a controller in common VR experiences. The goal of the training session was to familiarize the participants with the controls needed to provide input and to give them an opportunity to practice the experiment task. To satisfy the former goal, participants were asked to move in each of the four directions (right, left, forward, and backward) one after the other and move the slider all the way to the right and then all the way to the left. They were then asked to complete three triangle completion tasks after which the train-
ing session concludes. To give participants a sense of their performance during the training session only, an arrow was shown at the actual starting position after they provided their estimation.

Participants then took part in two experiment sessions; one for each condition. Group A participants started with the RN condition while Group B participants started with the RY condition. In each session, participants performed a total of 27 trials, each involving one triangle completion task. After each trial, participants were prompted with a slider [100] to provide their level of discomfort from 1 to 10, with level 10 signifying the highest level of discomfort [247]. Similar to Fernandes and Feiner [100], participants were told that when a value of 10 is selected, the experiment will be terminated, but they were assured that they will be fully compensated anyway. Participants were encouraged to strike a balance between time and accuracy. After each session, participants were asked to fill a post-exposure SSQ. A mandatory ten-minute break was given to each participant after the first session.

Upon the completion of both sessions, participants were asked to fill a post-experiment questionnaire at which they provided their demographic information that included sex, age, frequency of exposure to VR (five-point Likert scale), and tendency of being motion and/or VR sick (five-point Likert scale). The total duration of the study took approximately one hour and a half.
5.3.8 Measurements

In order to calculate HPE, we use the following formula [146]:

\[
HPE = \frac{|\vec{S} \vec{E}|}{|\vec{W}_2 \vec{S}|}
\]  \hspace{1cm} (5.2)

where \( |\vec{S} \vec{E}| \) is the magnitude of a 2D vector whose initial and terminal points are the horizontal plane coordinates of the starting and estimated positions, respectively, and \( |\vec{W}_2 \vec{S}| \) is the magnitude of a 2D vector whose initial and terminal points are horizontal plane coordinates of the second waypoint and the starting position, respectively (see Figure 5.4). We use the vector \( |\vec{S} \vec{E}| \) as it captures errors in both heading and distance while we use the vector \( |\vec{W}_2 \vec{S}| \) to normalize the differences across triangle designs. We also collected the 2D trajectory at the horizontal plane produced by each triangle completion task for all participants for further analysis, if needed. We use the data collected from the SSQs along with the self-reported discomfort scores in order to measure VR sickness. Data collected from the SSQ is used to calculate four associated scores, namely: TS, O, and N, and D scores. These scores were calculated as per the conversion formulas by Kennedy et al. [150]. The calculated scores of the baseline SSQ were subtracted from those of the first post-exposure SSQ to obtain the latter’s relative SSQ scores. Similarly, the calculated scores of the first post-exposure SSQ were subtracted from those of the second post-exposure SSQ to obtain the latter’s relative SSQ scores. We averaged the discomfort scores for each participant per FOV condition to calculate ADS and we used the last discomfort score for each participant per FOV condition to calculate EDS.
Table 5.2: Quantitative measures of the Home-Position-Estimation error (HPE), discomfort scores, and relative Simulator Sickness Questionnaire in terms of mean (standard deviation).

5.4 Results

Table 5.2 shows the mean and the standard deviation of the HPE, ADS, EDS, and SSQ results. We analyze the results in the remainder of this section.

5.4.1 Spatial Navigation Performance

The HPE was used as a measure of spatial navigation performance of the 28 participants in this study. Figure 5.5 shows a summary of the results. A 2-way mixed-model ANOVA did not find an interaction between sex and FOV ($F_{1,26} = 3.11, p = .09$). No significant effects of sex ($F_{1,26} = 2.13, p = .16$) or FOV ($F_{1,26} = 0.09, p = .77$) were found either. Spearman’s rank correlation did not find a significant association between HPE and frequency of using VR ($r_s = −.20, p = .31$). The correlation between HPE and motion/VR sickness tendency, however, was
significant \( (r_s = .52, \ p < .05) \), indicating a positive association between spatial error and motion/VR sickness history.

### 5.4.2 VR Sickness

VR sickness was measured in terms of the self-reported discomfort score, from which we measured ADS and EDS; and the SSQ questionnaire, from which we calculated the SSQ’s TS, N, O and D scores. Using the EDS score, we found that 6 out of the 28 participants (4 men) were asymptomatic. Unlike other studies [100] that used the ADS as a criteria to determine which participants were asymptomatic, we use the EDS due to its significant positive correlation with the SSQ-TS results (EDS: \( r_s = .52, \ p < .05; \) ADS: \( r_s = -.04, \ p = .77 \)) and because both EDS and SSQ results capture the participant’s discomfort level at the end of a session. As visual acceleration can affect the incidence of VR sickness [171], we analyzed the amount of time at which participants were travelling at a fixed speed. Both men and women travelled at a fixed speed more than 87% of the time (women: 87.60%, men: 87.56%;
Figure 5.6: Average and ending levels of discomfort. Results are aggregated per sex and FOV condition. RN = No FOV restriction; RY = Dynamically changing FOV. Error bars represent the standard deviation.

\[ F_{1,22} = .001, p = .97 \] We report the VR sickness results of our 28 participants as follows.

Discomfort Score

A 2-way mixed-model ANOVA did not find an interaction effect between sex and FOV condition on both the ADS \( (F_{1,26} = .5, p = .49) \) and the EDS \( (F_{1,26} = .70, p = .41) \). While no significant difference between sexes was found with respect to both ADS \( (F_{1,26} = .85, p = .37) \) and EDS \( (F_{1,26} = .90, p = .35) \), FOV restriction resulted in significantly lower ADS \( (F_{1,26} = 9.30, p < .05) \) and EDS \( (F_{1,26} = 7.23, p < .05) \). Figure 5.6 summarizes the results of the average and ending discomfort scores. Spearman’s rank correlation did not find a significant association between the reported frequency of using VR and neither ADS \( (r_s = -.24, p = .22) \) nor EDS \( (r_s = -.25, p = .20) \). Similarly, the reported tendency of motion/VR sickness was not found to be correlated with neither ADS \( (r_s = .36, p = .061) \) nor EDS \( (r_s = .36, p = .059) \).

\footnote{Due to a tracking error, we lost the speed data of the first 4 male participants, resulting in conducting this analysis using the speed data of the remaining 24 participants.}
A 2-way mixed-model ANOVA did not find an interaction effect between sex and FOV condition on all SSQ scores: TS ($F_{1,26} = .66, p = .42$), N ($F_{1,26} = .45, p = .51$), O ($F_{1,26} = 3.52, p = .072$), and D ($F_{1,26} = .006, p = .94$). No significant differences between men and women were found in any of the SSQ scores either: TS ($F_{1,26} = .27, p = .61$), N ($F_{1,26} = .032, p = .86$), O ($F_{1,26} = .64, p = .43$), and D ($F_{1,26} = .27, p = .61$). FOV restriction resulted in lower TS ($F_{1,26} = 5.04, p < .05$) and O ($F_{1,26} = 6.01, p < .05$) scores while no significance was found for the D ($F_{1,26} = 3.46, p = .074$) and N ($F_{1,26} = 3.80, p = .062$) scores. Spearman’s rank correlation did not find a significant association between the reported frequency of using VR and any of the SSQ scores: TS ($r_s = -.26, p = .18$), N ($r_s = -.28, p = .15$), O ($r_s = -.23, p = .23$), D ($r_s = -.30, p = .12$). A significant positive correlation was found between the SSQ-TS score and the tendency of getting motion/VR sick ($r_s = .39, p < .05$) while no significant correlation between motion/VR sickness tendency and the rest of the SSQ scores was found: N ($r_s = .35, p = .066$), O ($r_s = .33, p = .081$), D ($r_s = .36, p = .060$). As shown in Figure 5.7, both sexes experienced an SSQ symptom profile [275, 274] of D > N > O in the RN condition. While women
maintained the same profile in the RY condition, men’s profile changed to O > D > N.

5.5 Discussion and Future Work

Unlike our expectation, no significant sex difference was found in spatial navigation performance, contradicting with previous studies that suggested otherwise [93, 29, 256, 23, 254]. The experimental conditions of these studies, however, were different, e.g., they used a virtual water maze task in desktop environments [29, 93, 256], a path following task in CAVE environment with children [23], or a search task with an HMD having 48° FOV [254]. Aside from using a different FOV that ranged between 50° to 100°, our study differs fundamentally from previous evaluations in terms of the target navigation skill. Our study aimed to focus on the evaluation of the spatial updating skill which relies on survey representation that is affected by both landmark-based navigation and path integration [180]. Since women heavily rely on landmarks for navigation while men navigate mostly using geometric information [256], we designed a virtual environment that contains both subtle and salient landmarks to give men and women a fair chance to form a quality survey representation. This might have resulted in finding no significant sex difference in spatial updating.

Regardless of sex, we did not find a significant effect of FOV restriction on spatial navigation performance. This is different from what was reported in earlier studies [210, 233, 325]. Some of these studies used desktop VR [210, 233]. To evaluate spatial navigation performance, some studies used a search task [210, 233] while others used an obstacle avoidance task [325]. This is different from our study
at which we used an HMD and a triangle completion task. The restriction mechanism used in our study can be another contributing factor. Unlike previous studies that restricted FOV throughout the virtual experience, our restrictor only restricts FOV as a response to linear and angular speeds. This dynamic behavior of the restrictor might have given participants an opportunity to frequently gain a full view of the virtual environment while they were stationary, which might have caused them to perform efficiently in both FOV conditions.

Due to the physiology of women’s eyes [28] along with their vision perception [112] as we explained earlier, we expected that FOV restriction would not be an effective intervention to reduce the VR sickness symptoms in women. However, FOV restriction was shown to be effective in mitigating VR sickness symptoms in both sexes as shown in the ADS, EDS, and SSQ results, agreeing with the results of previous studies [262, 177, 44, 152, 100]. Our analysis did not find a significant sex difference in any of the VR sickness measures. This seems surprising when compared with previous studies [246, 274, 203] that report on the higher susceptibility of women to VR sickness. This contradiction could be due to the nature of the virtual task, which was shown to have an effect on the incidence of VR sickness in general [194] and among sexes [203]. A recent study agrees with our findings [178]. Unlike our study, however, sex groups in the former study are unbalanced (64 men vs. 43 women).

Visual acceleration input from the virtual experience can increase sensory mismatch between the visual and vestibular systems, which might lead to a greater incidence of VR sickness [171]. Both men and women travelled at a fixed speed more than 87% of the time. This low acceleration/deceleration rates might have contributed to masking potential sex differences in VR sickness.
Overall, although the results of this study did not support our hypotheses, they suggest a valuable implication: that FOV restriction seems to be an effective intervention in reducing the incidence of VR sickness without having a negative effect on spatial navigation performance in both sexes. However, we would like to reiterate that our study only tested participants’ spatial updating skills through a triangle completion task. Follow-up studies are indeed needed to test the effect of FOV restriction of other navigation skills such as spatial mapping and constrained route planning [180].

Six participants (21% of the total) were asymptomatic and only two of whom were women. Most of those asymptomatic participants reported very frequent use of VR, which could explain why they showed no symptoms [160, 275]. Considering that typically 5% to 10% of the participants in early VR studies are asymptomatic [275], having 21% of our participants reporting no VR sickness symptoms is relatively high. The use of a state-of-the-art HMD and having participants use body input to rotate in VR could be contributing factors to alleviating VR symptoms [172] and hence increasing the number of asymptomatic participants.

Our correlation analysis did not find a significant association between prior VR exposure and VR sickness symptoms. This contradicts with previous studies which showed a positive effect of prior experience with VR on the severity of VR sickness symptoms [133, 125]. This contradiction may stem from how we quantified prior VR exposure compared to previous studies. As the effect of prior VR exposure was not central to our study, we simply asked participants to rate their frequency of using VR on a 5-point Likert scale. This is different from previous studies that designed their experiments around a controlled exposure procedure. The use of a state-of-the-art HMD might also have made the need for prior VR
exposure to reduce VR sickness symptoms [160] less relevant, which might have led to the lack of correlation between prior VR experience and reported levels of discomfort in our study.

A significant positive association between motion/VR sickness history and VR sickness total severity was the only significant correlation that our analysis could find among the VR sickness measures we used in this study. This finding generally agrees with the findings by Stanney et al. [274] that showed a significant correlation between VR sickness symptoms severity, measured with the SSQ, and VR sickness history, measured with the Motion History Questionnaire (MHQ) [149]. Unlike their study, however, we did not find a significant correlation between the SSQ sub-scales (i.e., N, O and D scores) and the motion/VR sickness history. This difference might be due to our use of a 5-point Likert scale to measure motion/VR sickness history, which is fundamentally different from the MHQ.

The SSQ symptom profile was D > N > O for both sexes when no restriction was applied to the FOV. This is the same profile reported for men in a previous study [274], but the same study reported that women had a profile of D > O > N. While women’s profile was not affected by FOV restriction, it was interesting to observe the change of men’s SSQ symptom profile to O > D > N when the FOV was restricted. The notable difference between the two profiles seems to be related to Oculomotor discomfort. The cause of having Oculomotor discomfort higher in men than the other two SSQ scores when FOV was restricted is unknown to us. One future venue to explore that might help providing more explanation to this finding is the relationship between the differences in the visual system among men and women [313] and dynamic field of view restriction in virtual environments.

An improperly calibrated IPD on an HMD may lead to eye strain which is a
symptom of VR sickness that is measured by the SSQ’s oculomotor discomfort score. The measured IPD of most women participants in our study was below the Vive’s minimum supported IPD of 60.8mm. As a result, we had to set the IPD of the headset to a value slightly higher than theirs. However, even with an improperly calibrated IPD, we did not observe any significant difference across sexes in oculomotor discomfort, which was also the least observed score in women according to their SSQ symptom profile in both FOV conditions.

A few limitations may have affected the results of our study. We asked participants to perform the study tasks while standing up in order to control their virtual rotation using body input due to the reported adverse effects of virtual rotation using a controller [100]. Most of today’s VR experiences expect users to be standing up, especially the experiences offered by platforms that allow users to alternate between natural walking and artificial locomotion. However, having participants standing up for typically 25 minutes per session may have interfered with their perception of what accounts for discomfort, which may resulted in reporting fatigue due to prolonged standing as discomfort. Participants experienced both conditions one after the other, separated by 10 minutes. This may have affected their spatial navigation performance and reported discomfort at the second session. We mitigated this using counterbalancing as explained earlier in Section 6.3.3.

For future work, we would like to compare the effect of different FOV manipulations (e.g., independent rest frames [235] and non-salient objects blurring [211]) on sex differences. We also plan to study the effect of FOV restriction on sex differences in the context of spatial navigation tasks that differ from the one we used in this study. We also consider conducting follow-up studies with HMDs that has
FOV closer to that of humans such as the 210°-FOV StarVR®.

https://www.starvr.com/
CHAPTER 6
EFFECTS OF VR FIELD-OF-VIEW RESTRICTION ON SPATIAL LEARNING
AND VR SICKNESS AMONG MEN AND WOMEN USING A VIRTUAL
MORRIS WATER MAZE

6.1 Introduction

Virtual Reality (VR) is soon to become a mainstream medium. The mass usage of VR has already gone beyond gaming to cover domains such as education, tourism, surgical training, and architectural inspection. For it to be universal, VR experiences have to be equally accessible for all users, irrespective of their individual differences. Women are among the under-represented groups in the VR consumer market [13]. One way to increase the adoption of VR among women is to minimize the adverse effects associated with using VR, the most prominent of which is VR sickness, especially since it has been observed that women are more susceptible to VR sickness than men [116, 203, 223, 111, 103]. While there are several theories that attempt to explain its cause, VR sickness is commonly known to be elicited by a conflict between the visual, vestibular, and proprioceptive sensory inputs [246], leading to unpleasant symptoms that can be as mild as cold sweating or as severe as vomiting [151].

High-end consumer VR platforms, such as the HTC Vive and the Oculus Rift, do provide six degrees-of-freedom (DoF) positional tracking, enabling users to navigate virtual environments using real walking. Positional tracking, however, is limited to the confines of the physical environment. This calls for alternative virtual locomotion techniques that enable users to move virtually while being stationary in the physical world [171]. Virtual locomotion techniques that provide
continuous rate control, such as gaze-directed steering using a joystick, stimulate
the visual system with optic flow that leads to the illusion of self-motion, known as
vection [153], while corresponding physical motion is lacking. Such disagreement
between virtual and physical motion leads to a conflict between what is reported
by visual, vestibular and proprioceptive systems, leading to an increased risk of
VR sickness.

Various techniques have been proposed as countermeasures to VR sickness.
One recently proposed and common method is blocking the peripheral optic flow
during locomotion using field of view (FOV) restrictors (also known as tunnel-
ing), similar to those used in popular VR experiences like Google Earth VR [19]
(e.g., see Figure 6.1-right). The potential of this technique to reduce VR sickness
has been demonstrated by recent studies [100]. A limited FOV, however, was also
shown to impede spatial navigation performance [167, 210, 233], and this effect
was especially significant for women [87]. The reduction of VR sickness at the
cost of degraded spatial performance might be considered a fair trade-off. How-
ever, women have higher motion perception thresholds [112] and fewer rods for
peripheral motion detection [28]. This begs the question whether FOV restriction
reduces VR sickness equally in men and women. In this paper, we examine the
differential effect of dynamic FOV restriction on VR sickness and spatial learning
in men versus women using a virtual Morris Water Maze.

6.2 Background

Visually induced motion sickness (VIMS) is a negative physical response to digital
simulations and virtual experiences with symptoms that include headache, cold
Figure 6.1: Popular VR apps like Google Earth reduce the field-of-view during locomotion (e.g., tunneling) to block peripheral motion perception and mitigate visual-vestibular conflict and thus reduce VR sickness. However, prior studies suggest that reducing the field-of-view can impede spatial navigation performance in women.

sweating, vertigo, and nausea [150]. Although it shares many of the adverse effects associated with motion sickness, VIMS can be induced as a result of exposure to a visual stimulus without the need for physical motion [160]. As VIMS has been known in the VR research literature as VR sickness, cybersickness, and simulator sickness [145], we use the term VR sickness from this point forward.

To date, there is no definitive explanation of what causes VR sickness. The sensory conflict theory [246] attributes VR sickness to a conflict between the visual, vestibular, and proprioceptive senses. The postural instability theory [248] links VR sickness to a disruption of posture stability caused by the motion patterns of the visual stimulus of the virtual experience [248]. Other explanations of what causes VR sickness such as the eye movement [95] and the poison [307] theories also exist. The sensory conflict theory, however, seems to be the most widely accepted theory [151, 160].
Among individual traits that have been shown to increase the vulnerability to VR sickness [160, 274, 172, 114], sex is commonly reported, with the observation that women are more susceptible to VR sickness than men [246, 274, 203]. Hormonal differences [84], evolutionary differences [114], under-reporting of sickness symptoms by men [38], and differences in field-of-view [160, 172] have all been proposed as explanations of this gender disparity in the incidence of VR sickness.

Various methods have been proposed to reduce the incidence of VR sickness. It can be reduced by medical interventions [172], but effectiveness is limited due to adverse side effects [114]. An alternative is to use behavioral interventions that either regulate the user’s behavior (e.g., head movements or breathing), or manipulate the visual stimuli [151]. Manipulation of the visual stimuli has been achieved by controlling the FOV [44, 152, 177, 100, 40], using independent backgrounds and rest frames [235, 94, 328], dynamically controlling travel velocity [300, 109], freezing the virtual viewpoint rotations [148], and blurring non-salient virtual objects [211].

The effect of FOV on participants’ performance and quality of user experience in VR has been the focus of several studies. Reducing the FOV was shown as an effective intervention to reduce the incidence of VR sickness [262, 177, 44, 152, 100]. Other studies, on the other hand, showed that larger FOV can increase the sense of presence [262, 177, 167]. FOV restriction was also shown to impede spatial navigation performance [325, 210, 233] with a greater negative impact on women [87, 299].
6.3 User Study

This study examines the effect of dynamic FOV restriction on sex differences in VR sickness and spatial learning as participants navigate a virtual environment.

6.3.1 Participants

We recruited 31 participants, but three women exited the study due to severe discomfort. This left us with 28 participants (14 women) who completed the study (mean age: 22.96, SD: 5.41) and whose data was considered in our analysis. When they were asked about their frequency of getting motion or VR sick on a scale that goes from 1 (rarely) to 5 (very frequently), their responses were distributed as follows: 1 (50%), 2 (32%), 3 (7%), 4 (11%), and 5 (0%). When they were asked to rate their frequency of using VR on the same scale, their responses were distributed as follows: 1 (39%), 2 (36%), 3 (11%), 4 (7%), and 5 (7%). Participants were recruited from the academic institutions nearby through flyers and word of mouth. Each participant was given a $15 Amazon gift card as compensation for their participation. The study was approved by an IRB.

6.3.2 Tasks

To navigate an environment, real or virtual, humans need one or more spatial skills that include spatial updating, spatial cognitive learning and constrained route planning [249]. In this study, we evaluate the effect of dynamic FOV restriction on spatial cognitive learning using a virtual Morris Water Maze and an object place-
Virtual Morris Water Maze Task

First introduced by Richard Morris in 1981 [201], the Morris Water Maze task has been one of the gold standards to evaluate spatial memory [29]. Briefly, the task tests the quality of the subject’s spatial learning through their ability to remember the location of a hidden object in reference to distal cues that are not co-located with the hidden object [201]. A virtual Morris water maze was used to confirm sex differences in spatial cognition with men primarily relying on path integration and women on landmark-based navigation [337, 256]. Similar to these studies, we use the virtual environment we developed as specified in Section 6.3.5 to test participants’ spatial cognitive learning using a virtual Morris Water Maze task.

In each trial, participants start from a point at the edge of the pool while facing its wall. Participants are then asked to use the thumb-stick to move around in the pool to find a hidden platform under the water surface within 1 minute. The platform is found when participants cross its location under the water surface. This is when they hear a distinct sound feedback and see their viewpoint elevated with the platform. If the platform is not found within 1 minute, participants hear a different sound feedback and the platform is made visible. When that happens, participants are asked to move towards the platform until they stand on it. Participants are then moved to the next trial after 5 seconds.

To assess participants spatial learning over time, they are asked to locate the hidden platform over several trials grouped in blocks. Within a block of trials, the location of the platform is fixed while what changes is the starting point of the
participant in each trial. In blocks that follow, the location of the platform, the sequence of starting positions, and the distal cues (i.e., the placed images on the wall of the arena) are changed. Participants go through 6 blocks for each FOV condition, each having 6 trials. Insertion points were separated by 30° and a random sequence of insertion points was created for each block of trials.

Object Placement Task

Although the Morris Water Maze task has commonly been used to evaluate the quality of participant’s spatial memory through allocentric navigation, a recent study showed the location of the platform can be learned by depending on egocentric cues that are independent of spatial memory [320]. Therefore, we complement our study with an object placement task, inspired by the relative vector discrimination task proposed by Starrett and Ekstrom [276], that tests participants ability to estimate the location of the platform with less reliance on their egocentric cues by experiencing the virtual environment from a third-person perspective.

After a block of virtual Morris Water Maze trials, participants experience a top-down view of the virtual environment with the platform made visible and placed at the center of the pool. In this task, participants are asked to use the thumb-stick to move the platform to where they think it was during the previous block of trials. Participants confirm their estimation by pressing the controller’s A button.
6.3.3 Experiment Design

We designed a $2 \times 2$ mixed factorial study in which sex and FOV condition were the independent variables. Sex was the between-subjects variable with two levels: Men and Women. FOV condition, the within-subjects variable, also has two levels: no FOV restriction (RN) and dynamically changing FOV (RY). We considered eight dependent variables in this study, four of which measured spatial learning: learning rate, distance traveled, placement error, and placement latency; while the remaining four were the Simulator Sickness Questionnaire (SSQ) [150] scales: Total Severity (TS), Nausea (N), Oculomotor Discomfort (O), and Disorientation (D). Each participant experienced two sessions, one for each FOV condition. To minimize the transfer of VR sickness symptoms across sessions, each session was conducted on a separate day [100]. Each session contained six blocks of trials with each block containing six platform search trials and one object placement task. The order of blocks was changed across sessions to minimize learning effects. Half of the participants started with the RN condition (Group A) while the other half started with the RY condition (Group B). We alternated the assignment of men and women to each group to ensure that both groups had an equal number of men and women.

6.3.4 FOV Test Conditions

Each participant experienced two FOV conditions in this study: the full FOV allowed by the VR headset (RN) and the dynamic FOV restriction (RY). We dedicate the remainder of this section to describe the dynamic restrictor we used in the RY condition.
Similar to Bolas et al. [40] and Fernandes and Feiner [100], we dynamically manipulate the FOV as a function of the participant’s linear and angular speeds using a black texture with a transparent circular cutoff (Figure 6.1). The restrictor narrows down the FOV as a response to linear or angular movements by manipulating the radius of the circular cutoff according to the following formula\(^1\):

\[
FOV_{r,t} = FOV_{r,t-1} \times [1 - (RF_{max} \times \max(\frac{v_t}{v_{max}}, \frac{\omega_t}{\omega_{max}}))] 
\]

(6.1)

where \(FOV_{r,t}\) and \(FOV_{r,t-1}\) are the radii of the circular cutoff at times \(t\) and \(t-1\), respectively. \(RF_{max}\) is the maximum restriction applied to \(FOV_{r,t-1}\) at the peak linear or angular speeds. \(v_t\) and \(v_{max}\) are the current and maximum virtual linear speeds, respectively. \(\omega_t\) and \(\omega_{max}\) are the current and maximum virtual angular speeds, respectively. We empirically chose a value of 0.75 for \(RF_{max}\) to match the tunneling effect used in common VR experiences such as Google Earth VR [19]. We empirically found the values 4.2m/s and 180°/sec suitable for \(v_{max}\) and \(\omega_{max}\), respectively. To make participants less distracted [100], the edges of the restrictor were feathered and FOV restriction was applied gradually over time.

### 6.3.5 Instrumentation

We used the HTC Vive as the VR headset in this study which has a 110° diagonal FOV, 90Hz refresh rate, 2160×1200 pixels of combined resolution, adjustable interpupillary and focal distances, and six-degrees of freedom tracking for position and orientation, respectively. The headset was powered by a computer having a 3.4 AMD Ryzen 7 eight-core processor with 16GB memory and NVIDIA GeForce GTX 1080 Ti graphics running Windows 10. The XBox controller was used for

\(^1\)https://github.com/SixWays/UnityVrTunnelling
omni-directional locomotion using the left thumb-stick and for selection using the A button. We found this controller more suitable for our study as its thumb-stick offers better haptic feedback and granularity of control compared to the Vive trackpad.

6.3.6 Virtual Environment

The virtual environment consists of a 42m-wide circular pool filled with opaque water below which is a hidden $2m \times 2m \times 2m$ cubic platform. When made visible, the platform is elevated 1m above the water surface. The pool is positioned at the center of an open-roof 75m-wide circular arena with a wall that is 8m-high. Placed on the wall of the arena are four 2D images distributed evenly. Each of these images acts as a distal cue that participants can use to remember the position of the platform. Distances were mapped such that 1 Unity unit equals 1m. Participants
experienced the environment at a first-person-view and achieved navigation using the thumb-stick of the XBox controller at a speed that varied between 0 and 4.2m/s. Omni-directional steering was achieved using the same thumb-stick at a direction relevant to the virtual viewpoint’s forward vector. Participants always started the virtual experience from a point at the edge of the pool. To ensure that participants always stay in the pool, the water surface was made 1m below the pool wall. We used Blender to design the virtual environment that was imported to Unity3D, the virtual world generator used to run the study. We used the FOV restrictor developed by SixWays\(^2\) and configured it as explained in Section 6.3.4. Figures 6.1 and 6.2 show first-person and side views of the virtual environment, respectively.

6.3.7 Procedure

The study was conducted in a quiet space void of obstacles. On the first day, upon their arrival, participants were greeted and seated for an orientation at which participants were familiarized with the goal of the study, its duration, collected data, and tasks. Participants were then asked to fill the SSQ questionnaire to get a baseline reading of their VR sickness symptoms. The participants’ interpupillary distance (IPD) was then measured using a ruler and was used to calibrate the VR headset’s IPD. The IPD of the headset used in this study was limited to a value that varied between 60.9mm and 75.4mm. Accordingly, participants’ measured IPD was rounded up or down to the limits of the headset’s IPD when their IPD was outside the range of the headset’s IPD. Participants were then asked to stand at the center of the tracking space to take part in the training session, whose goal was to familiarize participants with the controls, the virtual environment, and the

\(^2\)https://github.com/SixWays/UnityVrTunnelling
tasks. The training session consisted of one block of three platform search trials followed by one placement task. For training purposes, participants were informed that the platform shall be positioned at the center of the virtual pool for all three search trials. To familiarize participants with the unsuccessful search scenario, we asked participants not to move during the first training trial until the deadline has passed, which we set for 20 seconds only for the training session. While they were briefed about the platform search task, participants were encouraged to use the four surrounding landmarks to remember the location of the platform. Participants were also encouraged to strike a balance between estimation time and accuracy while performing the placement task. Each participant performed six blocks, each consisting of six platform search trials and one placement task. Participants filled another SSQ questionnaire at the end of the session. On the second day, participants filled a baseline SSQ questionnaire, performed six experiment blocks, and then filled another SSQ questionnaire. Participants then filled a post-study questionnaire at which they provided their age, sex, frequency of experiencing motion or VR sickness (five-point Likert scale), and their experience with VR (five-point Likert scale). On average, the duration of the study took approximately one hour divided between day one (≈ 40 minutes) and day two (≈ 20 minutes).

6.3.8 Measurements

Virtual Morris Water Maze Measures

We measure participants performance in the virtual Morris Water Maze task in terms of learning rate and distance traveled. We quantify the learning rate as the mean of slopes of the normalized search completion times ($t_n$) across blocks. For
Table 6.1: Quantitative measures of virtual Morris Water Maze, Object placement, and simulator sickness questionnaire in terms of mean (standard deviation).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Women</th>
<th>Men</th>
<th>Total</th>
<th>Women</th>
<th>Men</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Virtual Morris Water Maze</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning rate</td>
<td>-.01 (.1)</td>
<td>-.03 (.1)</td>
<td>-.02 (.1)</td>
<td>-.01 (.1)</td>
<td>-.09 (.1)</td>
<td>-.05 (.1)</td>
</tr>
<tr>
<td>Distance</td>
<td>39.78 (8.0)</td>
<td>38.26 (8.6)</td>
<td>39.02 (8.2)</td>
<td>39.61 (6.2)</td>
<td>37.34 (8.4)</td>
<td>38.48 (7.4)</td>
</tr>
<tr>
<td><strong>Object Placement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>3.04 (.9)</td>
<td>3.38 (1.4)</td>
<td>3.21 (1.2)</td>
<td>2.81 (1.0)</td>
<td>3.02 (1.3)</td>
<td>2.91 (1.2)</td>
</tr>
<tr>
<td>Latency</td>
<td>8.84 (2.4)</td>
<td>10.53 (4.4)</td>
<td>9.69 (3.6)</td>
<td>9.36 (3.8)</td>
<td>10.92 (5.2)</td>
<td>10.14 (4.5)</td>
</tr>
<tr>
<td><strong>Simulator Sickness Questionnaire</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSQ-TS</td>
<td>20.57 (21.4)</td>
<td>13.36 (16.3)</td>
<td>16.96 (19.0)</td>
<td>10.15 (16.6)</td>
<td>11.22 (17.9)</td>
<td>10.69 (16.9)</td>
</tr>
<tr>
<td>SSQ-D</td>
<td>28.83 (35.2)</td>
<td>16.90 (22.6)</td>
<td>22.87 (29.6)</td>
<td>16.90 (26.9)</td>
<td>14.91 (26.5)</td>
<td>15.90 (26.2)</td>
</tr>
<tr>
<td>SSQ-N</td>
<td>20.44 (23.3)</td>
<td>12.95 (19.3)</td>
<td>16.70 (21.4)</td>
<td>10.90 (16.7)</td>
<td>8.86 (15.2)</td>
<td>9.88 (15.7)</td>
</tr>
<tr>
<td>SSQ-O</td>
<td>9.75 (9.6)</td>
<td>7.58 (10.3)</td>
<td>8.66 (9.8)</td>
<td>2.71 (8.2)</td>
<td>7.58 (14.0)</td>
<td>5.14 (11.5)</td>
</tr>
</tbody>
</table>

Each search trial, we calculate $t_n$ as follows:

$$t_n = \frac{t_c}{\vec{IP}}$$  \hspace{1cm} (6.2)

where $t_c$ is the trial completion time and $\vec{IP}$ is the distance between the starting position and the hidden platform position. We perform this normalization to factor out the time needed to travel directly to the platform from the search time. Because no place learning takes place during the first trial of a block, we exclude the first trial from the learning rate calculation. Because we are interested in the performance difference between the beginning and the end of a block, the slope is calculated from the second and last trials. A steep negative slope indicates learning. The distance traveled is measured as the mean distance traveled across all trials.
Object Placement Measures

We measure object placement performance in terms of placement error and placement latency. The placement error is measured as the Euclidean distance between the estimated platform position and its actual position. The placement latency is measured as the elapsed time, in seconds, from the start of the task until the participant presses the controller’s A button.

VR Sickness Measures

We use the collected VR sickness symptoms through the SSQ questionnaire and score them according to the procedure prescribed by Kennedy and Lane [150] to obtain the weighted Total Severity (TS), Nausea (N), Disorientation (D), and Oculomotor discomfort (O) scores. We subtract the scores of the baseline scores of a session from the post-exposure scores of the same session to obtain the relative SSQ scores of the corresponding FOV condition for each participant.

6.4 Results

We report on the analysis of the results of our 28 participants who fully completed our study in this section as per the measurement criteria highlighted in Section 6.3.7. Table 6.1 gives a summary of the results in terms of means and standard deviations.
6.4.1 Virtual Water Maze Task

Spatial performance in the virtual Morris Water Maze task was measured in terms of learning rate and distance traveled. A two-way mixed ANOVA did not detect a significant interaction between sex and FOV condition ($F_{1,26} = 2.41, p = .13$) with respect to learning rate. No significant difference was found between FOV conditions ($F_{1,26} = 1.86, p = .19$) while women had significantly slower learning rate than men ($F_{1,26} = 4.21, p = .05$).

A two-way mixed ANOVA did not find a significant interaction between sex and FOV condition ($F_{1,26} = .05, p = .83$) with respect to distance traveled. No significant main effect of sex ($F_{1,26} = .6, p = .45$) or FOV condition ($F_{1,26} = .10, p = .75$) either. Figure 6.3 shows a summary of the results.

6.4.2 Object Placement Task

We measure participants performance in the object placement task in terms of their estimation accuracy of the platform location and time delay to provide the
Figure 6.4: Left: Object placement accuracy in terms of placement error. Right: Object placement latency. Data in both plots are summarized by sex and FOV condition. RN= No FOV restriction. RY = Dynamic FOV restriction. Error bars represent the standard deviation.

estimation. A two-way mixed ANOVA did not find a significant interaction effect between sex and FOV condition \( (F_{1,26} = .13, p = .72) \) with respect to placement accuracy. No significant effect of sex \( (F_{1,26} = .47, p = .50) \) or FOV condition \( (F_{1,26} = 2.28, p = .14) \) was found either. A two-way mixed ANOVA did not find

Figure 6.5: The relative total severity, nausea, oculomotor discomfort, and disorientation results summarized by sex and FOV condition. RN= No FOV restriction. RY = Dynamic FOV restriction. Error bars represent the standard deviation.

a significant interaction effect between sex or FOV condition \( (F_{1,26} = .01, p = .92) \) with respect to placement latency. No significant effect of sex \( (F_{1,26} = 1.33, p = .26) \) or FOV condition \( (F_{1,26} = .54, p = .47) \) was found either. Figure 6.4 shows a summary of the results.
6.4.3 Simulator Sickness Questionnaire

We report the analysis of the relative SSQ results as per the details in Section 6.3.8. A two-way mixed ANOVA did not find an interaction effect between sex and FOV condition with respect to the TS ($F_{1,26} = 1.93, p = .18$), N ($F_{1,26} = .68, p = .42$), O ($F_{1,26} = 3.08, p = .09$), and D ($F_{1,26} = 1.14, p = .30$) scores. No significant main effect of sex was found on the TS ($F_{1,26} = .25, p = .62$), N ($F_{1,26} = .57, p = .46$), O ($F_{1,26} = .15, p = .70$), and D ($F_{1,26} = .53, p = .47$) scores. FOV restriction resulted in significantly lower TS ($F_{1,26} = 4.4, p < .05$) and N ($F_{1,26} = 4.25, p < .05$) scores while it had no significant effect on the O ($F_{1,26} = 3.08, p = .091$) and D ($F_{1,26} = 2.23, p = .15$) scores. Figure 6.5 shows a summary of the results.

6.5 Discussion and Future Work

Similar to previous studies that explored sex differences using a virtual Morris Water Maze task [29, 230], our study found a significant sex difference in spatial learning, with women performing worse than men. Specifically, our study showed that women had significantly lower learning rate, quantified in terms of improvement in place learning over time, given fixed distal landmarks. This finding is important because unlike earlier virtual Morris Water Maze studies [29, 230] that used desktop VR, ours employed an immersive VR headset, which offers a different spatial information fidelity profile than desktop VR [276]. This indicates the applicability of our findings to today’s consumer VR platforms. Our study, however, did not reveal any significant difference among men and women in the object placement task. Having participants experience the object placement task from a third-person view makes it a spatial task that depends on landmark-based navigation more than
the virtual Morris Water Maze task. As women often use landmark-based navigation more than men do [256], the heavy reliance of the object placement task on spatial inference using landmarks might have made the sex difference narrower. Participants took part in the object placement task after several repetitions of the virtual Morris Water Maze task. This might have given participants enough time to learn the position of the platform by the time they started the object placement task, which might be another reason why no sex difference was found in the object placement task. This is in line with previous studies which showed that environment familiarity through repetitions can make sex differences in spatial abilities fade away [220].

Our analysis did not find a significant effect of FOV restriction on spatial learning among men and women. On the surface, this seems to contradict previous studies which showed that restricting FOV impedes spatial navigation performance of both sexes in general [167, 210, 233] and of women in particular [87, 299]. There are, however, key differences in our study that might explain why we obtained different results. We used a VR headset that has an FOV of 110°. Participants in previous studies, however, experienced VR using a desktop monitor with significantly lower FOV. Moreover, the FOV was fixed throughout the experiment session in earlier studies. The restrictor that we used, however, only gets narrowed as a response to participants linear or angular head movement, giving participants an opportunity to experience the full FOV provided by the headset while they are stationary. These two differences in the degree of FOV and the behavior of the FOV restrictor might have given both men and women a fair chance in performing equally across FOV conditions in our study.

The use of a dynamic FOV restrictor seemed to have a positive effect in reduc-
ing VR sickness symptoms for both sexes. This is in line with a previous study that used a similar FOV restrictor [100]. However, unlike previous reports of higher susceptibility of women to VR sickness [116, 203, 223, 111, 103, 274], we did not find a significant difference among men and women on any of the SSQ scores. One explanation for this contradiction might be related to the nature of the task, which was found to have an effect on the rate of VR sickness incidence in general [194] and among sexes [203]. We also asked participants to perform the study standing-up for them to rotate in VR using their body input. We made this design decision to reduce the reported discomfort from rotation using a controller [100]. An implication of this decision is providing participants with limited proprioceptive and vestibular input which might have reduced the magnitude of sensory conflict, which might have led to low VR sickness scores across sexes.

Out of the 31 participants we recruited in this study, three participants exited due to severe discomfort in the first session. Two of these participants experienced VR with full FOV while the third experienced it with her FOV dynamically restricted. It is interesting to note that all of those who exited were women with very limited experience using VR. Two of these participants also reported having frequent occurrences of motion or VR sickness. Since frequent exposure to VR can reduce the symptoms of VR sickness and past experience with VR sickness [274, 160] can influence the incidence of VR sickness, these two factors might at least partially explain why these participants exited the study.

Seven participants (all women) had their IPD below 60.9mm, which is the minimum IPD of the VR headset we used in this study. We were concerned that such improper IPD calibration due to the headset limitation could have led to eye strain that would, in turn, lead to oculomotor discomfort. We did not find, however,
a significant difference between the Oculomotor Discomfort scores of the seven women in question and the rest of the participants Oculomotor Discomfort scores ($t_{33} = -1.94, p = .062$).

Overall, the outcomes of our study make valuable contributions to what we know about the implications of using common VR sickness prevention mechanisms on the accessibility of VR to women. Our results suggest that dynamic FOV restriction used with high-end consumer VR headsets seems to be effective for both sexes in reducing VR sickness without impeding spatial learning. We would like to point out, however, that our results should not be generalized to all spatial navigation tasks. The spatial tasks we used in this study only targeted participants’ ability to form an effective cognitive spatial map. Different spatial tasks should be used to target other spatial skills like constrained route planning [180]. In future work we would like to employ different spatial tasks to examine the effect of dynamic FOV restriction on spatial skills other than spatial learning. We would also like to study the effect of varying dynamic restriction parameters on VR sickness and spatial navigation performance in both sexes.
CHAPTER 7
CONCLUSION

In this dissertation, the universal usability of VR was presented in terms of two key themes. In the first theme, the gap in user experience quality between high-end and low-end VR platforms was addressed using two low-cost interaction techniques for mobile VR. Using an earbud and a microphone, PAWdio allows users to interact with the virtual environment using hand input gestures, which was shown to increase the immersiveness of the user experience compared to using indirect forms of interaction. With StereoTrack, users can navigate virtual environments by walking naturally using a pair of speakers and a microphone. Despite the limitations of StereoTrack, the ability to integrate it with existing techniques to improve the interactivity of gestural locomotion was promising. Both PAWdio and StereoTrack rely on readily-available hardware. With the millions of mobile VR headsets around the world, these techniques can introduce higher interaction quality for mobile VR at scale.

In the second theme, systematic evaluative research was conducted to investigate the effect of dynamic FOV restriction, a commonly used technique to mitigate VR sickness, on sex bias in terms of VR sickness and spatial navigation performance. Departing from previous studies in VR locomotion, this research aimed to disentangle the different navigation skills that contribute to spatial navigation performance. This was achieved by designing studies that target a specific navigation skill. In this first study, spatial updating was evaluated using a triangle completion task, at which participants ability to produce a new route to their starting position was tested. In the second study, spatial learning was the target navigation skill that was tested using a virtual version of the Morris Water Maze in addition to a point-
ing task. In both studies, dynamic FOV restriction was able to mitigate VR sickness without impeding spatial navigation performance in both men and women. Interestingly, a sex gap in spatial navigation performance was only found in the second study that evaluated spatial learning abilities, which shows the value of the systematic strategy followed in this research that strove to test different spatial navigation skills separately. This systematic analysis of spatial navigation performance is a steady step towards understanding the existence of a sex gap in the user experience of VR. More research is needed to investigate if a gap exists to ensure that the design of virtual experiences does not exclude half of the target user population of VR.

The statement that this research aims to make is that VR is a universal medium that is not, and should not, be restricted to one domain or one group of users. For this vision to be realized, a holistic paradigm that considers users capabilities, the context of use, and technology offerings must be adopted to design virtual experiences that are inclusive to all users irrespective of their differences. The themes presented in this dissertation are two steps towards this vision. Future research should continue to address more opportunities to improve the universal usability of VR. Example opportunities include the development of interaction techniques for people with visual, motor, auditory, and cognitive impairments. Pursuing this research direction will set the stage for VR to be a mainstream medium.
BIBLIOGRAPHY


space (WARP) for distributed multi-user walk-through application. In Proc.

[111] Andre Garcia, Carryl Baldwin, and Matt Dworsky. Gender differences in
simulator sickness in fixed-versus rotating-base driving simulator. In Pro-
ceedings of the Human Factors and Ergonomics Society Annual Meeting, vol-
ume 54, pages 1551–1555. SAGE Publications Sage CA: Los Angeles, CA,
2010.


[114] John F Golding. Motion sickness susceptibility. Autonomic Neuroscience: Basic

[115] Teresia Gowases, Roman Bednarik, and Markku Tukiainen. Gaze vs. mouse
in games: The effects on user experience. Proceedings of the 16th International

Annual Meeting, volume 46, pages 2109–2113. SAGE Publications Sage CA: Los
Angeles, CA, 2002.

using a walking interface. In Proc. RTO Human Factors and Medicine Workshop,
1997.

[118] Timofey Grechkin, Jerald Thomas, Mahdi Azmandian, Mark Bolas, and
Evan Suma. Revisiting detection thresholds for redirected walking: com-
bining translation and curvature gains. In Proc. ACM Symp. on Applied Per-

[119] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. Soundwave:
using the doppler effect to sense gestures. In Proceedings of the SIGCHI Con-


[141] Hiroo Iwata and Takashi Fujii. Virtual perambulator: a novel interface device


[162] Regis Kopper, Tao Ni, Doug A Bowman, and Marcio Pinho. Design and


[222] Charilaos Papadopoulos, Daniel Sugarman, and Arie Kaufmant. Nunav3D:


[263] Stefania Serafin, Niels C Nilsson, Erik Sikstrom, Amalia De Goetzen, and


[274] Kay M Stanney, Kelly S Hale, Isabelina Nahmens, and Robert S Kennedy. What to expect from immersive virtual environment exposure: Influences of


[305] Sam Tregillus, Majed Al Zayer, and Eelke Folmer. Handsfree omnidirectional


