AN EXAMINATION OF NAVIGATION METHODS FOR LARGE IMMERSIVE VIRTUAL ENVIRONMENTS WITH APPLICATION TO THE STUDY OF HUMAN-ROBOT TEAMS

By

Xianshi Xie

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Computer Science

May, 2015

Nashville, TN

Approved:

Robert E. Bodenheimer
Timothy P. McNamara
Julie A. Adams
Amy Shelton
Benoit Dawant

ACKNOWLEDGMENTS

My deep gratitude goes first to my advisor Bobby Bodenheimer. He guided me and shaped my research skills. He encouraged me when I faced stressful situations, especially during the dates when my daughter Elisa was born. He was patient with me when I made little progress on my projects. He enlightened me when I worked on challenging problems.

I'd like to express my deep appreciation to my wife Qiufeng Lin for her unconditional love, encouragement, and support. Without her this thesis would not have been finished.

I'd like to thank Timothy P. McNamara for his thoughtful psychology and statistics discussion. I'd like to thank Julie A. Adams and Amy Shelton for their very insightful suggestions to improve my work. I'd like to thank Benoit Dawant for his helpful big picture advice on my dissertation.

I'd like to thank the former and current members of Learning in Virtual Environments (LIVE) lab, specifically, Betsy Williams, Haojie Wu, and Qiang Fu. You helped me much and I learned much from you. I've been lucky to work with such an extraordinary group.

Last, I'd like to thank my family and my friends for their continual love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Virtual reality (VR) provides people with opportunities to experience places and situations that are different from their physical surroundings. Virtual reality has the potential to help people experience situations that are expensive, difficult, or impossible in the real world for reasons such as high cost or high danger. Virtual reality systems are becoming increasingly important as more applications are developed for them in many areas, such as learning, education, visualization, training, physical therapy, entertainment, and architecture. This dissertation focuses on immersive virtual environments (IVEs), a computer interface that surrounds an individual with sensory information, and allows interaction with a simulated environment (Loomis *et al.*, 1999a). Particularly, this work focuses on the Head-Mounted Display (HMD) based IVEs, because HMD-based VR systems are becoming readily available at the commodity level, for example Oculus Rift (Oculus, 2014), which costs just a few hundred dollars. In a HMD-based IVE, a person wears a portable stereoscopic display that renders the environment in real-time from the viewpoint that their head is oriented at.

In order to fully utilize IVEs, such as for learning and training, users often need to be able to navigate through them, rather than being passively manipulated by them. Navigating in IVEs is known to be challenging (Ruddle, 2001; Péruch *et al.*, 2000). The problem is compounded when the size of an IVE is larger than the physically tracked space which hosts the IVE, because a modified locomotion interface is required when users want to explore the virtual space beyond the boundary of physically tracked space. Previous literature shows bipedal locomotion is desirable for effective navigation because physical translation and rotation are able to provide important cues, such as proprioception and vestibular cues (internal cues, that is, the sense of the relative position of neighboring parts of the body, and balance). Such cues are important in maintaining spatial orientation, i.e., the ability to update one's own spatial location and orientation with respect to the surrounding environment. However, research shows people are able to adapt and recalibrate when there is conflict between the internal cues and visual cues, and that visual cues are dominant over internal cues when there is discrepancy between the two (Rieser *et al.*, 1995; Kuhl *et al.*, 2008). In this dissertation, we will leverage this phenomenon, combining and optimizing previous research done by Williams (2007), to develop a more complete navigation system that allows people to freely explore large IVEs while maintaining their spatial orientation.

However, some users may not have sufficient room for an IVE system to support a locomotion interface, or there may exist situations in which a locomotion is not viable. We also focus on the problem of designing locomotion interfaces for such constrained circumstances. Research shows three information sources are critical to effective spatial updating, i.e., body-based translation, body-based rotation, and visual information (Loomis *et al.*, 1999b). Therefore, in this thesis we investigate the relative contribution of physical translation and physical rotation to spatial navigation. A better understanding of this will give us suggestions on how to design locomotion interfaces for large IVEs and will be helpful to those users and designers of future IVEs.

Having developed and evaluated methods for navigating through large IVEs, we turn our attention to general tasks that involve locomotion and navigation. This dissertation focuses on nascent forms of human-robot teaming systems. Robotic systems were chosen as the scenarios because human-robot teaming is emerging as a viable practice in many domains, such as first response to disaster relief (Casper and Murphy, 2003) and wilderness search and rescue (Humphrey, 2009; Goodrich *et al.*, 2008, 2009). The training of human members of such teams is critical, but it is expensive and difficult to conduct such training in the real world due to several concerns, such as the high cost of purchasing and maintaining large number of robots, the difficulty in deploying large numbers of robots in reasonable spaces, and the difficulty of simulating suitable situations in veridical real world conditions. An IVE may be able to provide robust, challenging, and stressful task scenarios that have salient similarity to the real world conditions, while providing the capacity of a highly flexible robot-team and space scaling. Therefore, an IVE is potentially a good platform to test human-robot teaming methods and algorithms that do not currently exist. Originally, the intent of our research program was to investigate human-robot interaction technology in large IVEs, which may involve cutting-edge HCI technologies. However, it soon became apparent that there were challenging issues of a fundamental nature involving people's ability to locomote and navigate effectively through large IVEs when working with robots that should originally be addressed first. Therefore, I focused on a few of these fundamental issues related to the navigation and locomotion of a human-robot team in an IVE, such as: (1) how humans attend to robot teams in large IVEs when the teams are potentially large and/or distributed; and (2) how humans' ability to maintain spatial orientation and navigate is affected by a large number of robots.

This thesis chooses a search task for the human-robot teaming scenarios. Searching suspicious targets is a typical task in some human-robot teaming scenarios (Humphrey, 2009). In this task, human supervisors are required to supervise a robot team consisting of multiple robots, which potentially may become geographically distributed and separated into multiple groups. Therefore, the human supervisors may have to divide their attention between the robot-groups. McCormick *et al.* (1998) showed that most of the evidence favored a unified model of spatial attention: attention modulation is confined to a single, indivisible focal region in the visual field. So, a question we ask is how human attention is affected in a demanding scenarios where teams are divided. In particular, I would like to determine how human performance, perceived workload (Hart and Stavenland, 1988), and situational awareness (Endsley, 2000) are affected in the presence of large robot teams. Understanding the cognitive costs of the division of large robot teams in the field on a human supervisor has implications for the command and control structure of such teams.

Moreover, there are some situations in which a human supervisor may have to navigate and follow a robot team to work with it. For example, in searching an area for explosives or other dangerous objects, a robot team may flag items that are suspicious for examination by a human. IVEs provide an interesting test-bed for controlled studies of how a human's ability to maintain orientation and navigate during such tasks is impacted due to the presence of large number of robots. There is a considerable literature showing human spatial memory is view dependent (Shelton and Mcnamara, 1997; Diwadkar and McNamara, 1997), and our prior work showed that spatial memory of subjects in a simplified scenario is both view dependent and set-size dependent (see Chapter VI) when people navigate through a space involving a number of randomly placed indicators that they must search and recall the configuration of a subset. We explore how performance is affected when moving robots are added to such a search task. In addition, our work in Chapter VI examines whether performance may depend on an individual's navigation strategy. A better understanding of this phenomenon would have important implications on the design of human-robot teams and their command and control strategies.

I.1 Contribution and Organization

This dissertation uses an interdisciplinary approach. On the one hand, our studies leverage perceptual psychology to validate and improve virtual reality systems, such as the navigation system. On the other hand, our studies use virtual reality systems to study human perception, such as attention, spatial memory, and navigation.

This dissertation makes the following contributions in solving the problem of navigating in large IVEs within a limited physically tracked space, advancing cognitive findings of attention, navigation, and memory, and providing implications for the design and use of future human-robot teaming systems. Specifically, it:

1. **Develops an optimized navigation system that allows people to freely explore large IVEs within a limited physically tracked space.** This thesis extends and optimizes previous work done by Williams (2007) by providing a more complete solution for effective navigation of large IVEs within a limited physically tracked

space while maintaining users' spatial orientation. Therefore, users of future IVEs are able to freely explore large IVEs and make a better use of IVE system for learning and training, instead of just being manipulated by the authors of IVEs. This work is done in Chapter III.

2. **Determines the trade-off between various locomotion interfaces.** Some users of IVEs may not have enough navigable space, or locomotion may be not achievable in some scenarios. Therefore, other interfaces that do not require locomotion may be required. A better understanding of the trade-offs between various locomotion interfaces would be helpful for the IVE users to select proper navigation systems according to their needs and requirements. This work is done in Chapter IV.

3. **Determines how attention is affected in a large IVE when divided into multiple groups of objects, potentially geographically separated.** This thesis advances the cognitive findings of how people divide their attention, especially in demanding scenarios. In addition, the impact of locomotion and occlusion is further investigated. For robotic system scenarios, this work gives suggestions on the design of large human-robot teams, and provides implications for the command and control structure of such teams. This work is done in Chapter V.

4. **Determines how human performance at complex search tasks in a large IVE is affected when arrays of potentially moving indicators are included.** This work advances the cognitive findings of spatial memory and how individual navigation strategy impacts spatial memory and the performance. It also gives us a better understanding of the cognitive limitations and abilities of a human embedded with an array of moving indicators, and provides suggestions on pre-training in such navigation scenarios. Particularly, for human-robot teaming scenarios, it potentially provides suggestions on pre-training of a human supervisor who could be identified by a simple pre-test. Further, it provides guidance on how to train those people. This

work is done in Chapter VI.

The rest of this dissertation is organized as follows. Chapter II provides background information and related study to the present work. Chapter III discusses how to develop an optimized system to explore a virtual environment larger than physically tracked space. Chapter IV examines various locomotion interfaces in cognitive demanding navigation tasks. Chapter V determines how attention is affected in a large IVE when divided into multiple groups of objects, potentially geographically separated. Chapter VI investigated how performance at complex search tasks in a large IVE is affected when multiple moving indicators are included. Chapter VII concludes this work and discusses future directions.

CHAPTER II

BACKGROUND AND RELATED WORK

II.1 Navigation System for Large Immersive Virtual Environments

By their very nature, IVEs admit movement of a user through them. In large IVEs, the amount of space covered will be larger than the physically tracked space of the real-world facility housing the IVE. As discussed previously, the first part of this dissertation examines a method for effectively locomoting through a large virtual environment bipedally. There are two areas of work we will discuss, related to this problem: the first is the difficulty of navigating through IVEs, and the second deals with the locomotion interface, i.e., using bipedal locomotion.

Maintaining one's orientation while locomoting through a virtual environment can be difficult (Ruddle, 2001; Allen and Singer, 1997; Péruch *et al.*, 2000). Evidence shows that effective spatial updating relies on several external cues and internal cues. External cues include optical flow, auditory cues, environmental cues, etc. Internal cues include proprioceptive cues and vestibular cues, which are important to spatial updating (Klatzky *et al.*, 1998). When optical flow cues conflict with such internal cues, optical flow cues dominate over internal cues, and people are able to recalibrate their location and orientation (Rieser *et al.*, 1995; Kuhl *et al.*, 2008).

In virtual environments, systems that allow the exploration of virtual environments that are larger than the tracked physical space have typically manipulated motion cues, and the system used in Chapter III fits into this category as well. It is possible to manipulate optic flow with a haptic device such as a keyboard or joystick (Waller *et al.*, 1998; Ruddle *et al.*, 1999; Bowman *et al.*, 1999). But better spatial orientation is obtained by using bipedal locomotion rather than such devices (Chance *et al.*, 1998; Lathrop and Kaiser, 2002; Williams *et al.*, 2006; Ruddle and Lessels, 2009). A technique, called "walking in place" allows users to explore large virtual environments (Templeman *et al.*, 1999; Slater *et al.*, 1995; Nilsson *et al.*, 2014), but it lacks the same proprioceptive cues of walking. Williams *et al.* (2011) designed a method of walking in place — walking on a Wii balance board to explore a lab-sized virtual room. They compared it with both a joystick method and physical locomotion, and found their method is similar to physical locomotion but better than the joystick method at preserving a user's spatial orientation. Another method is walking on an omni-directional treadmill (Hollerbach *et al.*, 2003; Souman *et al.*, 2008), but such devices have been expensive. Commodity level omni-directional treadmills are becoming available, e.g., the Virtuix Omni (Virtuix, 2014), but how people's spatial cognition behaves in them is unknown. Virtual flying (Usoh *et al.*, 1999) is also an effective way of navigating inside large virtual environments, but it lacks the motor feedback associated with bipedal locomotion. Users do not actually move using the aforementioned technologies.

So, to provide for locomotion in IVEs and help maintain a user's spatial awareness, techniques that modify the optical flow are used. Virtual camera viewpoints can be manipulated so that users are able to walk continuously (with possible interruptions) through larger IVEs. These techniques are generally known as redirected walking techniques (Razzaque *et al.*, 2001; Nitzsche *et al.*, 2004; Razzaque, 2005; Bruder *et al.*, 2009; Peck *et al.*, 2009, 2010; Engel *et al.*, 2008; Steinicke *et al.*, 2008a, 2010). There are several different methods in this category, e.g., manipulating curvature gain, translational gain, or rotational gain.

Manipulating curvature gain (Razzaque *et al.*, 2001; Steinicke *et al.*, 2010; Neth *et al.*, 2012), adds offsets to real world movement when only one kind of movement, translation, or rotation, is tracked. In this method, the virtual camera is manipulated by such a small amount that users do not notice that the motion from the IVE and their physical motion diverge slightly. They unknowingly compensate for the offset, which results in a curved path. For example, when the user walks along a straight path, the camera will iteratively rotate to one side such that the user has to walk along a curved path in physical space in the opposite direction to stay on a straight path in the IVE (Figure II.1). When users turn their heads, an additional rotation of the virtual scene is applied based on the angular velocity of turning. Razzaque *et al.* (2001) tested subjects in a room-sized tracking space. The authors instructed subjects to walk along a 5-waypoint path in an IVE larger than the tracking space, and manipulated the rotational gain of the virtual scene based on the target's location and the linear speed, angular speed, position, and orientation of the subjects, such that subjects walked toward the furthest wall of the tracking space. The results showed that their system caused people to change their real walking direction without noticing it and that the method did not cause appreciable simulation sickness.

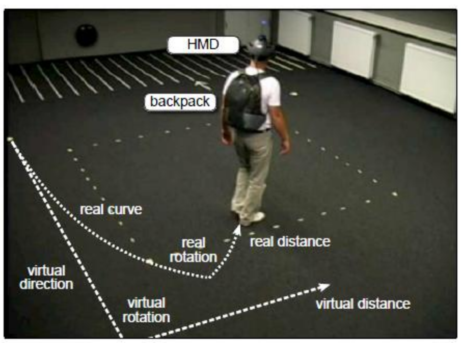

Figure II.1: Redirected walking scenario: a user walks in the real world on a different path in comparison to the perceptual path in the virtual world. Image from (Steinicke *et al.*, 2010).

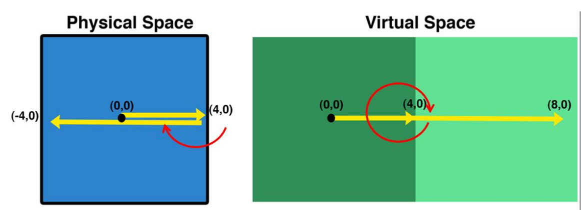

Figure II.2: Description of resetting method: the left picture is of physical space and the right one is of virtual space. When users reach the boundary of physical space (4,0), they turn back by 180 degrees and walk to (-4,0) while making a 360 degrees turn in the IVE and continuing forward.

Since the curvature gain method seeks to manipulate the scene in a subtle and imperceptible way, a large physical tracking space is required. Steinicke *et al.* (2008a) used a tracked space with a 24m radius, and Razzaque (2005) used a tracked space with a 15m radius, which is larger than many virtual environment facilities. It is also difficult to modify redirected walking to permit free exploration, although some methods make inroads here (Engel *et al.*, 2008; Hodgson *et al.*, 2008). Hodgson *et al.* (2008) developed a generalized redirected walking algorithm that allowed spontaneous and unconstrained navigation in a gym-sized tracking space.

Manipulating rotation overtly (Williams *et al.*, 2007; Peck *et al.*, 2009, 2010; Hodgson *et al.*, 2014), such that the locomotion of the subjects fits within the limits of the HMD tracking space, is another method of navigating inside a large virtual environment. The method I develop and examine in Chapter III, called "resetting", allows free exploration in large IVEs (Williams *et al.*, 2007). Resetting is an intervention method of manipulating a user's orientation by scaling the rotation by 2 when the user reaches the boundary of the physical tracking space. It can also be called the 2:1 rotational gain (RG) method. The rotational gain is increased by a factor of two so that a 180-degree turn visually appears to be a 360-degree turn, i.e., a half-turn seems to be a complete turn. Thus a person reaching the boundary of the physically tracked space is able to turn back into the tracking space while in the virtual space it seems that they have simply turned around. Figure II.2 illustrates this algorithm.

Peck *et al.* (2009, 2010) added distractors, i.e., objects in the virtual environment, for the users to focus on while the virtual environment rotates, so that users walk toward the center of the tracking space instead of colliding into the wall. Using distractors has been shown to maintain user's spatial orientation and enhance a user's feeling of presence (Peck *et al.*, 2009).

A third method manipulates the virtual camera in the translational direction, such as the method of scaling the optical flow (Williams *et al.*, 2006; Steinicke *et al.*, 2008b), and "seven league boots" (Interrante *et al.*, 2007). Scaling the optical flow is also called the translational gain (TG) method. It works by multiplying the optical flow experienced by a user as they locomote through an IVE. For example, if the TG is 10, that means that a user experiences walking 10 meters in a IVE by walking 1 meter in physical space. Williams *et al.* (2006) showed that users were able to maintain their spatial orientation when the gain was scaled up to 50. In the "seven league boots" method (Interrante *et al.*, 2007), the authors determined a user's intended direction and scaled the gain of that direction only using a wand control. The authors compared the method with virtual flying, a normal gain of 10, and normal walking, and found participants preferred the "seven league boots" to the other methods.

Steinicke *et al.* (2010) also sought to find the threshold for imperceptible curvature gain, rotational gain, and translational gain, under which users are not able to tell the difference between their real motion and motion in virtual environment. According to those authors, "finding detection thresholds have essential implications for the design of future locomotion user interfaces, which are based on redirected walking," (Steinicke *et al.*, 2010) which may "improve the sense of natural walking." Their results showed users can be turned physically about 49% more or 20% less than the virtual rotation. Users can be manipulated physically by about 14% more or 26% less than the perceived virtual translation. Users can be oriented by 13 degrees to the left or to the right after walking 5m distance, which corresponds to walking along a circular path of 22m radius. This idea, that users are not sensitive to the amount of rotational gain, is a key idea we explore in Chapter III.

Another type of navigation method for experiencing a large IVE manipulates the virtual scene instead of the virtual camera (Bruder *et al.*, 2009; Suma *et al.*, 2011, 2012). Bruder *et al.* (2009) combined redirected walking with virtual portals, which transport the users from one location to another instantaneously, for exploring large spaces. Suma *et al.* (2011) leverage change blindness illusions to reorient users without their noticing the changing of the IVE structure. Suma *et al.* (2012) propose a self-overlapping architecture, called "impossible space", to maximize natural walking in large IVEs, and evaluate the overlapping threshold at which users begin to notice when the IVE structure is compressed.

Most redirected walking methods allow a user to explore a large IVE and maintain the user's spatial awareness, but these methods needs a target ahead of time (Razzaque *et al.*, 2001; Razzaque, 2005; Bruder *et al.*, 2009; Steinicke *et al.*, 2010) or need very large tracking space (Razzaque, 2005; Steinicke *et al.*, 2010; Hodgson *et al.*, 2008). Our focus will be on free exploration of an IVE, and thus I focus on the "resetting" method (Williams *et al.*, 2007) combined with the translational gain method (Williams *et al.*, 2006). By combining and optimizing these two methods, we will have a deeper understanding of how they work together.

II.2 Locomotion Modes

Occasionally, however, the users of virtual reality systems may just have a small physical space in which to move. Therefore, effective alternatives without locomotion should be further investigated. This is the topic of Chapter IV, but here we review the key concepts. The previous section showed that locomotion could be decoupled into physical translation and physical rotation. Research shows both information sources play a critical role in spatial navigation together with visual information (Loomis *et al.*, 1999b). Physical rotation has been considered more important than physical translation, because in experiments in which physical versus imagined turning were compared, physical turning was more salient (Rieser, 1989; Mou and McNamara, 2002).

Compared with the optical flow of locomotion alone, physical rotation adds significant benefits for basic spatial tasks, such as when updating a subject's mental heading after simulated or imaginary navigation through a two-leg triangular path (Klatzky *et al.*, 1998). In this work, Klatzky and colleagues investigated spatial updating of self-position and orientation during real, imagined, and virtual locomotion. They instructed subjects to locomote a two-leg triangular path with a turn, and asked them to point back to the start position as they would if they had walked the path and were at the end of the second leg of path, in five conditions, including real walking, imagined walking from a verbal description, watching another person walk, and experiencing optical flow that simulated walking, with or without a physical turn between the path segments. They found only the two conditions with a physical turn produced correct performance, while the remaining conditions produced overturning by the magnitude of the turn between the path segments, which indicates that subjects encoded the pathway trajectory but failed to update an internal representation of heading.

However, for more complex search tasks, physical rotation and optic flow alone are not sufficient. Ruddle and Lessels (2006, 2009) conducted experiments in which subjects were asked to find eight targets among many objects. Chance *et al.* (1998) asked subjects to remember objects encountered during the traversal of a maze. In both experiments, locomotion by walking was found superior to physical rotation and optic flow under joystick control. The walking interface also caused the least motion sickness. Therefore, the authors suggested the advisability of having subjects explore virtual environments using full-body locomotion in tasks involving spatial orientation. Ruddle *et al.* (2013) also studied how people learn to walk in VR worlds using various interfaces, such as joystick, treadmills, and real working. The authors measured the travel time, number of collisions, and speed-profile, and concluded real walking is the most proficient way to navigate VR worlds.

Riecke *et al.* (2010) performed a similar experiment as Ruddle and Lessels (2009), under what they considered to be more rigorous controls. In this paper, all locomotion conditions were experienced in an HMD (Ruddle and Lessels used a desktop display for one condition), removed environmental cues that may have aided in the search task (Kelly *et al.*, 2008, 2009), and removed the regular grid structure of the searched objects, randomizing it in a two-dimensional Poisson disk. Finally, the mode of locomotion using the joystick was changed from a button-press mode to angle control using the stick that allows continuous linear velocity control. Under these conditions, subjects performed equally well in the walking and physical rotation conditions but performed worse in the visual-only condition, contrasting with Ruddle's results that subjects performed best in the walking condition while worse in the physical rotation and visual-only conditions. The authors conjecture that the existence of environmental cues and features in the Ruddle experiment may obscure potential effects of physical rotation because users were able to orient themselves using such cues both in the visual-only and rotation conditions. They suggest that rotation cues may become more important under cognitively demanding tasks and limited availability of visual (re-)orienting cues. Young *et al.* (2014) replicated Riecke's experiment except in their experiment the subjects only did joystick translation plus physical rotation condition, but in both a commodity-level HMD system (i.e., Oculus Rift) and a standard HMD system (i.e., nVisor SX60, the same HMD used in this thesis). The authors found the commodity-level system outperformed the standard system, in terms of task completion time and total object visits, but did not find a difference in other measures, such as the total number of targets found and the number of revisits. Their results suggest that the Oculus Rift might be a good alternative for a high-cost system, although users may suffer more simulator sickness.

Our plan in this dissertation is to extend the results of the Riecke *et al.* (2010) experiment to a larger space, which will require some form of redirected walking so that users are able to navigate the space beyond the boundary of the physical tracked space.

II.3 Spatial Updating and Spatial Memory

Another key to navigating and wayfinding is a cognizance of where one is and what is around one. How object locations are stored and represented spatially in human memory has been studied by many researchers. In a large-scale environment, people need to know their self-location and self-orientation before finding a route to their destination. Understanding the reference systems that people use in facilitating spatial memory can help us design better interfaces for navigating in large-scale virtual environments. There are two kinds of reference system useful to facilitate spatial memory: egocentric systems (Klatzky, 1998; Wang and Spelke, 2000; Shelton and McNamara, 2001; Wang and Spelke, 2002) and allocentric systems (Klatzky, 1998; Shelton and McNamara, 2001; Mou *et al.*, 2004). Egocentric systems specify the locations and orientations with respect to the observer (e.g., self-centered coordinates, self-to-objects) (Wang and Spelke, 2000, 2002). Allocentric systems specify locations and orientations with respect to the properties of spatial structure of the surrounding environment, independent of the observer (Klatzky, 1998).

As we will see in Chapter VI, I plan to investigate whether spatial memory is view dependent in our task scenarios, e.g., when people explore multiple targets in a near-to-far manner among many objects. Evidence shows people form mental representation of spatial structures in an orientation-dependent manner, especially when they learn the environment from a single or small number of viewing perspectives (Shelton and Mcnamara, 1997; McNamara, 2003; Greenauer and Waller, 2008). When only a small number of viewing perspectives are experienced, people perform better at experienced perspectives but have worse performance from novel viewing perspectives, in terms of accuracy and latency, in various tasks, e.g., judgment of relative directions (i.e., imagine you are standing at position A and facing object B, now point to object C) (Shelton and Mcnamara, 1997; Shelton and McNamara, 2001), object recognition, and scene recognition (Diwadkar and McNamara, 1997; Greenauer and Waller, 2008). After learning from multiple perspectives, either egocentric or imaginary, people can form multiple mental representations of learned perspectives, while spatial memory from novel perspectives requires normalization to the closest and most familiar representations (Shelton and Mcnamara, 1997; Diwadkar and McNamara, 1997; Shelton and McNamara, 2001).

Diwadkar and McNamara (1997) tested viewing dependence in scene recognition. In their study, participants learned the locations of objects on a desktop from a single perspective and then were tested to recognize the arrangement from different perspectives, including familiar and novel views. They found the recognition latency was a linear function of the angular difference between the study view and a test view. They conducted a second experiment to instruct subjects to learn from an egocentric perspective and three other training views. A subsequent recognition test showed both the study view and the three training views were represented in memory, and the recognition latency was a linear function of the angular difference between the test view and the nearest study or training view. They concluded that the inter-object spatial relations are encoded in a view dependent way, and that the recognition of novel views requires normalization to the most similar representation in memory. Thus, if people need to normalize the mental representation of a novel view to a learned view in our task scenarios (as discussed in Chapter VI), we would expect decreased performance when the final viewing angle differs from the learned viewing perspective.

Shelton and Mcnamara (1997) tested viewing dependence in a navigable room-sized real-world environment. In their study, subjects learned the layout of objects from a certain perspective, and subsequently blind-walked to a new viewpoint to learn the layout. After that, subjects were taken to a different room to make a judgment of relative directions of the learned objects. Their results showed that people had the lowest pointing error and shortest latency when the imagined heading aligned with the two learned perspectives. They further concluded that the two views of a spatial layout produced two viewing-dependent mental representations in memory. Participants in the above study did the task in a navigable environment, which is the same case we will use in our task scenarios. In addition, even though people in their study experienced walking from the first viewpoint to the second viewpoint, their spatial memory is view dependent, while people in our task scenario will be teleported to a new viewpoint without locomotion. We would expect their spatial memory to be view dependent as well.

Waller (2006) examined the effects of controlled viewpoint changes on judgments of scene recognition and found similar effects. For example, users responded with lower accuracy and longer latencies when they viewed images of an arrangement of objects taken from a single viewpoint and subsequently were tested to recognize the arrangement from novel viewpoints with lateral or forward-depth translation but with constant orientation. However, when the test-viewpoint is translated laterally, resulting in a centered “canonical” view of the arrangement, people performed equally well as the original viewpoint, which implies this “canonical” view was stored in memory.

As described above, when people use an allocentric system to orient themselves, they interpret and represent the locations of objects according to the properties of spatial structure and surrounding environments, such as the relative location to a salient landmark (McNamara, 2003), the relative location to intrinsic axes or directions formed by the collection of objects (Mou and McNamara, 2002), the proximity to other objects, local and global coordinates formed by surrounding environments, e.g., walls of a room (Shelton and McNamara, 2001), or other environmental cues (Kelly *et al.*, 2008, 2009).

An example of a landmark reference system (McNamara *et al.*, 2003) is the following: people walking along a path that encircles a large rectangular building inside a park tend to have the smallest pointing errors during judgment of relative direction task when their imagined facing direction is parallel to the path or point to a salient landmark (e.g., a lake). However, they may have never experienced the direction to the lake along the path; therefore, this result implies non-egocentric spatial information was coded in memory. People will likely use landmarks to orient themselves in the spatial memory task scenarios discussed in Chapter V and VI.

Mou and McNamara (2002) showed that people were able to interpret and represent a room-sized layout of objects from non-egocentric views aligned with natural intrinsic directions of a set of the objects. In one experiment, the authors instructed people to learn the layout from either a non-egocentric perspective aligned with some natural and salient intrinsic axes (0-180 condition) or an egocentric view misaligned with the intrinsic axes (315-135 condition). Subjects were then taken to another room to do a judgment of relative directions test. The authors found that subjects performed better when their imagined headings were parallel and orthogonal to the intrinsic axes (e.g., a sawtooth pattern) under the 0-180 condition, which implies that people use intrinsic structures as a reference frame. However, in the 135-315 condition, when they were instructed to learn the layout from egocentric views misaligned with intrinsic axes of the layout, they performed best when the imagined heading was parallel to the egocentric learning view. The authors filtered out the potential impact of the local and global cues (e.g., a mat and the room walls) by placing the objects into a round room in another experiment, and found that the sawtooth pattern remained. Greenauer and Waller (2008) noted that a symmetric structure is not necessary for a non-egocentric coding, but the authors did not produce the sawtooth pattern of Mou and McNamara (2002). This finding suggests that people encode spatial representation in non-egocentric intrinsic axes directions. The relation of the intrinsic cues to their structure in our work is complex: environmental cues exhibit symmetry whereas salient cues are randomly distributed (see Chapter VI).

People often use a common global coordinate to localize familiar but unseen targets when they are put into a highly familiar environment, e.g., a virtual city model of a real city (Frankenstein *et al.*, 2009). In this work, subjects were randomly teleported to a location inside a virtual city corresponding to the city of their residence (Tübingen, Germany) and asked to point to familiar locations. People performed best when they were oriented globally north. However, some evidence shows local coordinates are preferentially coded in memory over global coordinates under some conditions (Shelton and McNamara, 2001) in a room-sized environment. This work is noteworthy in that it involved teleporting in a virtual environment. In some of the work in Chapter VI, this thesis employed teleportation.

People also employ environmental cues, including geometric cues and featural cues, to facilitate spatial memory (Cheng, 1986; Kelly *et al.*, 2008, 2009). For example, the shape of a room gives cues to spatial updating. In a virtual environment, Kelly *et al.* (2008) instructed people to walk along some segments of path and to point to the target that they learned in the beginning of the trial in different shapes of rooms, e.g. trapezoidal, rectangular, square, and circular rooms, which have 1-fold, 2-fold, 4-fold, and infinite-fold symmetry, respectively. Kelly and colleagues also manipulated the number of the segments of the path. Their results showed that people performed worse when the number of segments of the path increased in a circular room condition. To find out why people performed equally well in an angular room, the authors conjectured that users used the shape of the room to reorient themselves throughout the task, or used the corner and flat surfaces to improve self-motion perception. The authors then changed the orientation of each wall when participants faced directly away from it, and found the shape-changing room indeed affected people's performance, which implied that the constant shape of the room was important to spatial updating.

Kelly and colleague's other study showed that the quantity and ambiguity of environmental cues are more important than cue types (2009). The procedure was similar to the one described previously (Kelly *et al.*, 2008); however, in this study they manipulated the environmental cue type and cue quantity. They had four conditions: a circular room, a circular room with four evenly distributed identical stripes on the wall, a square room, and a square room without corners. The authors found that males performed equally well when they walked up to a six-segment path in all conditions except the circular room condition, but that females performed worse when the number of path-segments increased in all conditions that only had one type of cue. The authors conjectured that the ambiguity of the cue caused worse performance and constructed two new environments with unambiguous cues, a circular room with four uniquely colored stripes and a kite-shape room without corners. The authors had only females participate this time and found the dis-ambiguity of the cues facilitated performance. Based on this work, the work we do in Chapter VI will involve strong environmental cues, and also examines for (other) individual differences described in the following.

For large navigable environments, some researchers suggest that allocentric maps are built from two-level processes, e.g., Gallistel (1990). One way might be to construct an egocentric representation from an early perceived stage. The second is path integration, the process by which the internal sensory cues (e.g., proprioception and vestibular cues) and external cues (e.g., optical flow, or auditory cues) accumulate. The more locomotion, then the more familiar with the environment a person would be.

Some research also suggests that people interpret and represent spatial structures using both an egocentric reference system and an allocentric reference system (Burgess, 2006;

Waller and Hodgson, 2006; Shelton and McNamara, 2001; Kelly and McNamara, 2008b,a). More reliance on allocentric reference frames depends on the amount of locomotion involved, the number of objects to be remembered, and the size, familiarity, and intrinsic structure of the environment (Burgess, 2006).

Shelton and McNamara (2001) systematically investigated the possible interactions among the reference frames determined by the viewer's perspective (egocentric system) and the external environment (allocentric system). In their study, subjects learned the locations of objects in a room from two room-wall-aligned views and one room-wall-misaligned view, and then they were taken to another room for testing. The results showed people only formed representations from the two aligned views, but the misaligned view appeared not to be stored in memory. Then they manipulated the number of views, and the congruence and existence of external local and global reference frames (e.g., a local frame of a mat, a global frame of the room walls). In the second experiment, subjects learned the layout from a single view, either aligned or misaligned with external reference frames (e.g., mat and room walls). The results showed that the aligned learning view produced a sawtooth pattern across novel headings while the misaligned learning view produced a smooth pattern, which indicates that alignment with the environmental reference systems influenced the inferential processes necessary to access novel views.

The authors expanded their work to address the role of alignment versus misalignment with multiple views. The results showed that only the aligned learning view was represented; there was no evidence in judgments of relative direction that participants even saw the misaligned view. The authors next conducted experiments to investigate the relative importance of local and global reference systems. In a follow-up experiment, subjects learned the layout in the room where the local reference system (e.g., a mat) and the global reference system (e.g., room walls) was not congruent, from a single view either aligned with the mat or with the walls. These results showed that alignment with the local reference system at learning also produced some facilitation during testing on novel headings aligned with the local reference system, while alignment at learning with the global reference system did not produce such a benefit, which indicates that the local reference system may have precedence over the global reference system. The next experiment extended these investigations to multiple views, and the results showed the local and the global reference system were equally important in terms of judgment of relative directions. Finally, the authors investigated how a spatial layout would be mentally represented in the absence of salient environmental reference systems, either from a single view or from multiple views. The results showed that the learning view was represented in memory, and only the first view was represented when multiple views were experienced. In the work of Chapter VI, I am also interested in assessing individual differences in egocentric and allocentric navigation.

People have a preferred reference system for spatial updating (Gramann *et al.*, 2005). As we will see in Chapter VI, I investigate how these different strategies may affect people's performance. Goeke *et al.* (2013) investigated different strategies that people used for virtual path integration, in the yaw and pitch directions. In their online study, about 300 participants watched 24 short videos of virtual passages through a star field with one turn in either the left-right direction or the up-down direction. At the end of the passage they selected one of four homing arrows to indicate the initial start location. Their results showed participants used two distinct strategies to solve the task, indicating they used either an egocentric reference frame or an allocentric reference frame. Those using an egocentric reference frame, so-called "turners", were able to update their mental heading even when they just watched the video but did not update their physical heading, while those using allocentric reference frame, so-called "non-turners", do not update their mental heading during the task. The relative proportion of their test base was around 33% turners, 47% non-turners, and 9% switchers (those would could move flexibly between allocentric and egocentric strategies) in their study. This work is significant to us because it provides a simple test to evaluate a person's preferred navigation strategy.

Much research studies spatial memory and navigation, but less is known when people need to divide their attention between multiple groups of spatial objects in large IVEs or when locomotion is involved. McCormick *et al.* (1998) argues that most of the evidence favors the unified model of spatial attention: attention modulation is confined to a single, indivisible focal region in the visual field, but other researchers argue that observers are able to divide their attention to multiple spatial locations (Adamo *et al.*, 2008). Jans *et al.* (2010) suggested dividing attention might not be easily achieved by a naive human observer, but it is a skill that can be acquired only through training. In this dissertation I will further investigate how people divide their attention on multiple groups of objects, especially when they are placed under demanding task scenarios. Literature shows that people are able to track multiple moving objects (Pylyshyn and Storm, 1988; Tombu and Seiffert, 2008), and the concurrent tasks that demand attention reduce this tracking ability (Tombu and Seiffert, 2008). It is known that self-motion, either active or passive, impairs the ability to track multiple objects (Thomas and Seiffert, 2010). In this dissertation I will further determine how locomotion affect human attention in demanding human-robot teaming task scenarios.

II.4 Human Robot Interaction

As mentioned previously, many of the applications and scenarios in this dissertation were originally motivated for purposes of humans and robots working together in large-scale virtual environments. There is a considerable literature on human-robot teaming and interaction in the real world; humans and robots working cooperatively as a team are increasingly important in many scenarios. In addition to the first response and search and rescue applications mentioned in Chapter I, there are service tasks (Iwamura *et al.*, 2011) in which humans and robots team. We give a review of this literature as it pertains to the situations and scenarios relevant to this dissertation.

Effective teamwork requires training under realistic situations, but due to the expense and difficulty, it is hard to implement large scale training with large robot teams (NBJ, 2006; Humphrey, 2009). In addition, it is expensive to acquire and maintain a large number of robots and the associated large deployment space. Current human robot teams are often limited in size, e.g., one human to one robot or many humans to one robot (Goodrich *et al.*, 2007; Harriott *et al.*, 2011).

However, IVEs may be able to provide a good test-bed for human-robot interaction because they are able to provide robust, challenging, and stressful simulations that are similar to real-world situations. They also allow the number of robots to be scaled to a large number and a robot team to be flexibly deployed in large space. Prior research on human-robot interaction in IVEs has investigated Cave Automatic Virtual Environments (CAVEs) (Odashima *et al.*, 2003; Livatino *et al.*, 2009) and semi-immersive environments (Boudoin *et al.*, 2008). Robots and virtual environments have been combined before into mixed-reality (MR) simulations (Azamasab and Hu, 2007; Chen *et al.*, 2009; Michael *et al.*, 2008; Xie *et al.*, 2012), but the primary use of much of this work has been to visualize the state of the robot or its sensory data (Chen *et al.*, 2006; Anderson and Baltes, 2007; Nishiwaki *et al.*, 2008; Chen *et al.*, 2009). In this dissertation, I will examine scenarios suitable for human-robot teaming that involving a large number of virtual robots. The goal is to gain knowledge for evaluating the effectiveness of large human-robot teaming, which may have important implications for the design of human-robot interaction methods and the command and control structure of robot teams.

As we will see in Chapter V, I will use perceived workload and situational awareness as some of my evaluation metrics because they are commonly accepted human-robot interaction evaluation metrics (Hart and Stavenland, 1988; Endsley, 2000; Steinfeld *et al.*, 2006). Workload and situational awareness play key roles in determining the effectiveness of human-robot interaction (Drury *et al.*, 2003). Endsley defined situational awareness as the "perception of the elements of the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1995). Workload, defined by Hart and Staveland, is "a hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance"

(Hart and Stavenland, 1988).

CHAPTER III

LOCOMOTION SYSTEM

III.1 Introduction

To enable free and effective exploration of large immersive virtual environments within a limited physically tracked space, this chapter continues and extends the previous work done by Williams (2007) by combining and optimizing the translational gain method and resetting method. In Williams *et al.* (2006) the effect of manipulating the motion cue of optic flow, or, as it was called in that paper, scaling the translational gain of a freely locomoting user was investigated. This work was extended and refined in Williams *et al.* (2007). The basic finding there was that users could reliably maintain spatial orientation in a system where optic flow was (non-linearly) scaled by up to a factor of 50. This factor enables the virtual environment to be expanded by a factor of 50.

In Williams *et al.* (2007), a different method, the resetting method, also called the 2:1 method, is used to expand the navigable size of IVEs. In this method, the yaw rotational gain of the user is changed at key moments when the user reaches a boundary of the tracking space. The rotational gain is increased by a factor of two so that a 180° turn visually appears to be a 360° turn, i.e., a half-turn seems to be a complete turn. Thus a person reaching the boundary of the physically tracked space is able to turn back into the tracking space while in the virtual space it seems that they have simply turned around. Results from Williams *et al.* (2007) showed that users could maintain spatial orientation while navigating a path and performing these 2:1 resets. These interventions are consistent with the findings of Jerald *et al.* (2008) on inserting yawing scene motion into virtual environments.

This chapter tests both methods in combination, and thus scales the linear optical flow of a subject and the angular optical flow of a subject, although not simultaneously, by design. The goal of this chapter is to assess if and how the spatial orientation of users is affected by the combination of the methods in tandem, and to determine, more generally, if such a combination of methods is a viable means for exploring large virtual environments. Walking at a normal rate places inherent limitations on the size of virtual environments that could reasonably be explored; scaling the translational gain increases this size. However, even at high gains, the limits of the tracked physical space will inevitably be reached; resets present one method for remedying this, although there appears to be a cognitive cost associated with them. Thus, the combination of the two techniques may allow the construction of a system for free exploration of vast virtual environments.

III.2 Experiment 1

The first experiment was designed to test whether subjects could reasonably maintain spatial orientation when resetting and scaled translational gain were combined, up to scaled gains found reasonable by Williams *et al.* (2007). Additionally, we wanted to assess if there was a cost associated with more resets. Twelve subjects, six males and six females, with ages ranging from 20-45 years old, participated in the experiment. Participants all had normal sight and hearing. Subjects were not familiar with the experiment and the virtual environment, or had not participated in a similar experiment within the last year. Subjects were compensated for their participation at the rate of $10 per hour. The virtual environment was viewed through a full color stereo NVIS (Reston, VA) nVisor SX head mounted display(HMD) with 1280×1024 resolution per eye, manufacturer's specification of a field of view of $60°$diagonally, and a frame rate of 60Hz. All experiments in this thesis used the same devices as Experiment 1.

III.2.1 Method

For this experiment, we asked subjects to remember three objects at the beginning of a trial, then to walk a path, and turn blindly (by making all objects in virtual environment disappear) to face one of those learned objects, without moving. To measure the spatial orientation of the subjects, we recorded the turning error and latency. Turning error was measured as the angle between the direction that should be turned and the actual direction to which the subject turned. Latency was measured by the response time between the time when the target was identified by the subject and the time when the subject completed the task.

In this experiment, we tested scaled translational gains of 1:1, 10:1, and 50:1, because this represented a reasonable sampling of the scales with which subject could remain effectively spatially oriented according to Williams (2007), with 50:1 being a coarse upper bound. Williams tested these scaled translations under two conditions, which she called filtered and unfiltered. The high gain translations were visually noisy, with lots of jitter. Williams (2007) developed a method for reducing the perceptual effects of the jitter, which was caused by unintended head movement. This method set a threshold speed (i.e. 0.5m/s), under which virtual speed is scaled up nonlinearly and slowly, and above which virtual speed is scaled linearly. Williams showed that better spatial updating performance was obtained with the filtered condition than in the unfiltered condition. A filtering function used for the 10:1 gain is shown in Figure III.1. We used these filtering functions in our experiments.

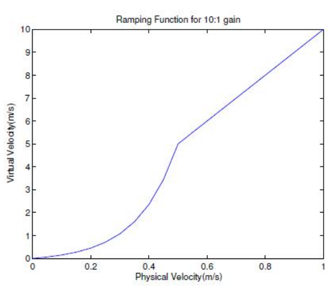

Figure III.1: Experiment 1: This figure shows the ramping function used for the 10:1 gain experiment.

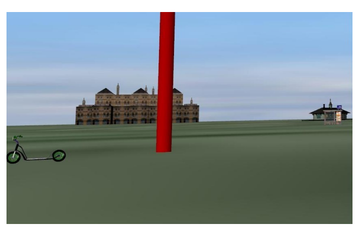

Figure III.2: Experiment 1: An image of the virtual environment showing the red target rod and a sample of target objects.

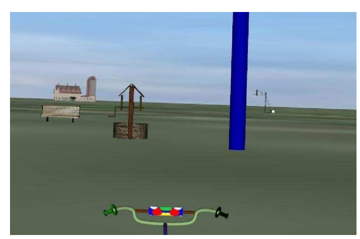

Figure III.3: Experiment 1: An image of the virtual environment showing the blue target rod and a sample of target objects.

During each trial, the subject was told to remember the positions of three specific objects in the environment. Once the subject indicated to the experimenter that he or she had done this, a target red rod appeared in front of the subject, and the subject was told to walk toward it. When the subject reached the rod, it disappeared, and a blue target rod appeared somewhere else in the virtual environment. The subject turned to the blue rod and walked toward it. When the subject reached the blue rod, all objects in the IVE disappeared, and the subject was asked to turn to face one of the three learned objects. The subject's turning error and latency for this trial were recorded.

The travel path consisted of a two segment route: from the start to the red rod, and from the red rod to the blue rod. There were zero, one, or two resets on each segment, symmetrically, so each trial had zero, two, or four resets. To eliminate uncertainty and bias, each gain and reset combination was performed three times. For each of these trials, the turning angle at the red rod to find the blue rod was 90°, 120°, or 150°, respectively, with a left or right turn being randomly selected. Thus each subject performed three gains times three resets times three angles for a total of 27 trials. To counterbalance orders of the reset conditions, each possible order of reset condition was presented to two subjects, one male and one female.

Figure III.4 shows an example of a travel path of a 90°-turn two-segment path in the virtual environment traveled by a user in this study during a two-reset trial. The subject walked to the red rod, and experienced a reset intervention once at position 1 along their path to the rod. Once the subject reached the red rod at position 2, the subject turned 90° to the right to find the blue rod and walked toward it. Along the way to the blue rod the subject experienced a reset at position 3 and then continued to reach the blue rod at position 4. The gray bars represent boundaries of the underlying tracked space. Figure III.5 (A) shows the actual path taken in physical space corresponding to the virtual path in Figure III.4. The subject was reset at position 1 and turned 180° to continue to the red rod positioned at 2; then turned 90° to the right and continued toward the blue rod and was reset at position 3. Figure III.5 (B) shows the physical path taken by the subjects when traveling the virtual path of Figure III.4 but doing two resets along each path segment, for four resets total. Note that in the 1:1 gain condition, the virtual travel distance was the same as physical distance; but when the gain was 10:1 or 50:1, the distance between is magnified by a factor of 10 and 50, respectively.

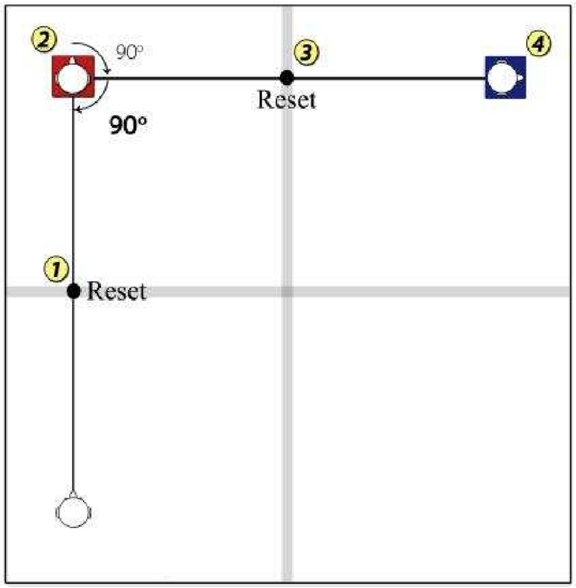

Figure III.4: Experiment 1: This figure shows an example of a travel path of a 90° turn two segment path in the virtual environment. The subject starts the trial positioned in the bottom left corner. Image from Williams *et al.* (2007).

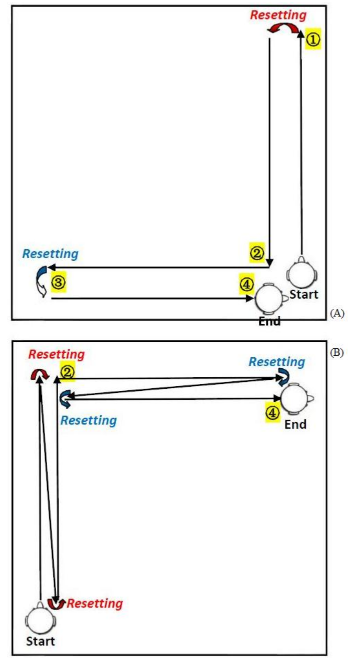

Figure III.5: Experiment 1: Figure (A) shows the physical path taken by the subjects when traveling the virtual path of Figure III.4. Figure (B) shows four resets case.

III.2.2 Results

The mean turning errors and latencies across gains and resets are shown in Figures III.6 and III.7, respectively. We performed a repeated measures analysis of variance (ANOVA) on turning error. The dependent variable, turning error, was measured across the three reset conditions (zero, two, and four resets) by three gain conditions (1:1, 10:1, and 50:1), while order (zero-reset first, two-reset first, or four-reset first) and sex (male, female) were between subjects variables. The only significant main effect we found was that of reset, $F(2,22) = 6.06$, $p = .015$. A post hoc paired-sample t-test with Bonferroni correction

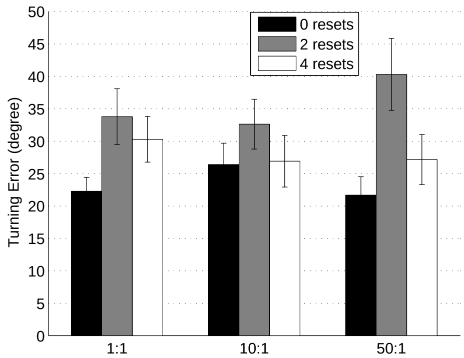

Figure III.6: Experiment 1: Mean turning error as a function of number of resets and translational gain. The error bars show one standard error of the mean.

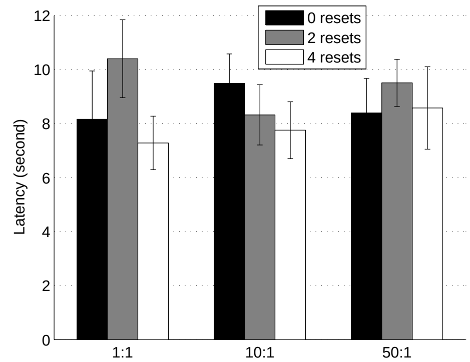

Figure III.7: Experiment 1: Mean latency as a function of number of resets and translational gain. The error bars show one standard error of the mean.

shows the 2-resets (mean=35.6, SD=11.4) had higher errors than the 0-resets condition (mean=23.5, SD=6.6). There were no other main effects or interactions. Note that the baseline mean turning error of about 22° in the 1:1 condition is consistent with errors in the turn-to-point modality in virtual environment experiments, both in our group's prior work (Williams *et al.*, 2006, 2007) and in other groups work doing similar experiments (Chance *et al.*, 1998; Peck *et al.*, 2010).

A similar analysis on the dependent variable latency showed no main effects except sex, $F(1,10) = 7.5$, $p = .03$. Male participants were faster on average (mean=7.0s, SD=2.7) than female participants (mean=10.3s, SD=2.3).

Gains	0 reset	2 resets	4 resets
1:1	33.3%	47.2%	47.2%
10:1	50.0%	69.4%	55.6%
50:1	72.2%	38.9%	50.0%

Table III.1: Experiment 1: Percentage of trials across all subjects in which the subject under-turned, that is, did not turn far enough to reach the target. Each cell consists of 36 trials.

Gains	0 reset	2 resets	4 resets
50:1	69.4%	33.3%	63.9%

Table III.2: Experiment 2: Percentage of trials in the second experiment across all subjects in which the subject under-turned, that is, did not turn far enough to reach the target. Each cell consists of 36 trials.

As noted, we recorded whether subjects under-turned or over-turned in their responses. A systematic bias in response might indicate an interesting side effect of our system on the spatial updating process. Of 36 trials for all 12 subjects in each condition, the under-turning data are shown in Table III.1. A χ^2 analysis on whether participants over-turned or under-turned showed no significance, χ^2(df = 4) = 6.14, p = .18. Participants were as equally likely to under-turn as over-turn.

III.3 Experiment 2

In Experiment 1, two resets had significantly more turning error than zero or four resets. This effect was particularly pronounced in the 50:1 gain condition. We conjectured that this was due, somehow, to the fact that subjects ended up in different physical locations in our lab in the two and four reset conditions. These locations are illustrated in Figure III.5 for the virtual path of Figure III.4 with two and four resets, respectively. The obvious candidate is that there are environmental sound cues that help the person spatially update in the four-reset condition but not in the two-reset condition (although we do not have a theory of how such cues would work), so we reproduced a subset of the conditions of Experiment 1 with subjects wearing noise-masking headphones and obtaining any verbal navigation commands through the headphones, over the noise.

Twelve subjects, 6 male and 6 female, aged 24-28, participated in this experiment. Subjects were not familiar with the experiment and the virtual environment. Subjects were compensated for their participation, \$10 per hour. The experimental procedure and equipment were identical to that of Experiment 1 except in this experiment only the 50:1 gain condition was experienced with zero, two, and four resets. As noted previously, subjects wore earphones that played masking noise and all verbal instructions were given through the earphones.

The mean turning errors and latencies versus reset condition are shown in Figures III.8 and III.9, respectively. Similar to Experiment 1, a repeated measures ANOVA on the turning error in the three reset conditions revealed a main effect of resetting, F(2,22)=12.2, p<.001. A post-hoc paired-sample t-test with Bonferroni correction shows that the 2-resets condition (mean=54.9, SD=21.1) is different from the 0-resets condition (mean=24.6, SD= 12.1), t(11)=3.70, p=0.003; and the 4-resets (mean=51.5, SD=21.1) is different from the 0-resets, t(11)=3.91, p=0.002.

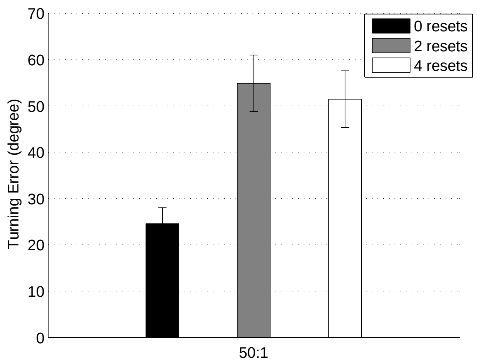

Figure III.8: Experiment 2: Mean turning error as a function of number of resets. The error bars show one standard error of the mean.

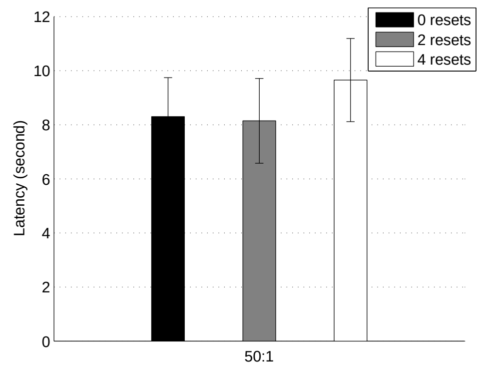

Figure III.9: Experiment 2: Mean latency as a function of number of resets. The error bars show one standard error of the mean.

There was no effect of the reset condition in the repeated measures analysis of participants response latencies. There were no significant effects of sex, or interactions. As in Experiment 1, there was no significant effect based on whether participants under-turned or over-turned (see Table III.2 for details).

We did a between-subjects comparison of the turning error and latencies of participants in the 50:1 gain conditions of Experiments 1 and 2. The analysis of turning error showed that reset was significant, F(2,18)=13.48, p<.001 and that the experiment was significant, F(1,9)=8.4, p=.02. Participants found the 2-resets condition in Experiment 1 the most difficult (mean=35.6, SD=11.4), whereas the 2-resets (mean=54.9, SD=21.1) and the 4-resets (mean=51.5, SD=21.1) conditions were more difficult than the 0-resets condition in Experiment 2.

For latencies, there were no main effects of reset and experiment conditions, and no significant interaction of the two factors. Participants took similar amounts of time across trials in the three different reset conditions, and experimental condition.

III.4 Experiment 3

As we have seen previously, walking is a desirable means of navigating a large virtual environment. In Experiments 1 and 2, we presented a system that combined increased translational gain and resetting to present a system capable of supporting navigation and free exploration in such an IVE. We saw that there were cognitive costs to both resetting and translational gain under various conditions, and that there are limits to scaling the translational gain arbitrarily, i.e., it is not a complete solution. So, while the previous experiment validated a system that combined these two methods (Xie *et al.*, 2010) to provide affordances that allow users to explore an unlimited IVE efficiently, it needs refinement. Translational gain is relatively easy to manipulate, so we do not explore it here. However, we can also reduce the number of resets. Consistent with prior work (Steinicke *et al.*, 2010), we have found in pilot studies that users are relatively insensitive to the manipulation of the rotational gain, which means that during a reset the rotational gain need not be exactly 2:1. We can manipulate the gain to facilitate users turning to the largest open space, e.g., the center of the tracking space, so that users are able to walk more within the tracking space, ideally reducing the total number of resets.

III.4.1 Method

To instruct users to turn to the center of the tracking space, we have to know the turning direction because a left turn and right turn require different rotational gains. For example, in Figure III.10, users experience a 360/α gain when turning left to complete reset, and experience a 360/(360-α) gain when turning right to complete reset. Additionally, this determination of a user's turning direction should be done in real time. We performed this calculation by measuring users' yaw orientation at a frequency of 60 Hz. A positive increment of two consecutive measured yaw angles indicated a right turn; a negative increment indicated a left turn. In our experiment, we did not encounter any false starts to turns that would have confused this algorithm.

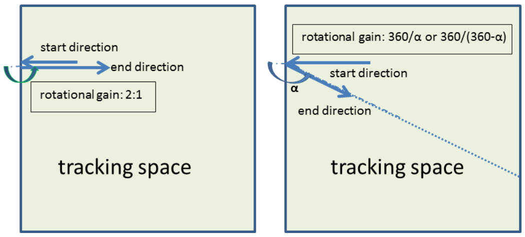

Figure III.10: Experiment 3: This figure shows a case of varying the rotational gain. The left figure is of 2:1 rotational gain, and the right picture of varying gain. For varying rotational gain, users walk toward the center of the tracking space, by which the number of resets may be reduced.

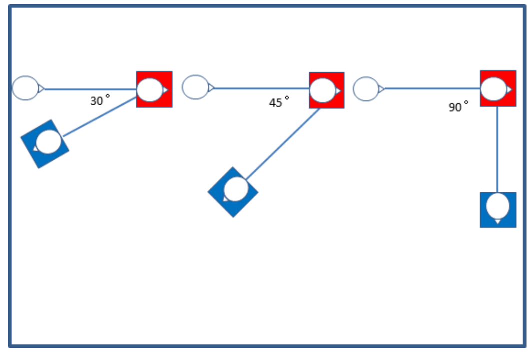

Figure III.11: Experiment 3: This figure shows paths of three different turning angles. Users walk to the red pole first and then turn to walk to the blue pole.

In this experiment, the varying rotational gain was compared with the 2:1 rotational gain method. We varied the number of resets as zero, one, or two, along a two-segment path; we chose these numbers of resets because after a varying reset users will walk a back-and-forth path through the center of physical tracking space and experience 2:1 resets while walking a straight path in IVE.

A mixed design was adopted for this experiment. Twenty-four subjects were used, aged from 18 to 32. Twelve did the Varying Gain condition; twelve did the 2:1 Gain condition, with gender balance in each group. Each subject did all three Number-of-Resets conditions.

The material and equipment were the same as those of Experiments 1 and 2. We used a 10:1 translational gain for this experiment. The layout of the objects were the same as in Experiments 1 and 2. Users wore noise-masking headphones and heard white noise through the whole study. They heard commands from experimenter through the headphones as well.

The procedure was similar to Experiment 1 and 2. Users spent several minutes getting familiar with the locomotion system at the beginning of the study. Once users became familiar with the system, one experimenter led the user to the start positions. Each subject did nine trials, three trials for each Number-of-Resets condition, in a randomized order. Users experienced three different turning-angle (90°, 135°, 150°) paths for each Number-of-Resets condition, shown in Figure III.11. The turning error and latency were recorded.

III.4.2 Results

The mean turning errors and latencies across Type-of-Reset and Number-of-Resets are shown Figure III.12 and III.13.

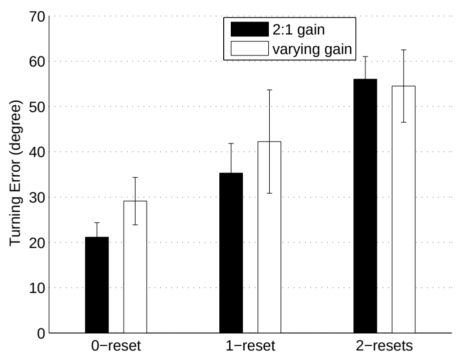

Figure III.12: Experiment 3: Mean turning error as a function of number of resets. The error bar shows one standard error of the mean.

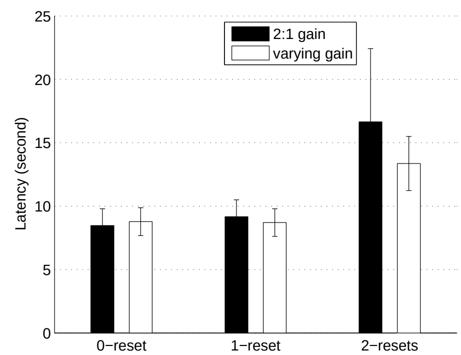

Figure III.13: Experiment 3: Mean latency as a function of number of resets. The error bar shows one standard error of the mean.

In terms of turning error, a mixed ANOVA shows a main effect of Number-of-Resets, F(2,44)=11.03, $p<0.001$, but no main effect of Type-of-Reset or interaction between the two. A post-hoc paired-sample t-test with Bonferroni correction shows difference between the 0-resets (mean=25.1, SD=15.3) and the 2-resets (mean=55.3, SD=22.6) condition, t(23)=6.4, $p<0.0001$. In terms of latency, a mixed ANOVA shows a main effect of Number-of-Resets, F(2,44)=4.2, p=0.02, but no main effect of Type-of-Reset or interaction between the two.

From the results, the 2:1 Gain condition was consistent to that of Experiments 1 and 2, and there continued to be a cost to resetting. We found no difference between the Varying Gain condition and the 2:1 Gain condition in terms of how people performed. However, the varying gain method is also able to reduce the total number of resets over long path lengths compared to the 2:1 method.

III.5 Discussion

Performance in all conditions was better than chance, indicating that subjects could successfully spatially update under combined high translational gains and resets. At the highest gains, subjects were navigating a virtual environment of dimensions 750m $\times$ 750m, quite large. Overall performance did not deteriorate with gain, and this finding is consistent with that of Williams (2007). In addition, the errors in turning seemed to be distributed randomly between overturning and underturning, something that indicates that the system was not causing the subjects to exhibit a biased turning response, as has been noticed in other navigation tasks in virtual environments when subjects were confused (Riecke, 2008).

The two-reset condition was systematically worse than the four-reset condition in Experiment 1, and we devised Experiment 2 to understand if subjects were extracting cues from the environment to help orient themselves. From the results of Experiment 2 it appears that they may be using some cues to orient themselves, although a compelling explanation as to what this mechanism is eludes us. It may be due to the two-reset condition requiring a spatial updating of subject's current spatial location that requires a 180°rotation of earlier representation, whereas the four-reset condition required a 360°turn, and that may be easier to accommodate. However, Experiment 2 shows that there is a cost to resetting in terms of turning errors, and again this finding is consistent with the findings of Williams (2007). Note that there appears to be no combined cost due to the coupling of resetting with translational gain, or at best it is only loosely coupled.

To address the resetting effect, in Experiment 3 we optimized the resetting method by varying the rotational gain to guide users back to the center of the physically tracked space during resetting. In this way, users are able to walk longer distances through the center and thus the the total number of resets may be reduced for long distance walking. Our results did not show any difference between the varying gain method and the 2:1 gain method, in terms of turning error and latency. But since the varying gain method has the potential to reduce the number of resets, this method would be superior to the 2:1 gain method.

III.6 Conclusion

This chapter presented a combination and optimization of two techniques designed to allow users to navigate freely through large immersive virtual environments when the physically tracked space is small. Our results showed users can perform spatial updating using the combined system at better than chance in all conditions, although there is a cognitive cost to resetting. An optimization of resetting to minimize the number of resets seems to be desirable for the combined system. We noticed that users are insensitive to the exact rotational gain of resets, i.e., the rotational gain of resetting does not have to be exactly 2:1, which is consistent with prior work (Kuhl *et al.*, 2008; Steinicke *et al.*, 2010). Therefore, we varied the rotational gain of resetting in real time and redirect users toward the largest open space of the tracked space, such as the center of the tracked space, which is similar to the work of Engel *et al.* (2008). The results showed users are able to maintain their spatial orientation under such variable gain resetting.

However, in the long run, there is still a cognitive cost to resetting, so some approaches to remedy or reduce resetting effect is still desirable. As one moves into free exploration and navigation, an interesting question is how the errors in spatial updating we have measured here translated into performance in more complicated navigational tasks. Ruddle and Lessels (2009) have studied this in the context of body-based cues, but a clearer understanding of this in the context of exploring large virtual environments is needed.

CHAPTER IV

LOCOMOTION MODE

IV.1 Introduction

The previous chapter built a bipedal locomotion system for large IVEs while maintaining users' spatial orientation. However, there are situations where bipedal locomotion as a method of moving through an IVE is not viable, e.g., where the IVEs do not support a tracking interface. This chapter studies an alternative locomotion interfaces and compares that with a bipedal locomotion interface. The goal is to have a better understanding of the trade-offs between various interfaces in different navigation scenarios. However, we necessarily limit our investigation of these interfaces. In particular, we consider locomotion with joysticks, but not using the plethora of other motion control devices that exist in the gaming community, e.g., the Xbox controller, Playstation controller, or Razer Hydra (Young *et al.*, 2014).

Prior work has shown that bipedal locomotion is often desirable to maintain a user's spatial orientation (Chance *et al.*, 1998; Ruddle and Lessels, 2009) because bipedal locomotion is able to provide proprioceptive and vestibular cues that are important. In particular, both physical translation and rotation are important for spatial navigation (Ruddle and Lessels, 2009). In this chapter we extend this study to various complex spatial navigation tasks applied to very large virtual spaces.

IV.2 Experiment 4

IV.2.1 Method

The experimental setup was that of Riecke *et al.* (2010). We compared three locomotion modes: pure joystick (J), joystick translation with physical rotation (JR or joystick rotation condition), and free walking (W). In the joystick condition, subjects used a wireless joystick to achieve translation and rotation in a virtual environment. In the JR condition, subjects physically turned while using the joystick to translate in the virtual environment. In the walking condition, subjects were able to freely navigate inside the virtual environment.

The task was to find eight randomly distributed targets among sixteen randomly distributed locations. In each possible location was a birdhouse, and eight red balls were used as targets and placed inside eight of the sixteen birdhouses. The environment was a featureless plane so that subjects were not able to get any orienting cues from the surrounding environment (Figure IV.5). The sixteen possible locations for targets were randomly distributed so that subjects were not able to get cues from intrinsic reference frames based on the layout of the targets (Mou and McNamara, 2002).

There were three different gains in this experiment, i.e., 1:1, 2:1, and 10:1. Therefore, the sizes of the IVEs were scaled correspondingly. For the 1:1 gain, the sixteen birdhouses were placed within a circle with radius 2m. For the 2:1 gain, the radius of the circle was 4m. For 10:1 gain, the radius of the circle was 20m. We used a mixed design: between-subjects for the different gains and within-subjects for the locomotion modes. Three groups of subjects, 12 subjects for each group, were assigned to the three different translational gain conditions. All subjects did the three locomotion methods, i.e., joystick, joystick plus rotation, and walking condition. In this experiment, each subject did one training trial before each locomotion mode condition that contained three experimental trials; in total, each subject did twelve trials for all three locomotion modes. The locomotion mode order was balanced; the experiment was gender balanced overall.

During the experiment, subjects started in the center of the virtual space. When they approached a birdhouse, subjects clicked the joystick (subjects carried a joystick in the Walking condition also), and the birdhouse became transparent, so that subjects could see whether there was a blue ball inside or not. If the birdhouse was a target birdhouse, a success audio cue would play, the ball would turn red for one second, and then return to its blue color. If the birdhouse was not a target birdhouse, then a blue ball would appear. Thus, if they revisited a birdhouse, subjects could tell they were revisiting the birdhouse from the presence of the blue ball and a revisit audio cue. The task ends when all eight targets have been found or eight consecutive revisits occurred. A message in the upper right hand corner of the screen displayed the current number of targets left to find in a trial. The subjects were asked to complete the task as efficiently as possible, that is to try to minimize the number of revisits and minimize the travel distance and time taken.

We measured the number of times that subjects revisited the same birdhouses that had been visited before, the number of targets found by subjects, the number of targets found before a revisit, the number of targets revisited, the total time spent on the task, the accumulated turning angles, total travel distance, and number of perfect trials.

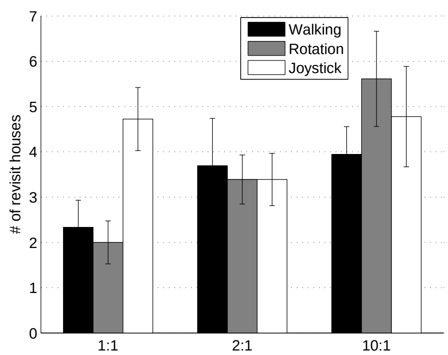

Figure IV.1: Experiment 4: The number of revisits across conditions. Error bars show standard errors of the mean.

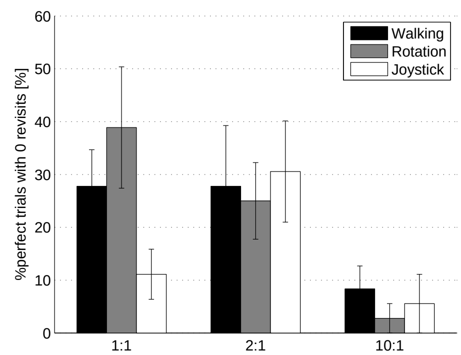

Figure IV.2: Experiment 4: Percentage of perfect search trials across conditions. Error bars show standard errors of the mean.

IV.2.2 Results

The number of revisits in the 1:1 gain condition is highest in the pure joystick condition, which is consistent with Riecke *et al.* (2010), while in a larger virtual space there is no difference across locomotion methods. Regarding the number of revisits, the mixed model ANOVA showed no main effect on locomotion method or gain, but an interaction between locomotion and gain conditions, $F(4,66)=2.8$, $p=0.03$. Examining Figure IV.1 for the interaction, we can see different patterns under different translational gains. In the 1:1

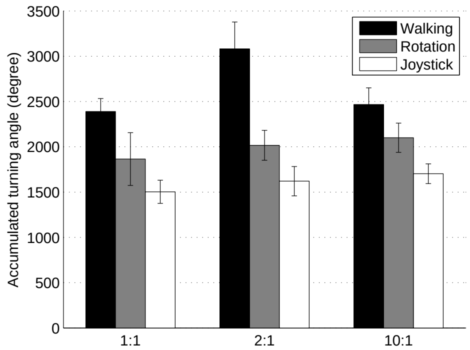

Figure IV.3: Experiment 4: Accumulated turning angle during the search. Error bars show standard errors of the mean.

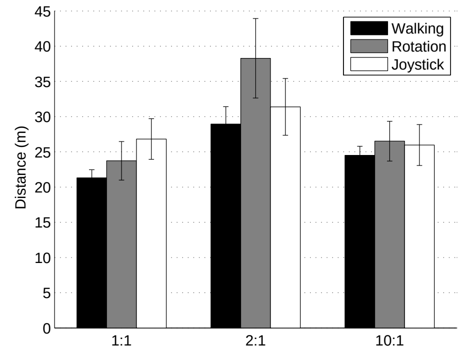

Figure IV.4: Experiment 4: Normalized travel distance across conditions. Error bars show standard errors of the mean.

gain, the joystick condition had worse performance than the walking and rotation conditions; while in the 2:1 and 10:1 gains, there was no such difference. Additionally, trials in the rotation condition seemed to have worse performance when the gain increased. Examining for simple effects, a within-subject ANOVA on the subjects who did the 1:1 gain showed a main effect of locomotion mode, F(2,22)=11.39, p<0.001. A post-hoc paired-sample t-test with Bonferroni correction showed difference between the JR (mean=2.00, SD=1.65) and J conditions (mean=4.72, SD=2.41), t(11)=4.02, p=0.002, and between the J and W conditions (mean=2.33, SD=2.07), t(11)=3.62, p=0.004, which is consistent with the Riecke *et al.* (2010), which showed rotation may suffice for such complex spatial orientation task. However, in the 2:1 gain and 10:1 gains, a within subject ANOVA showed no main effect on locomotion methods. Subjects performed equally well under walking, rotation, and joystick conditions.

Increasing the gain increased the number of revisits under the JR condition, but not under the walking and joystick conditions. A between-subject ANOVA for the rotation condition showed a main effect of gain, F(2,33)=6.129, p=0.005, but no gain effect under the J and W conditions. Under the JR condition, a post-hoc unpaired-sample t-test with Bonferroni correction showed a main effect between the 1:1 gain (mean=2.00, SD=1.65) and 10:1 gain conditions (mean=5.61, SD=3.64), t=3.13, p=0.007. These results indicate that the performance decreased when the gain increased under JR condition, but the performance was not detectably changed under J and W condition.

The 10:1 gain condition has the worst percentage of perfect search trials. Regarding the percentage of perfect trials, a mixed model ANOVA showed a main effect of gain, F(2,33)=4.844, p=0.01. Please see Figure IV.2 (on a scale of 100). A post-hoc unpaired-sample t-test with Bonferroni correction on gain showed difference between the 1:1 gain (mean=25.9, SD=20.4) and 10:1 gains (mean=5.6, SD=10.0), t(16.093) = 3.1169, p=0.007, and between the 2:1 gain (mean=27.8, SD=24.9) and 10:1 gain conditions, t(14.506) = 2.8723, p=0.01. Therefore, the 10:1 gain reduced the number of perfect search trials which had zero revisits. A conclusion one might draw from this is that while a 50:1 translational gain is reasonable for navigation (Williams, 2007), it cannot be used in more challenging tasks.

Walking led to increased orienting motions. For accumulated turning angle (we measured this by recording users' real time yaw orientation and accumulating the difference of every two consecutive orientation records) during the search, a mixed model ANOVA showed a main effect on locomotion method, F(2,66)=36.59, p<0.0001. Please see Figure IV.3. A post-hoc paired-sample t-test with Bonferroni correction showed a difference between the JR (mean=1994°, SD=726) and W (mean=2646°, SD=798) conditions, t(35)=5.02, p<0.001, and between the J (mean=1609°, SD=462) and W conditions, t(35)=7.12, p<0.0001, and between the J and JR conditions, t(35)=3.93, p<0.001. The fact that subjects looked around more during search under walking and rotation conditions indicates that they may employ a qualitatively different navigation strategy: by looking around more subjects were able to optimize the trajectory. This result is somewhat different from Riecke *et al.* (2010) in that participants looked around more only in the walking condition in their study.

No time difference across conditions but the 2:1 gain led to the overall highest

normalized travel distance. Normalized travel distance is calculated by the accumulated optic flow distance over the translational gain. Therefore, the normalized travel distance is the physical walking distance in walking condition, and the optic flow distance over the translational gain the two conditions involving the joystick. A mixed model ANOVA showed a main effect of gain condition, F(2,33)=3.556, p=0.04. Please see Figure IV.4. A post-hoc unpaired-sample t-test on gain showed difference between the 1:1 (mean=23.9m, SD=6.1) and 2:1 (mean=32.9m, SD=12.1) gains, t(16.153)=2.265, p=0.038, and a marginal difference between the 2:1 and 10:1 (mean=25.7m, SD=6.4) gains, t(16.59) = 1.8, p=0.09.

Navigation strategy relied on the size of virtual space. According to the answers to the post-task survey, most subjects typically employed two distinct navigation strategies to complete the task. We adopt here the terms used by Ruddle and Lessels (2009): (a) perimeter (subjects initially checked the birdhouses around the perimeter, and then checked the ones in the center.); and (b) lawnmower (searching in a series of parallel lanes), although, because of the Poisson disk nature of our birdhouse distribution, a strategy only approximating this can be effected. We manually categorized subjects into a perimeter or lawnmower strategy based on their travel paths. In the 1:1 gain, 58% of subjects employed the perimeter strategy, and the other 42% people used the lawnmower strategy. In the 2:1 gain, 50% of people employed perimeter strategy, and 50% people used lawnmower strategy. In the 10:1 gain, 83% of people employed perimeter strategy, and 17% people used lawnmower strategy. Therefore, when the gain was increased to 10:1, there were more people using perimeter strategy. We computed the results by strategy, and found in 2:1 condition those people using the perimeter strategy had fewer visits (mean=4.2, SD=2.2) in the walking (mean=2.7, SD=2.0) and rotation (mean=2.3, SD=0.7) conditions compared with that of joystick condition. However, people using the lawnmower strategy had a higher number of revisits in the walking (mean=4.7, SD=4.7) and rotation (mean=4.4, SD=2.1) conditions compared with that of the joystick condition (mean=2.6, SD=1.5). Therefore, there is a trend that the perimeter strategy facilitated the task in the walking and rotation conditions, while the lawnmower strategy facilitated the task in the joystick condition. Unfortunately, we did not have enough power to obtain statistical significance due to the small number (around six) of subjects in each group.

The walking interface was preferred by most subjects in all three gain conditions. According to the answers to the post-task survey, in the 1:1 gain condition, eight people liked the walking interface best, while five people preferred the joystick rotation and nobody liked the pure joystick interface; in the 2:1 gain condition, eight people preferred the walking interface, while four people liked the joystick rotation and one person liked the pure joystick; in the 10:1 gain condition, six people preferred the walking interface, while three people liked each of the joystick rotation and pure joystick interfaces, respectively. Therefore, the majority preferred the walking interface, and the fewest people liked the pure joystick interface, particularly in a room-sized virtual space with 1:1 gain.

IV.3 Experiment 5

The goal of this experiment is to investigate the relative importance of body-based rotation and translation in a more complex memory and search task scenario. Because the pure joystick (J) condition was no better in any gains in Experiment 4, we only compare the walking condition and the joystick rotation condition (joystick translation plus physical rotation) here. The scenario is similar to how people form spatial memory in complex navigation and search tasks, where it has been shown that spatial memory is view dependent (Shelton and Mcnamara, 1997).

IV.3.1 Method

In this experiment, subjects saw a number (twenty) of trashcans scattered about a virtual plaza. Some of these trashcans contained balls. Trashcans containing balls are called "suspicious" trashcans. The task for the subjects was to memorize the locations of the suspicious trashcans among all trashcans. In particular, subjects searched a few (eight) of the trashcans. Balls were located in some number of these. After searching all eight of the indicated trashcans, subjects were asked to indicate where the suspicious trashcans were. Subjects searched the trashcans sequentially, that is, a trashcan to search was indicated to the subject, and after that trashcan was searched the next was indicated until all eight had been searched. Thus the order in which the trashcans were searched was controlled.

More specifically, subjects started from home position (position 1 in Figure IV.6), and the task started when subjects clicked the trigger of the joystick. The search was conducted in a near-to-far manner. At that time one of the trashcans would turn red. Subjects then approached it. When they were close, subjects clicked the joystick again. The trashcan would momentarily turn transparent, and they would be able to see if a ball was inside the trashcan or not. If a ball was inside, they were to note the location of that (suspicious) trashcan. When they were finished looking inside the trashcan, they looked around to find the next trashcan, which would be red and ready for searching. There were potentially a different number of target balls in the eight trash cans on each trial. The variable number of balls used we called the *set size* condition of the experiment. In this experiment we used set sizes of 3, 5, and 7 balls. The set size condition places different demands on a subject's working memory.

After the search phase was completed, subjects were teleported to a new location from which they would be asked to recall the trashcans that were suspicious. The position to which they were teleported was the *viewing* position and in this experiment there were three different viewing positions. We varied the final view position because prior work has shown that spatial memory is view dependent (Diwadkar and McNamara, 1997). These viewing positions were a 0° view (the original start position, called the 0-view), a 90° view (orthogonal to the main direction of motion, called the 90-view), and a 135° view (at 135° to the main direction of motion, called the 135-view). Refer to Figure IV.6 for further reference. Subjects used the joystick to select the trashcans that they thought contained balls.

We used a within-subject design for this study. Thus, each subject completed both the walking condition and the joystick rotation condition. Within either condition, subjects did nine trials including three set-sizes by the three view positions. Random configurations of trashcans and balls were generated for each trial. In the experiment, the size of the trashcan array is around 50m by 40m, which is much larger than the size of our physical lab. We used a translational gain of 10:1 in this experiment. We want to explore thoroughly how people perform in this gain condition, because prior work has shown that people can perform reasonably well in 10:1 gain (Williams, 2007; Xie *et al.*, 2010).

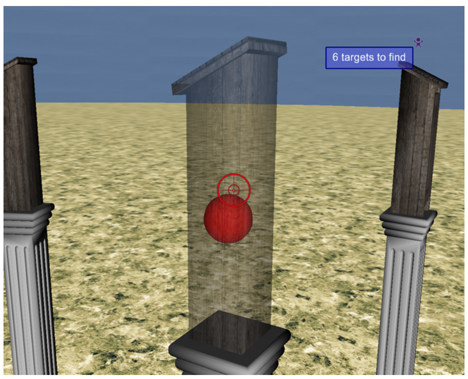

Figure IV.5: Experiment 4: Screen shot of the bird houses and red ball.

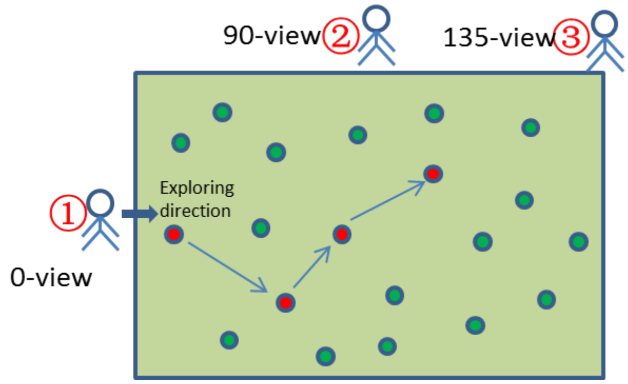

Figure IV.6: Experiment 5: Subjects search in a near-to-far manner. Position 1, 2, and 3 indicate final viewing positions. Position 1 is the original start position.

Eighteen subjects, 9 male and 9 female, aged from 18 to 30, participated the experiments and were paid 15 dollars. Before the actual experiment began, each subject did a few practice trials to make sure they were familiar with the basic environment and procedures.

We measured the correct selection percentage (CSP) of the balls and the time used to make the selection. We also calculated a two dimensional similarity between the correct configurations of the suspicious trashcans and the configurations of subjects' selection, using bi-dimensional regression[1] (Tobler, 1994; Carbon, 2013), which is suitable for a two dimensional configuration similarity comparison. Specifically, we used the Euclidean form of the regression, that transforms one configuration to another through scaling, translation,

[1]The bi-dimensional regression package we used is found in R.

and rotation. For the correspondence of the anchor points of the two configurations, we assumed the correctly selected targets as pairs of points (e.g., we assume subjects made the correct choice intentionally), iterated all possible permutations for incorrectly selected targets, and picked the configuration with highest r^2 (e.g., this measure indicates correspondence between two 2D configurations, ranging from 0 to 1; the higher, the more correspondence) among all permutations.

IV.3.2 Results

For the correct selection percentage (CSP), a three way repeated measures ANOVA shows main effects of locomotion mode (F(1,17)=5.6, p=0.03), set-size (F(2,34)=5.6, p=0.008), and view-angle (F(2,34)=8.6, p=0.001). For locomotion mode, the collapsed mean CSP is 0.62 (SD=0.15) in the walking condition, and 0.54 (SD=0.17) in the joystick rotation condition. For view-angle, the collapsed mean CSP is 0.62 (SD=0.16) in the 0-view condition, 0.60 (SD=0.15) in the 90-view condition, and 0.52 (SD=0.16) in the 135-view condition. A post-hoc paired sample t-test with Bonferroni correction showed difference between the 0-view and 135-view, t(17)=3.5, p=0.003, and between the 90-view and 135-view, t(17)=3.4, p=0.003. For set-size, a post-hoc paired-sample t-test shows difference between the 5-ball (mean=0.53m SD=0.11) and 7-ball (mean=0.64, SD=0.12) conditions, t(17)=5.7, p$<$0.001. Please refer Figure IV.7 for details.

For latency, a three way repeated measures ANOVA shows main effects of set-size, F(2,34)=48.9, p$<$0.0001, and view-angle, F(2,34)=8.8, p=0.001. The results make sense for set-size because subjects have to use longer time to choose more targets. The latency is 26.7s, 36.2s, and 45.4s for the 3-ball, 5-ball, and 7-ball, respectively. For view-angle, the collapsed mean latency is 32.2s (SD=7.2) in the 0-view condition, 36.4s (SD=10.7) in the 90-view condition, 40.0s (SD=10.2) in the 135-view condition. A post-hoc paired sample t-test with Bonferroni correction shows difference between the 0-view and 135-view, t(17)=6.0, p$<$0.001. Please refer Figure IV.8 for details.

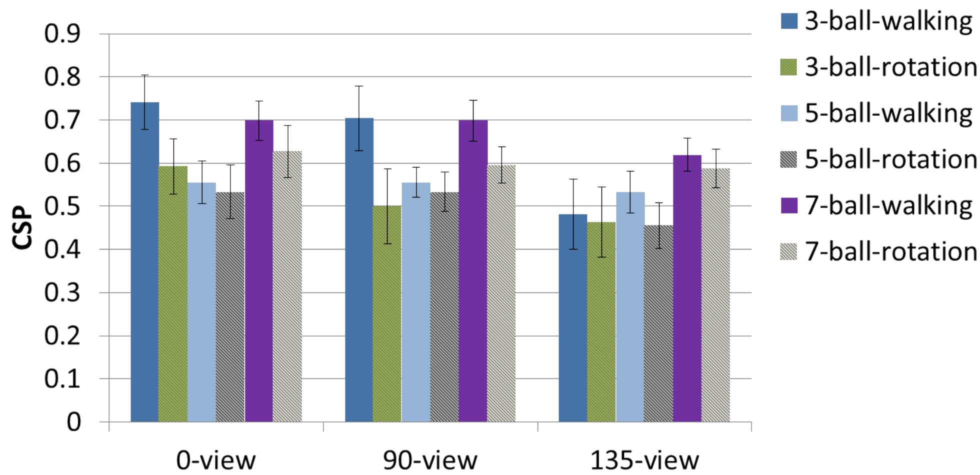

Figure IV.7: Experiment 5: CSP across all conditions. Error bars show standard errors of the mean.

For the Bidimensional regression (BDR) metrics, a three way repeated measures ANOVA for r^2 shows main effects of set-size (F(2,34)=13.5, p<0.001), and view-angle (F(2,34)=4.6, p=0.02). For view-angle, the collapsed mean r^2 is 0.82 (SD=0.09) in the 0-view, 0.83 (SD=0.07) in the 90-view, 0.76 (SD=0.09) in the 135-view. A post-hoc paired sample t-test with Bonferroni correction shows a difference between the 90-view and 135-view, t(17)=3.5, p=0.003. For set-size, the collapsed mean r^2 is 0.86 (SD=0.08) in the 3-ball, 0.80 (SD=0.07) in the 5-ball, and 0.75 (SD=0.09) in the 7-ball. A post-hoc paired sample t-test shows a difference between the 3-ball and 5-ball (t(17)=1.7, p=0.02), 5-ball and 7-ball (t(17)=2.53, p=0.02), 3-ball and 7-ball (t(17)=5.13, p<0.001). Please refer Figure IV.9 for details.

A three way ANOVA for rotation component shows main effects of locomotion mode (F(1,17)=7.6, p=0.01), and view-angle (F(2,34)=7.0, p=0.003). For locomotion mode, the collapsed mean rotation is 10.8 (SD=4.7) in the walking, and 15.5 (SD=7.6) in the joystick rotation. For view-angle, the collapsed mean rotation is 10.5 (SD=6.6) in the 0-view, 11.4 (SD=7.0) in the 90-view, 17.4 (SD=7.7) in the 135-view. A post-hoc paired sample t-test with Bonferroni correction shows a difference between the 90-view and 135-view, t(17)=3.4, p=0.003, and a difference between the 0-view and 135-view, t(17)=3.4, p=0.003. Please refer Figure IV.10 for details.

A three way ANOVA for translation component shows a main effect of locomotion mode, F(1,17)=21.59, p<0.001. The collapsed mean translation is 3.65 (SD=1.29) for the walking condition, and 4.87 for the joystick rotation condition (SD=1.68).

The above BDR results show the walking condition has equivalent r^2 as the joystick rotation condition, but the former has less rotation and translation components than the latter, which indicates people were able to remember the shape of the ball configuration equally well in both conditions, but there were larger angular offsets and linear offsets of the ball configuration in the joystick rotation condition.

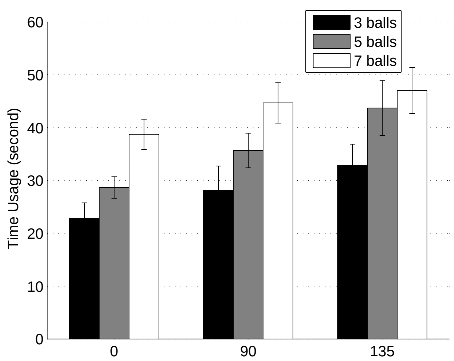

Figure IV.8: Experiment 5: Latency across the conditions. Error bars show standard errors of the mean.

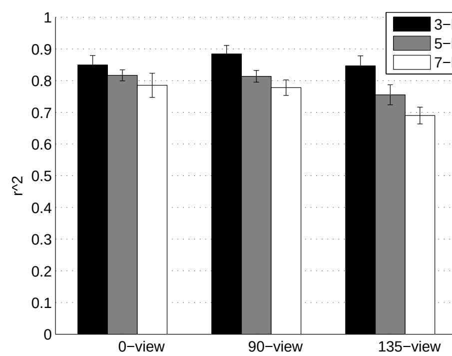

Figure IV.9: Experiment 5: r^2 across the conditions. Error bars show standard errors of the mean.

Subjects also filled out a NASA-TLX (task load index) questionnaire (Hart and Stavenland, 1988) for each locomotion condition. NASA-TLX is a subjective workload assessment tool. It is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six sub-scales: Mental demands, Physical demands, Temporal demands, Own performance, Effort and Frustration. We found no difference among questions between the walking condition and the joystick rotation condition.

From these results, we may conclude subjects performed better with the walking interface than with the joystick interface, which indicates the importance of physical translation

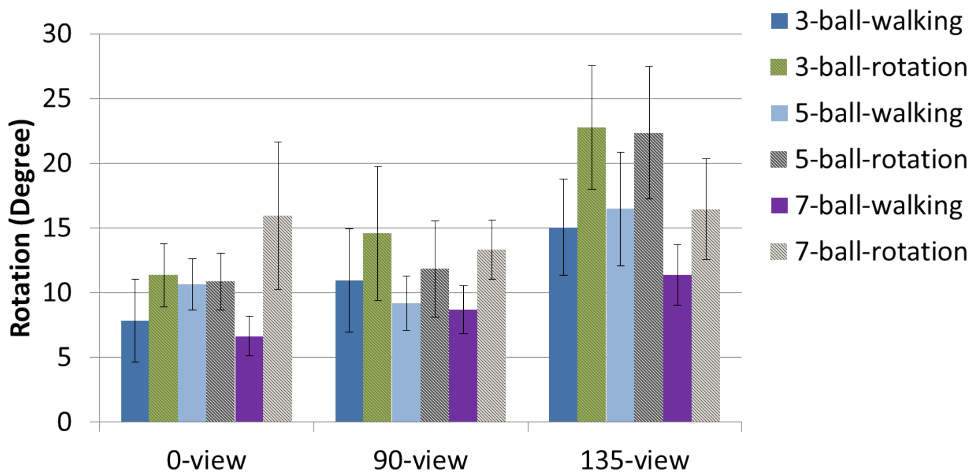

Figure IV.10: Experiment 5: Rotation across the conditions. Error bars show standard errors of the mean.

in spatial navigation, especially in complex memory and search tasks. Subjects also performed in a view dependent way, in all measures, i.e., the CSP, latency, BDR metrics, which is consistent with previous research (Diwadkar and McNamara, 1997). In terms of set-size, the 5-ball condition was worse than the the 7-ball condition in the CSP measure. However, the 5-ball condition was better than the 7-ball condition in the r^2 measure, which indicates that subjects did not remember the exact locations of the five balls but remembered the shape of the configuration in this configuration better than in the 7-ball condition. This pattern is strange, but was robust, as it was noticed in other pilot experiments, and will also be seen in Experiments 9 and 10. It is, in fact, more pronounced in Experiment 9, and we reserve a complete discussion of it until that Experiment, in Chapter VI.4 on page 83.

IV.4 Discussion

The previous two experiments showed that task complexity and the gain of IVE can influence the effectiveness of a locomotion interface. In Experiment 4, where subjects had to find eight targets from 16 randomly placed objects, translational gain and locomotion interfaces interact; people had fewer perfect trials in 10:1 gain, which tends to indicate that for more complicated tasks higher gains are not as good; for orienting motions, the walking is better than the joystick rotation, and the joystick rotation is better than the pure joystick, but it is not clear how this effects task. There are suggestions that navigation strategy changes with size of environment, but nothing conclusive. In Experiment 5, where subjects had to search and remember the locations of a subset of some objects among many distractors, subjects performed better in the walking condition than in the joystick rotation condition. Some subjects reported that they were not able to tell how far they had traveled when only using the joystick to move and thus it was hard for them to remember the path they had been through, thus, it was harder for them to memorize the locations of the targets. Their reports are consistent with some previous research that indicates physical translation is critical to path integration (Ruddle and Lessels, 2009) and spatial navigation. In Experiment 4, the traveled path and orientation data of subjects were recorded; the data showed subjects had more orienting motion and less collision with objects when they walked, consistent with Riecke *et al.* (2010) and Ruddle and Lessels (2009).

IV.5 Conclusion

This chapter presented two experiments and tried to understand better the trade-offs between a walking interface and a joystick interface as users navigated in complex task scenarios through large virtual environments. The results showed task complexity may influence the effectiveness of locomotion interfaces. Particularly, walking was significantly better than joystick translation plus body-based rotation in the scenario of Experiment 5. Therefore, our results may give guidance to users of large IVEs that there is a trade-off between the two locomotion modes. Large physically tracked spaces (e.g., room-sized), are suggested if users want to gain better navigation ability in large IVEs. In this case the physical space needs to be open for users to walk freely, with a motion tracking system. While a joystick interface just requires a small space, no need to walk, and no position tracking, users' spatial performance may have to suffer worse learning experiences in IVEs due to worse navigation ability. In recent work, Young *et al.* (2014) compared a standard HMD system (which is the same as the one that was used in Experiments 4 and 5) and a commodity-level HMD system (i.e., Oculus Rift), only with gain 1:1 and the joystick rotation condition, in a scenario that is the same as Experiment 4 and Riecke *et al.* (2010). Their results showed the low-cost system outperformed the standard system in terms of users' performance measures, but users may have to suffer worse simulator sickness in the low-cost system.

CHAPTER V

ATTENTION DIVISION

V.1 Introduction

Humans working with robots as a team are becoming increasingly important in many areas, as has been previously discussed. IVEs provide a good platform to test human-robot teaming methods and algorithms that do not currently exist. In Chapter III, a locomotion system for large IVEs was built. This chapter will investigate how people behave and respond when they are working with large robot teams in a demanding task scenario in large IVEs. This chapter focuses on attention to spatial detail, human spatial attention, and especially, how people's attention is divided when they are put into demanding task scenarios.

This chapter takes human-robot teaming as a task scenario, in which a human has to supervise a large robot team consisting of multiple robots that are potentially distributed into multiple groups in the field. The goal is to investigate how people's attention is divided when the robot team is geographically separated. In such scenarios, the human supervisor may need to keep track of multiple groups of robots; when the robot team is separated into multiple groups and the human supervisor has to locomote between the groups, we conjecture that a supervisor's attention and performance will be adversely affected by this attention division and the need to locomote. Similar to the previous chapter, we chose our team task to be a search task, that of searching suspicious objects. In this case, the search will be performed automatically. Automatic search through large areas is a complicated problem that could involve a large number of robots. There are at least four types of search strategy, e.g., hasty and heuristic search, constraining search, high probability region search and exhaustive search (Adams *et al.*, 2007). This chapter adopted exhaustive search as the searching scenario.

V.2 Experiment 6

Experiment 6 was designed to test how people divide their attention when they are placed in a demanding task scenario, i.e., supervising a large robot team in a typical search task scenario. In particular, we want to test how people divide their attention when the robot team is deployed in different ways.

For the search scenario we employed an exhaustive search strategy. The robots we modeled have a limited range of communication and sensing. Therefore, our robot team moved and searched side by side within a fixed range so that the robot team would cover all regions of the search area. The robots in the team must be able to avoid the obstacles as well. For ground mobile robots, Latimer *et al.* (2002) investigated the multi-robot coverage problem based on a single robot coverage algorithm, called the boustrophedon approach, which divides the planar area into regions called cells; each cell is covered by simple back-and-forth motions. Then the whole area coverage is achieved when all cells are done. Multi-robot coverage uses the same planar cell-based approach. Multiple robots move side by side and sweep a cell at the same time; robot teams are allocated among different cells. Robots within a team communicate and share information with one another while teams cover cells independent of each other. The advantage for path planning lies in a 2D configuration space for a team of n robots instead of planning in a $2n$ dimensional configuration space. Rekleitis *et al.* (2004) extended this approach by allowing the robots to operate under the restriction that communication between two robots is available only when they are able see each other. In this thesis, we adopted the approach of these authors (Latimer *et al.*, 2002; Rekleitis *et al.*, 2004) for our searching strategy and adapted the general shape of searching space to a rectangular space. For all experiments in this thesis there were only small obstacles which are randomly distributed over the entire area.

V.2.1 Method

We used twelve subjects for this study, six males and six females, aged 18 to 30. No subjects were familiar with the virtual environment or the experiment. All subjects were compensated for their participation, $10 per hour.

This experiment consisted of the following scenario: a robot team, possibly distributed into two groups, searched a set of objects (trash cans) to see if any contained suspicious objects. A human observer was required to memorize the location of the trash cans that the robots designated as suspicious. The conditions of the experiment varied in how the robot team was distributed. In the first condition, the entire robot team acted together as a single unit over a single search area of 100×50m. In the second condition, the robot team was broken into two groups over two search areas, each 50×50m, searching simultaneously, with each area viewable from a single position in which humans observers had to turn their heads approximately 90 degrees. In the third condition, the robot team was broken into two groups over two search areas, each 50×50m, searching simultaneously, but human observers had to locomote to view each area, as neither area was viewable from the view of the other one. These areas were approximately 30m apart, and observers locomoted between them using a scaled translational gain of 15:1. We chose this gain because 15:1 is the lower bound of gain that allowed subjects to walk a 30m distance in the IVE within two seconds, which would allow them to have enough time switch from one robot group to the other. Within each area trash cans are laid out in a random (Poisson disk) manner. As described previously, the robot team conducted an exhaustive search of the area. For example, in the first condition, as shown in Figure V.1, eight robots started from the left half of search area, 5 m apart from each other, moved and searched forward along straight reference paths, with a 0.5 m/s speed, until reaching the boundary of search area. Then the robot team moved to the other half of search area, and did the same search. The robots used simulated laser sensors to detect obstacles along the path, and would pass around the obstacles once one was detected. All other experiments involving robots used the same search strategy. Along the path, when a suspicious object was found, the robot stopped and emitted a beeping sound for about 15 seconds to notify the observer. During this time, the trashcan changed color. In each condition there were six suspicious objects, three for each area in the second and third conditions. A trial consisted of one complete search, and after each trial, subjects were asked to indicate which trash cans were suspicious. The design of the experiment was within subjects, with gender and the order of conditions balanced. Five trials of each condition were experienced, as a block.

In this experiment and all other experiments, the Player/Stage robot framework was used to model, control, and simulate the actual robot behavior, running on an Ubuntu linux server platform (Gerkey *et al.*, 2003; Collett *et al.*, 2005). Player was used to set up robot servers; Stage was used to clone the trashcan environment. Player enabled us to use a built-in motion controller and simulated laser sensor of the robots. Then in VR, we created one robot client (proxy) to control each robot, by leveraging the built-in controller. The communication between robot servers and robot clients were established through a middle layer written in C++.

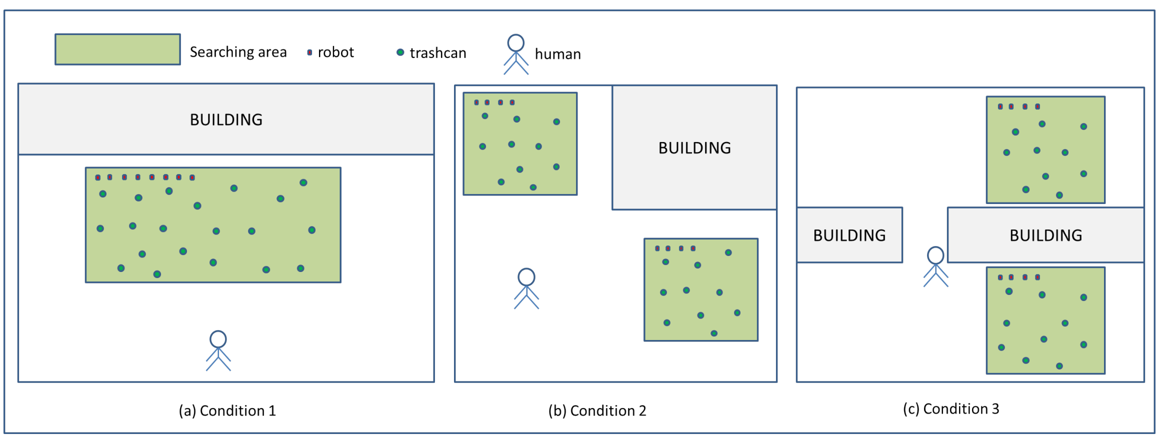

Figure V.1: Experiment 6: Search area layout of all three conditions. Human observer viewed at a distance of about 60m from the center of search area. In Condition 3, the center of two areas is about 90m apart.

We measured the correct selection percentage (CSP) of the trashcans and the latency. We also calculated a two dimensional similarity between the correct configurations of the suspicious trashcans and the configurations of subjects selection, using bi-dimensional regression (Tobler, 1994; Carbon, 2013), as in the previous chapter. For the correspondence of the anchor points of the two configurations, we assumed the correct selected targets as pairs of points (e.g., we assume subjects made the correct choice intentionally), iterated all possible permutations for incorrect selected targets, and picked the configuration with highest r^2 (e.g., this measure indicates correspondence between two 2D configurations, ranged from 0 to 1; the higher, the more correspondence) among all permutations. All subjects also filled out NASA-TLX questionnaires (Hart and Stavenland, 1988) and 3D SART (i.e., Situational Assessment Rating Technique) questionnaires (Humphrey *et al.*, 2007), as described in Chaper II, after completing each condition. In Experiments 7 and 8, subjects completed these questionnaires after each condition, identically to Experiment 6.

V.2.2 Results

Figure V.2 shows that the average ratio of correctly selected trashcans to the total number (i.e., correct selection percentage, CSP) in the first condition (the Standing & 1Group, S1G) is about 0.9, indicating subjects incorrectly selected about 0.6 trashcans on average (less than 1) out of the array of 20 trashcans. In the second and third conditions, (Standing & 2Group, S2G, and Locomotion & 2Group, L2G, respectively) the ratio of correctly selected trashcans to the total is about 0.76 and 0.72, respectively, which means subjects incorrectly selected about 1.5 trashcans on average. A one-way repeated measures ANOVA shows a main effect on these ratios of condition, $F(2,22)=11.33$, $p<0.001$.

A post-hoc analysis using a paired sample t-test with Bonferroni correction, shows a significant difference between the S1G and S2G conditions, $t(11)=3.45$, $p=0.005$, and between the S1G and L2G conditions, $t(11)=3.95$, $p=0.002$. Thus, separating the teams decreases people's performance (by over a factor of 2), but the addition of locomotion does not further affect the performance. There were no significant differences in the amount of time it took subjects to select the configurations of objects.

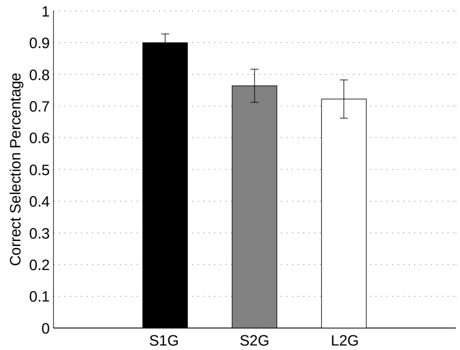

Figure V.2: Experiment 6: The correct selection ratio across conditions. S1G stands for the Standing & 1Group condition; S2G, the Standing & 2Groups condition; and L2G, the Locomotion & 2Groups condition. Error bars show standard errors of the mean.

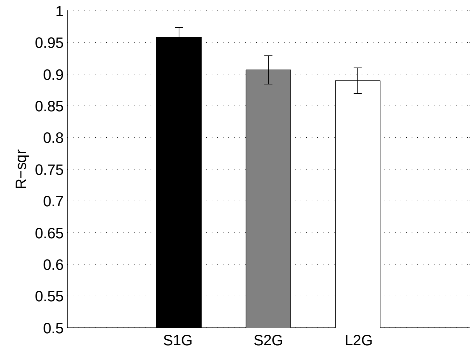

Figure V.3: Experiment 6: The r^2 across conditions. S1G stands for the Standing & 1Group condition; S2G, the Standing & 2Groups condition; and L2G, the Locomotion & 2Groups condition. Error bars show standard errors of the mean.

The r^2 resulting from the bi-dimensional regression analysis, which shows the degree of correspondence between the two configurations ranged from 0 to 1 after translation, scaling, and rotation, was analyzed in a one way repeated measures ANOVA. The results shows a main effect on the conditions, F(2,22)=8.07, p=0.002. A post-hoc paired sample t-test with Bonferroni correction shows a difference between the S1G condition and the S2G condition (t(11)=2.9, p=0.01), and between the S1G condition and the L2G condition (t(11)=4.01, p=0.002). From Figure V.3, the S1G condition has higher r^2 (r^2=0.96) than the S2G (r^2=0.91) and the L2G (r^2=0.89) condition.

In measuring perceived workload, we used the NASA-TLX survey questionnaire (Hart and Stavenland, 1988), administering it at the end of each condition. We see there is an increasing trend of overall workload from Condition 1 (S1G condition) to Condition 3 (L2G condition), which have means 41.09, 51.23, and 67.90 (on a scale of 100), respectively (Figure V.5). An ANOVA shows the three conditions are significantly different (F(2,22)=16.94, $p<0.001$). A post-hoc paired-sample t-test with Bonferroni correction between the S2G and L2G conditions shows a significant difference (t(11)=3.76, p=0.003). This result is ex-

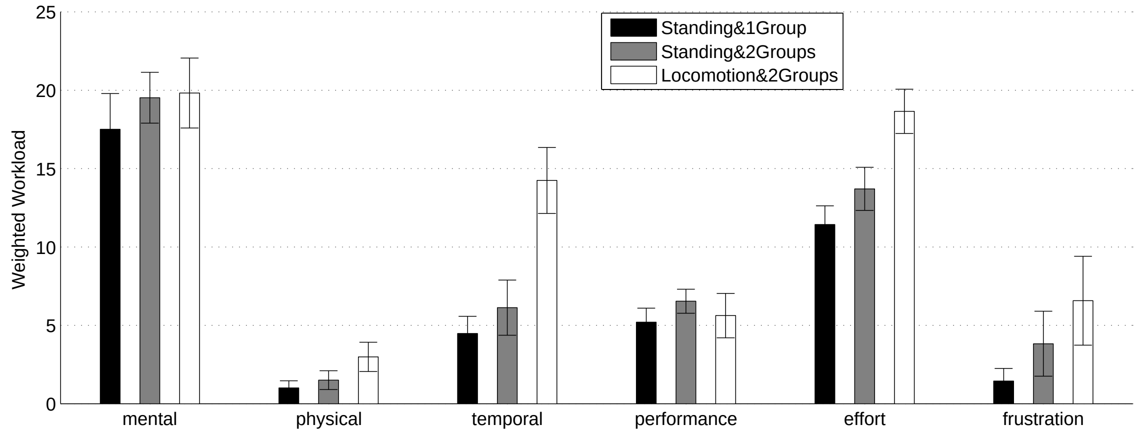

Figure V.4: Experiment 6: Perceived workload across conditions. Error bars show standard errors of the mean.

pected since subjects not only have to divide their attention but also have to walk back and forth to keep track of both robot teams. The overall workload increases by 32% due to locomotion from the S2G to the L2G condition, and this result is significant (a paired-sample t-test with Bonferroni correction, $t(11)=5.8$, $p=0.0001$).

There are six factors that contribute to the overall perceived workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each factor is analyzed using ANOVA analysis across conditions. The results show significance exists in physical demand ($F(2,22)=3.66$, $p=0.04$), temporal demand ($F(2,22)=11.89$, $p<0.001$) and effort ($F(2,22)=9.95$, $p<0.001$). From Figure V.4 we see all these three factors are obviously higher in the L2G condition than the other two conditions.

To assess situational awareness (Figure V.6), we administered a Likert scale questionnaire for demands on attentional resources, supply of attentional resources, understanding of the situation, and overall situational awareness (Humphrey *et al.*, 2007). A within-subjects ANOVA shows that there is a significant effect of demand ($F(2,22)=6.48$, $p=0.006$), but no effect of supply, understanding, or overall situational awareness. A post-hoc paired-sample t-test with Bonferroni correction on the demand factor reveals a significant difference between the S1G and L2G conditions ($t(11)=2.9164$, $p=0.01$).

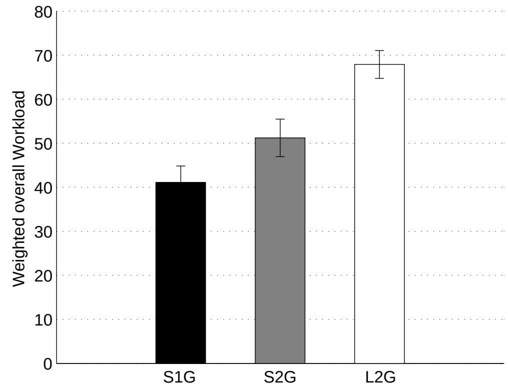

Figure V.5: Experiment 6: Overall perceived workload across conditions. S1G stands for the Standing & 1Group condition; S2G, the Standing & 2Groups condition; and L2G, the Locomotion & 2Groups condition. Error bars show standard errors of the mean.

V.2.3 Analysis

Separating the areas decreased performance (mis-selection almost doubled) and increased the perceived workload. The reason is that subjects have to keep track of both areas simultaneously and have to move their head a lot, which causes more disruption during their memorization process. Note that the total number of robots, trashcans, and suspicious trashcans are identical across the conditions. Another observation is that locomotion did not appear to further decrease the performance but did increase the perceived workload dramatically. We conjecture there are several reasons why the performance was not further decreased. First, the locomotion interval was quite short, i.e., less than a couple of seconds are required to move from one area to the other. Second, the search process of robots last about two minutes, which was long enough for subjects to scan and rehearse the locations of the suspicious trashcans found by the robots. Third, subjects were able to see most of the search area at the standing or observing location, which facilitated their employing a simple strategy of memorizing the suspicious trashcans. For example, some subjects reported they divided the trashcans into groups by the proximity of the trashcans, which helped select suspicious trashcans since subjects stood at roughly the same location with the same viewpoint to choose those trashcans. However, the perceived workload increased from the S2G

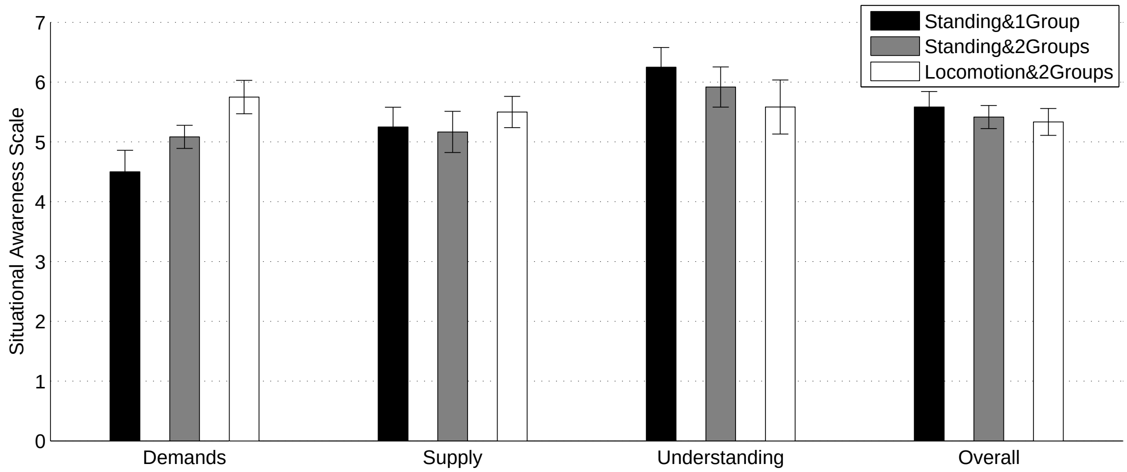

Figure V.6: Experiment 6: Situational awareness across conditions. Error bars show standard errors of the mean.

condition to the L2G condition, especially in temporal demand, effort, and physical demand. The results suggest that subjects have to work harder in the locomotion condition to achieve the same level performance as the S2G condition. The BDR measure is consistent with the CSP measure, which means that the geometry similarity between the selection configuration and the correct configuration is decreased by the robot team separation as well.

V.3 Experiment 7

In the previous experiment, for the S2G condition, the robot team split into two groups searching for two areas and the two groups were located at some distance away (see Figure V.1(b)). Subjects had to turn 90 degrees to see one of the two areas. But in the first condition subjects need only to turn small angles to see the whole search area. Thus two factors were inherently conflated in this condition: the separation of the robot team, which divides attention, and the need to use head movements, which has the entire team out of the field of view at one time. In this experiment, we controlled for these factors independently.

V.3.1 Method

The method is basically the same as that of Experiment 6, with the difference being that we have only two conditions. The first condition is the same as the first condition of Experiment 6. But in the second condition, the two search areas are located side by side, and subjects have the same field of view of the whole search area for both conditions. Thus, the only difference between the two conditions is the robot team splits into two groups in the second condition. Thus, the difference of the S2G condition between Experiment 6 and 7 is that the two areas in Experiment 6 were far apart and subjects had to make a larger head turn angle (around 90 degrees) to see the two areas in the S2G condition, but only make a small head turn angle (around 50 degrees) to see the whole area in Experiment 7. In this experiment, we used 12 subjects, six male and six female. The experiment was order and gender balanced.

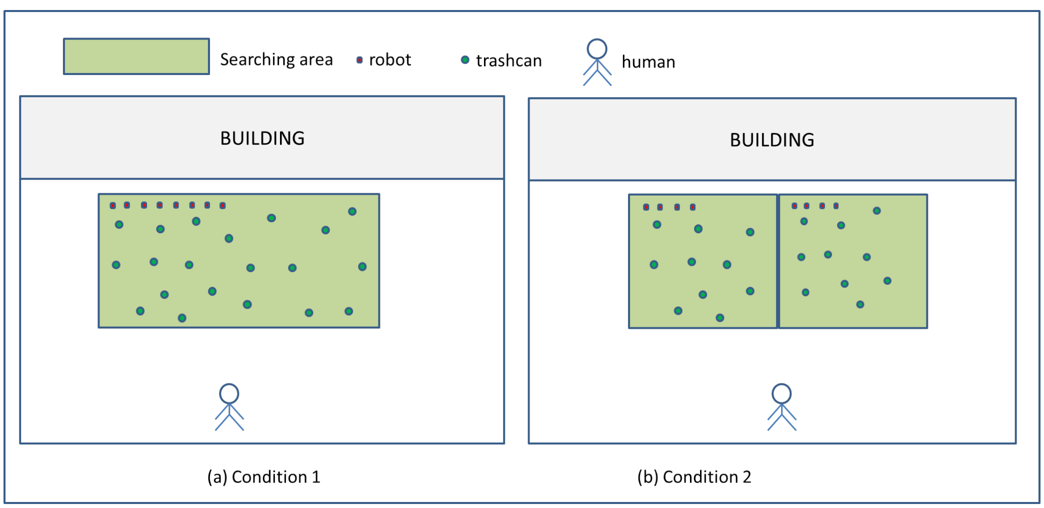

Figure V.7: Experiment 7: The search area layout for the two conditions. The subject viewed the robots from a distance of about 60m from the center of search area(s).

V.3.2 Results

The correct selection percentage (CSP) has a mean of 0.87 in the S1G condition and 0.79 in the S2G condition (Figure V.8), which indicates that subjects mis-selected 0.78 trashcans out of 6 suspicious trashcans in the S1G condition, and mis-selected 1.26 trashcans in the S2G condition. A one way repeated measured ANOVA shows a main effect of the separation, F(1,11)=7.86, p=0.017. Therefore, separating the robot team into two groups, even with the same field of view of whole search area, significantly decreased subjects' performance by mis-selecting about 0.5 more trashcans, around 61% more. There was a gender effect in this experiment, F(1,10)=9.20, p=0.013. Male subjects did better, with a CSP of 0.95 and 0.86 for each condition, respectively, than female subjects, with a CSP of 0.80 and 0.72 for each condition, respectively. It took about 33 seconds on average to complete the selection and there was no significance between the two conditions in duration to complete the task. There was no effect of the order in which the subjects did the conditions.

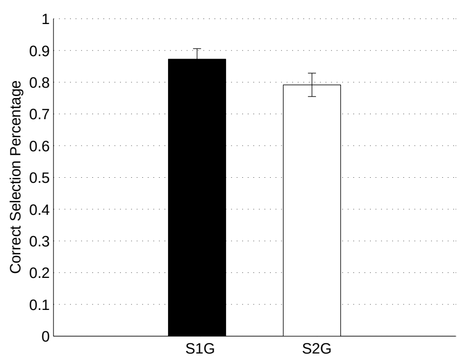

Figure V.8: Experiment 7: The correct selection ratio across conditions. Error bars show standard errors of the mean.

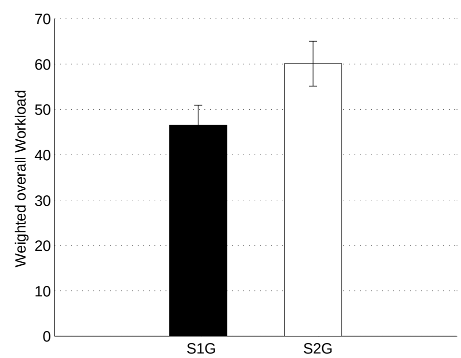

Figure V.9: Experiment 7: Overall perceived workload across conditions. Error bars show standard errors of the mean.

For the NASA-TLX questions, an ANOVA analysis shows a main effect on overall weighted workload, F(1,11)=15.0, p=0.0026, which shows subjects perceived overall workload is higher in the S2G condition than in that of the S1G condition (Figure V.9). For the six factors, we found significance on mental demand, F(1,11)=10.2, p=0.008. Please see Figure V.10 for details.

For situational awareness, there was a main effect of demands on attentional resources, F(1,11)=7.3, p=0.021, which showed that the S2G condition requires much more demand placed on subjects attentional resources by completing the task, than that of the S1G condi-

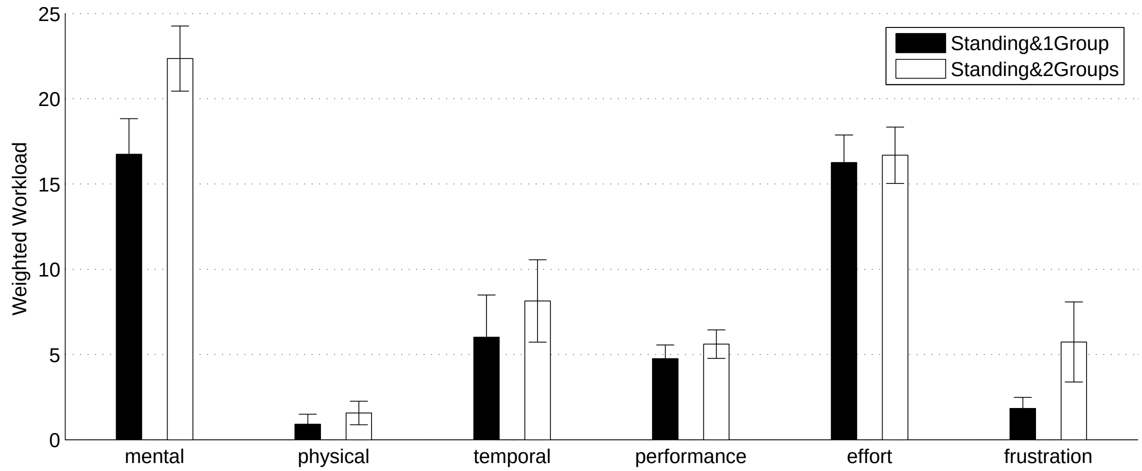

Figure V.10: Experiment 7: Perceived workload components across conditions. Error bars show standard errors of the mean.

tion. Results showed no effect on the factors of supply, understanding or overall situational awareness. Please see Figure V.11 for more details.

We also did a mixed model two-by-two ANOVA for Experiments 6 and 7, Experiment (between) $\times$ Condition (within), in which we dropped the L2G condition of Experiment 6. We find a main effect of condition, $F(1,22)=19.72$, $p<0.001$, but no effect on experiment. Therefore, the separation of the team decreased performance, not the turning angle.

V.4 Experiment 8

In Experiment 6, we found that although there was no performance penalty for locomotion, there was a trend that locomotion between distributed areas affected human performance, increasing the human perceived workload (especially the temporal workload), and increasing the demands of attentional resources. Therefore, we conjecture people will perform worse in a more demanding task. We want to assess the role of locomotion in a more demanding scenario.

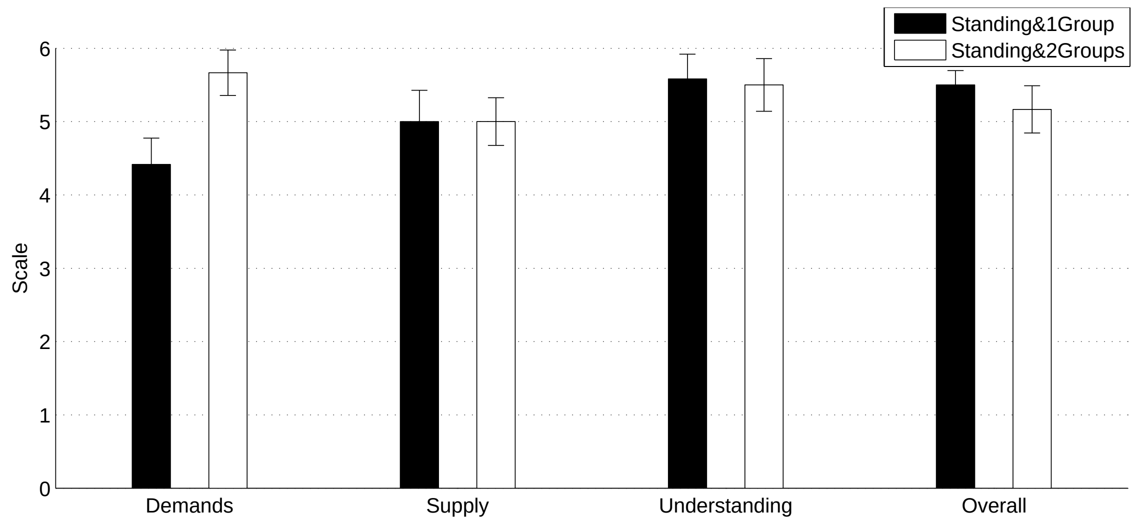

Figure V.11: Experiment 7: Situational awareness measures across conditions. Error bars show standard errors of the mean.

V.4.1 Method

This experiment has a similar experimental setup and procedure to Experiment 6 and 7 with some parameters are changed. In this experiment, we compared various conditions, all with two groups of robots, which are similar to the S2G and L2G conditions in Experiment 6. The differences in this experiment were that the conditions require roughly the same turning angle by the supervisor between the two groups of robots (Figure V.12); the task was designed to be more demanding in that the robots will complete the search in about one minute, while in Experiment 6 the search time is around two minutes. The other factor we decided to test is occlusion between the separated robot groups. When the observer walks between the two areas, if there is a building in between, we need to determine whether any effect was due to locomotion or the presence of a building (occlusion). Therefore, we tested this occlusion effect as well. Twelve participants did the study, six males and six females, aged 19 to 38. All participants got compensation for doing the study, $10 per hour. The materials were the same as Experiments 6 and 7. There were four conditions for the Locomotion and Occlusion factors: Standing w/o Occlusion, Standing w/ Occlusion, Walking w/o Occlusion, and Walking w/ Occlusion. Figure V.12 shows the layout for

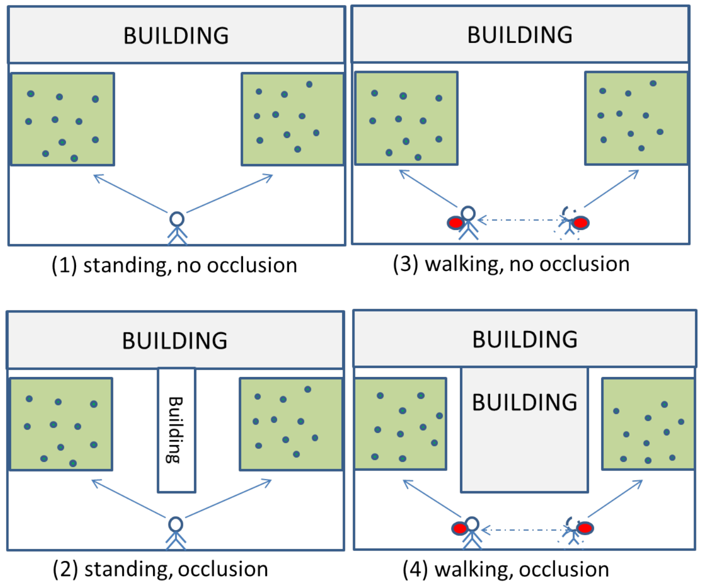

Figure V.12: Experiment 8: The search area layout for the four conditions. In the Walking conditions, subjects were required to walk to either of the two red poles to watch the closer group of robots.

each condition. Each subject did all conditions, four trials in each. The trials were blocked by conditions. Six subjects did the Standing condition first; six subjects did the Walking condition first. In both groups, one-half of the subjects did the Occlusion condition first; one-half did the w/o Occlusion first. In the Standing conditions, both areas are viewable from a single location. In the Walking condition, as shown in Figure V.12, we placed two poles as the stopping and observing spots. Subjects had to walk to a closer pole to see either group of robots or the trashcans. The rest of the procedure was the same as Experiments 6 and 7.

V.4.2 Results

The mean CSP across the conditions is shown in Figure V.13. In terms of CSP, a two-way repeated measures ANOVA shows a main effect of Locomotion, $F(1,11)=7.1$, $p=0.02$; but no main effect on Occlusion or interaction between the two factors. For the Standing condition, the collapsed mean CSP is 0.83; for Walking condition, the collapsed mean CSP is 0.75. There is no difference among the conditions in terms of latency. For r^2, a two-way repeated measures ANOVA did not find any main effects.

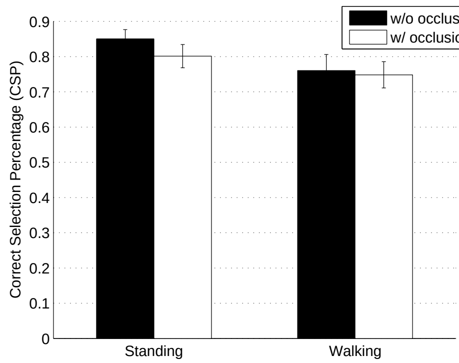

Figure V.13: Experiment 8: CSP across conditions. Error bars show standard errors of the mean.

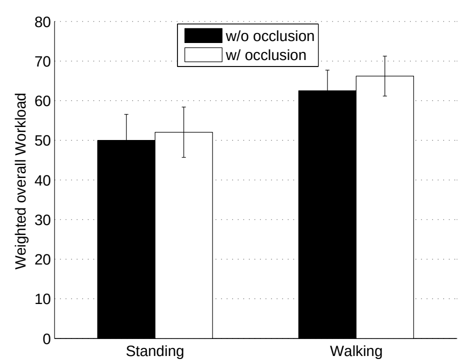

Figure V.14: Experiment 8: Overall perceived workload across conditions. Error bars show standard errors of the mean.

We analyzed situational awareness using a two-way repeated measures ANOVA for the factors demand, supply, understanding, and overall situational awareness. For Demand, the results shows a main effect of Occlusion, $F(1,11)=8.25$, $p=0.02$; the w/ Occlusion condition (mean=5.6, SD=1.5) required more demand than the w/o Occlusion condition (mean=5.1, SD=1.7). The mean demand across conditions is shown in Figure V.15.

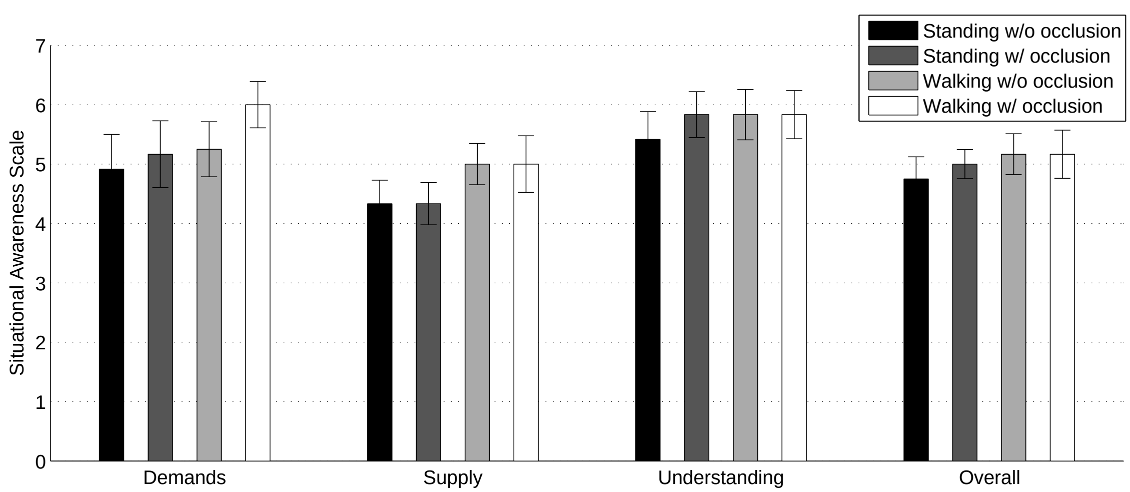

Figure V.15: Experiment 8: Situational awareness measures across conditions. Error bars show standard errors of the mean.

For perceived workload, a two-way repeated measures ANOVA found a main effect of

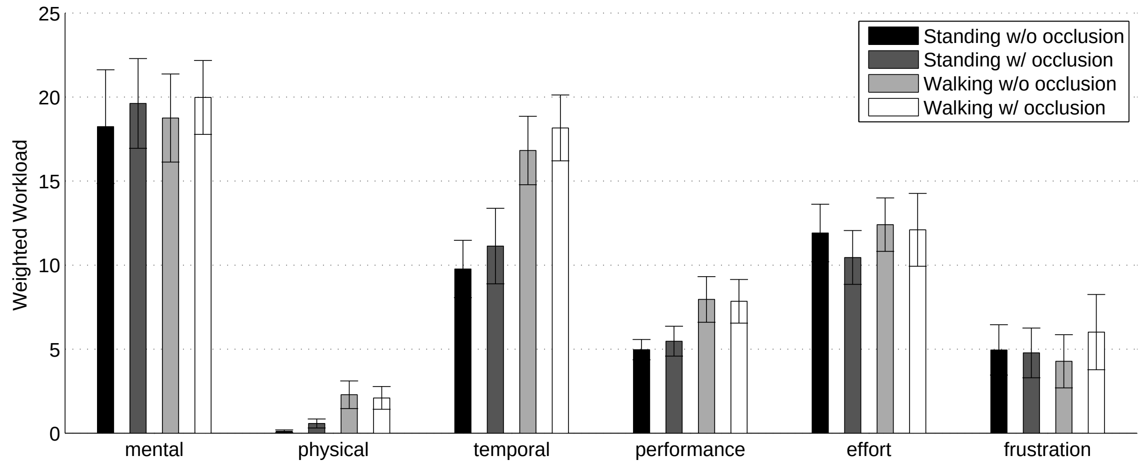

Figure V.16: Experiment 8: Perceived workload components across conditions. Error bars show standard errors of the mean.

Locomotion for overall workload (Figure V.14), F(1,11)=16.1, p=0.002; the Walking condition (mean=64.4, SD=17) required more perceived workload than the Standing condition (mean=51.0, SD=22). We also found a main effect of Locomotion for Physical workload (Walking (mean=2.2, SD=2.2), Standing (mean=0.35, SD=0.59)), F(1,11)=10.0, p=0.009; a main effect of Locomotion for Temporal workload (Walking (mean=17.5, SD=6.2), Standing (mean=10.4, SD=6.5)), F(1,11)=46.9, p<0.001; and a main effect of Locomotion for Performance workload (Walking (mean=7.9, SD=4.3), Standing (mean=5.2, SD=1.9)), F(1,11)=5.67, p=0.036. Figure V.16 shows more details.

V.5 Discussion

Overall, subjects memorized more than 70% of the suspicious trashcans in all experiments. In the S1G condition of both Experiments 6 and 7, the CSPs are around 90%. Although eight robots and the large search area with 20 trashcans inside at first seemed overwhelming to subjects, it turned out that subjects performed quite well in this condition. We think there are several factors contributing to this result. First, eight robots acted as a single team, and subjects were able to see all of them within the field of view or with only a slight head movement, which allowed subjects to use a simple strategy to memorize the locations of suspicious trashcans, such as grouping several trashcans nearby and memorizing the relative locations between the trashcans. Subjects also had some time to anchor their memory and strengthen their mental representation of the trashcan layout. Second, the eight robots moved side by side, swept the left half of the area, moved to the right half of the area, and swept the right half of the area. Therefore, subjects saw the suspicious trashcans in the left half of the area first and then in the right half. Searching in the two subareas was in tandem, so subjects have mental representations of the two subareas in tandem, therefore, subjects memorized the trashcans one by one with few disruptions. Third, the viewpoint did not change when subjects were required to identify the suspicious trashcans.

However, the separation of the robot team, not the amount of head movement, decreased subjects' performance. When the robot team split into two groups and the search area split into two areas, subjects had to have two mental representations of the trashcan layouts at the same time. During searching, subjects had to switch between the two mental representations, which is different from the S1G condition with the two mental representations in tandem. Therefore, the frequent switch between the two mental representations requires more effort and resources than that of tandem process of the two mental representations. The switching process increased the difficulty level of the task and perceived workload.

In Experiment 6, we noticed that locomotion might be a second factor affecting the performance, because there is a trend (not significant) that subjects performed worse in the locomotion condition than in the standing condition. In addition, locomotion increased the perceived workload significantly. A possible reason for this trend is that locomotion makes the task more rushed and thus increased the temporal demand. Therefore, we designed and conducted Experiment 8. The results of Experiment 8 showed locomotion indeed decreased performance and increased perceived workload (especially physical and temporal), in a more demanding task scenario, but occlusion had no effect on performance. Similarly, some evidence shows that self-motion, either active or passive, impairs the ability of multiple-object tracking (Thomas and Seiffert, 2010); and people employ a common mechanism to track changes both to the locations of moving objects around them and to keep track of their own positions. In our task scenarios, subjects needed to maintain their orientation and to keep track of their own locations during the locomotion, which demands attention, and thus suffered a cognitive cost due to locomotion, regardless of occlusion or not. We noticed that although the task is more demanding than that of Experiment 6 and 7, subjects still had comparable performance, with around 75% CSP, which means subjects were still able to do the task even when the search time was reduced to around 1 minute.

V.6 Conclusion

This chapter presented how people divide attention when they are placed under demanding task scenarios in large immersive virtual environments. Particularly, this chapter investigated the scenarios where a human is working with large robot team consisting of multiple robots that are potentially distributed into multiple groups in a large space. We discovered that how the robot team is separated is relevant. Particularly, the separation of the robot team affects human attention, e.g., decreased a supervisor's performance, increased the perceived workload, and increased the demand on attentional resources; but the amount of head movement was not relevant to that. Locomotion, in a more demanding task, further decreased the performance, increased the perceived workload, and increased demand on attentional resources; but occlusion in our experiments did not affect attention. The findings of these experiments, that there is primary cognitive cost to robot team separation and locomotion, could give ideas and suggestions on how to design and use such large robot team system in large IVEs, and in real world conditions.

CHAPTER VI

NAVIGATION STRATEGY

VI.1 Introduction

In some situations, the human supervisor of a robot team may need to be embedded with a team of robots. For example, to search for suspicious items, it may be best for a human supervisor to follow a robot team and examine the items flagged by robots as suspicious targets. IVEs can provide a good platform for controlled studies of such situations. In this chapter, we focus on how a supervisor's ability to maintain spatial orientation and to navigate is impacted by the presence of moving robots. In particular, this chapter studies how a supervisor performs with a large robot team in intensive memory and search tasks. For example, previous research shows people form mental representations of spatial structures in an orientation-dependent manner when they learn the environment from a single or small number of viewing perspectives (Shelton and Mcnamara, 1997; Diwadkar and McNamara, 1997). This chapter will investigate the orientation dependency of spatial memory, with and without the presence of a large robot team. Research also shows that people may have a preferential navigation strategy (Goeke *et al.*, 2013), by which we mean they favor an egocentric or allocentric approach. This chapter will examine that question, seeking to determine how an individual's navigation strategy impacts performance in dynamic robot team scenarios.

VI.2 Experiment 9

This experiment investigated people's spatial memory, in terms of view dependency and set-size dependency, in a simplified scenario where people have to search multiple suspicious targets from a subset of many objects in a large virtual space. This experiment does not involve robots.

VI.2.1 Method

In this experiment, subjects saw a number (twenty) of trashcans scattered about a plaza. Some of these trashcans contained balls. The task for the subjects was to search the trashcans, find the trashcans containing balls, and memorize the locations of these trashcans. Consistent with previous chapters, we call a trashcan containing a ball a "suspicious trashcan".

The task was similar to that of Experiment 5, in Chapter IV, as well as the procedure. In particular, subjects searched a few (eight) of the trashcans in a near-to-far manner. Please refer to Figure VI.1. Balls were located in some number of these. After searching all eight of the indicated trashcans, subjects were asked to indicate where the suspicious trashcans were. Subjects searched the trashcans sequentially, that is, a trashcan to search was indicated to the subject, and after that trashcan was searched, the next was indicated until all eight had been searched. Thus the order in which the trashcans were searched was controlled.

More specifically, subjects started from home position (position 1 in Figure VI.2), and the task started when subjects clicked the trigger of the joystick. The search was conducted in a near-to-far manner. At that time one of the trashcans would turn red. Subjects then approached it. When they were close, subjects clicked the joystick again. The trashcan would momentarily turn transparent, and they would be able to see if a ball was inside the trashcan or not. If a ball was inside, they were to note the location of that (suspicious) trashcan. When they were finished looking inside the trashcan, they looked around to find the next trashcan, which would be red and ready for searching. There were potentially a different number of target balls in the eight trash cans on each trial. The variable number of balls was (again) called the set size condition of the experiment. In this experiment we again used set sizes of 3, 5, and 7 balls.

After the search phase was completed, subjects were teleported to a new location from which they would be asked to recall the trashcans that were suspicious. The position to which they were teleported was the *viewing* position and in this experiment there were three different viewing positions. We varied the final view position with a 0° view (called the 0-view), a 90° view (called the 90-view), and a 135° view (called the 135-view), identical to Section IV.3 on page 45. Refer to Figure VI.2 for further reference. Subjects used the joystick to select the trashcans that they thought contained balls.

In the experiment, the environment was a virtual plaza of a virtual city, and the size of the trashcan array was around 50m by 40m, which is much larger than the size of our physical lab. Therefore, we increased the translational gain of the virtual environment to 10:1, consistent with the results of Chapter III.

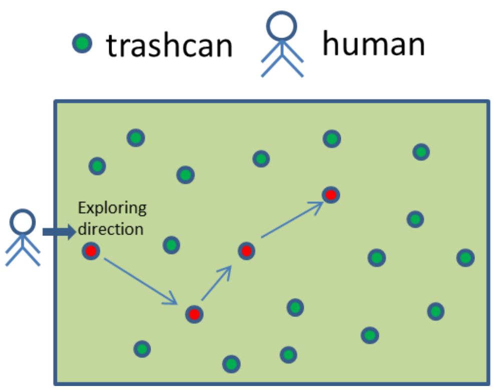

Figure VI.1: Experiment 9: Subjects explore in a near-to-far manner.

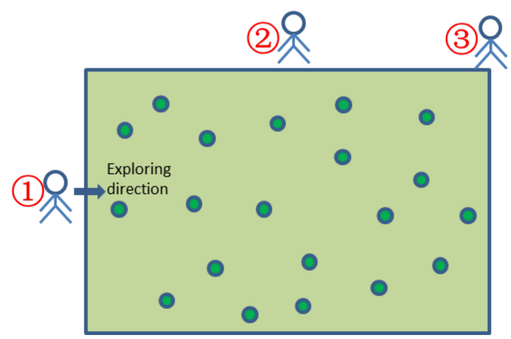

Figure VI.2: Experiment 9: Final viewing positions. Position 1 is the original start position.

The trials were divided into three sets, which included nine trials in each set. Within sets, trials were blocked by set size. Within sets, the view-angle condition was varied in a Latin square. Across sets, each number of balls in a set size occurred first, second, and third. Between every continuous two sets, subjects took a short break. Eighteen subjects, 9 male and 9 female, aged from 18 to 28, participated the experiments and were paid $15. Before the actual experiment began, each subject did two to three practice trials to make sure they were familiar with the basic environment and procedures.

We measured the correct selection percentage (CSP) of the balls and the time used to make the selection. We also calculated a two dimensional similarity between the correct configurations of the suspicious trashcans and the configurations of subjects selection, using bi-dimensional regression, similar to that of Experiment 5 in Chapter IV.

VI.2.2 Results and Discussion

The correct selection percentage (CSP) was analyzed in a two-way within subjects ANOVA. Please refer to Figure VI.3. The result showed main effects of both set size ($F(2,34)=13.46$, $p<.001$) and view angle ($F(2,34)=5.448$, $p<0.01$). For set size, a post hoc paired-sample t-test with Bonferroni correction shows difference between the 3-ball and 5-ball conditions, $t(17)=2.8$, $p=0.01$, and between the 5-ball and 7-ball conditions, $t(17)=3.29$, $p=0.004$. From Figure VI.3, we see the baseline condition (3-ball-0-view) has the highest CSP (mean=0.75, SD=0.16); and the 3-ball (mean=0.66, SD=0.17) and 7-ball (mean=0.65, SD=0.09) conditions have higher CSP than the 5-ball condition (mean=0.58, SD=0.1). For view-angle, a post hoc paired-sample t-test with Bonferroni correction shows a difference between all three pairs of viewing angles, the 0-view and 90-view ($t(17)=2.85$, $p=0.01$), the 0-view and 135-view ($t(17)=4.71$, $p<0.001$), and the 90-view and 135-view ($t(17)=2.78$, $p=0.01$). The collapsed mean CSPs for the 0-view, 90-view and 135-view are 0.69 (SD=0.1), 0.62 (SD=0.1), and 0.58 (SD=0.1), respectively. From Figure VI.3, we see the CSP decreased when viewing angles increased. Thus, people's performance is both view dependent and set-size dependent. We will probe these findings more deeply in the discussion section (page 83).

Task completion time was analyzed with a two-way within-subjects ANOVA as well. There are main effects of both set size ($F(2,34)=100$, $p<0.00001$) and view angle ($F(2,34)=4.22$, $p=0.02$). Refer to Figure VI.5. Regarding set size, a post hoc paired-sample t-test with Bonferroni correction shows differences between all three pairs of set sizes, the 3-ball (mean=21 s, SD=5) and 5-ball (mean=30 s, SD=6.8), $t(17)=7.20$, $p<0.001$; the 3-ball and 7-ball (mean=40 s, SEM=6.8), $t(17)=12.12$, $p<0.0001$; and the 5-ball and 7-ball, $t(17)=8.37$, $p<0.0001$. This result makes sense because subjects spent more time on the

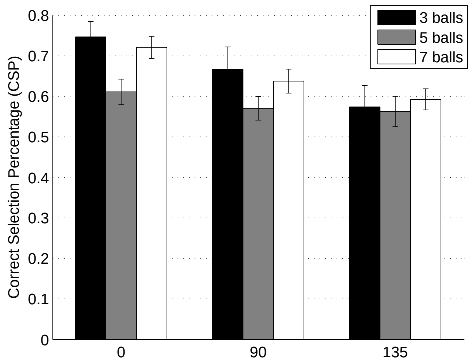

Figure VI.3: Experiment 9: The correct selection ratio across conditions. Error bars show standard errors of the mean.

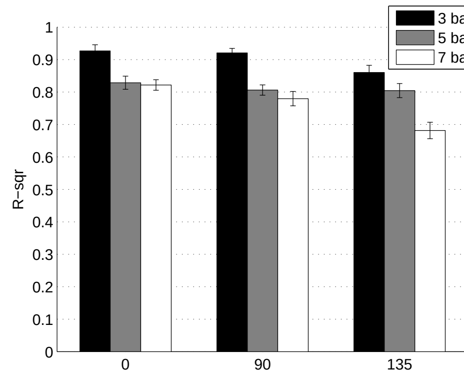

Figure VI.4: Experiment 9: r^2 from a bidimensional regression analysis across conditions. Error bars show standard errors of the mean.

task when they needed to select more balls. For the view-angle, a post hoc paired-sample t-test with Bonferroni correction shows a difference between the 90-view and 135-view, t(17)=3.80, p=0.001. The average time usage for the 0-view, 90-view, and 135-view is 29.7 s (SD=4.2), 29.5 s (SD=5.9), and 31.8 s (SD=6.4), respectively. The viewing angle results at our experimental power do not offer much insight.

The r^2 resulting from the bi-dimensional regression analysis, which shows the degree of correspondence between the two configurations from 0 to 1 after translation, scaling, and rotation, was analyzed in a two-way within subjects ANOVA. The results show main effects of both set size, F(2,34)=59.13, p<0.0001, and view-angle, F(2,34)=14.8, p<0.001. There is an interaction as well, F(4,68)=2.9, p=0.03. The weak interaction occurs because the 5-ball condition has essentially constant behavior across viewing conditions, whereas the 7-ball condition exhibits a downward trend, significant at the 135-view. Refer to Figure VI.4 for visual details. Regarding the set size, a post hoc paired-sample t-test with Bonferroni correction shows a difference between all three pairs of set sizes: the 3-ball and 5-ball (t(17)=6.36, p<0.001), the 3-ball and 7-ball (t(17) = 10.92, p<0.0001), and the 5-ball and 7-ball (t(17)=4.14, p<0.001). The collapsed mean r^2 for the 3-ball, 5-ball and 7-ball conditions are 0.90 (SD=0.05), 0.81 (SD=0.06), and 0.76 (SD=0.06), respectively. From

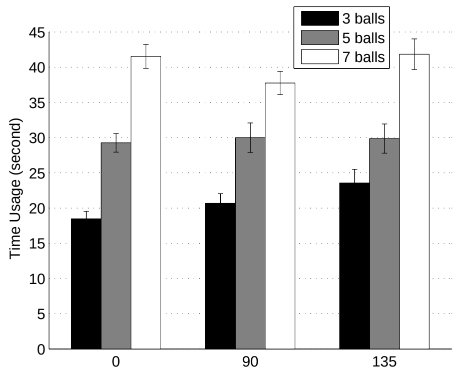

Figure VI.5: Experiment 9: Time usage across conditions. Error bars show standard errors of the mean.

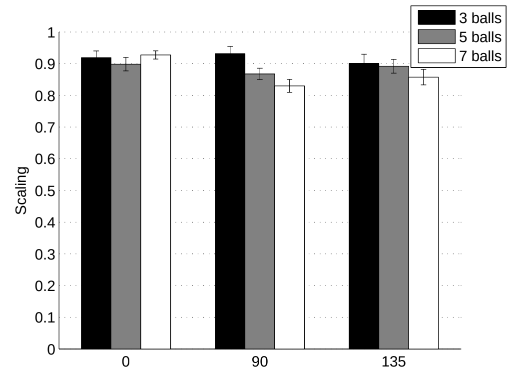

Figure VI.6: Experiment 9: Scaling from the bidimensional regression across conditions. Error bars show standard errors of the mean.

the Figure VI.4, we see the baseline 3-ball-0-view condition has the highest r^2, around 0.93, and the r^2 decreases when the set-size increases. Regarding view-angle, a post hoc paired-sample t-test with Bonferroni correction shows differences between the 90-view and 135-view (t(17) = 3.65, p=0.002) and the 0-view and 135-view condition (t(17)=4.68, $p<0.001$). The collapsed mean r^2 for the 0-view, 90-view and 135-view conditions are 0.86 (SD=0.06), 0.84 (SD=0.05), and 0.78 (SD=0.06), respectively. From Figure VI.4, the 0-view has the highest r^2, slightly higher than the 90-view and significantly higher than the 135-view, especially in the 3-ball condition and the 7-ball condition. The bi-dimensional regression results further show that people's spatial memory is both view dependent and set-size dependent, in terms of the geometry of the spatial layout.

The type of bi-dimensional regression we performed is a Euclidean one, and as mentioned, is composed of a scaling, translation, and rotation. We next tried to analyze the overall transformation in terms of these components to understand what was happening with the data. For the scaling component, a two-way within-subject ANOVA shows main effects for both set size (F(2,34)=4.71, p=0.02) and view-angle (F(2,34)=4.27, p=0.02). Please see Figure VI.6 for details. Regarding the set size, a post hoc paired-sample t-test with Bonferroni correction shows difference between the 3-ball (mean=0.92, SD=0.07) and the 7-ball (mean=0.87, SD=0.05) conditions, $t(17) = 3.29$, $p=0.004$. For view-angle, a post hoc paired-sample t-test shows a difference between the 0-view (mean=0.91, SD=0.05) and the 90-view (mean=0.88, SD=0.06), $t(17) = 2.59$, $p=0.02$. Scaling tends to decrease as view angle or set size increases.

For the rotation component, a two-way within-subjects ANOVA shows a main effect of view-angle ($F(2,34)=12.34$, $p<0.001$). See Figure VI.7. For the view-angle, the collapsed mean rotation for the 0-view, 90-view and 135-view conditions are 8.7 (SD=5.3), 11.0 (SD=5.9), and 16.2 (SD=6.5), respectively. A post hoc paired-sample t-test with Bonferroni correction shows a significant difference between the 90-view and 135-view ($t(17) = 3.16$, $p=0.006$), and between the 0-view and 135-view condition ($t(17)=4.59$, $p<0.001$). The rotation component was highest for the 135-view.

For the translation component, a two-way within-subjects ANOVA shows a main effect of view-angle ($F(2,34)=10.14$, $p<0.001$). The collapsed mean translation for the 0-view, 90-view and 135-view is 3.2 (SD=1.0), 3.4 (SD=1.3), and 4.4 (SD=1.4), respectively. See Figure VI.8. A post hoc paired-sample t-test with Bonferroni correction shows a significant difference between the 0-view and 135-view ($t(17) = 4.4$, $p<0.001$), and between the 90-view and 135-view conditions ($t(17)=3.13$, $p=0.006$). The translation component was highest for the 135-view.

From the above bi-dimensional regression analysis, we see that all three components, scaling, rotation, and translation, are view dependent, which causes the r^2 view dependence. All three components determine the geometric property of 2D configurations. Here the 2D configuration refers to the geometry of selected trashcans or the geometry of true suspicious trashcans. The scaling measure shows the size ratio of the two 2D configurations. The rotation measure shows the rotational angle between the two 2D configurations. The translation measure shows the offset between the 2D configurations.

We also looked for learning effects. After collapsing the time usage data in each set of trials (subjects did the experiment in three sets of trials), a one-way repeated measures

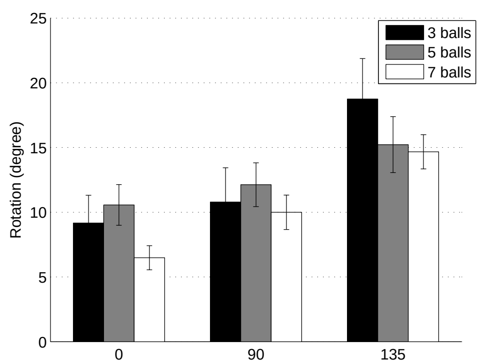

Figure VI.7: Experiment 9: Rotation from the bi-dimensional regression across conditions. Error bars show standard errors of the mean.

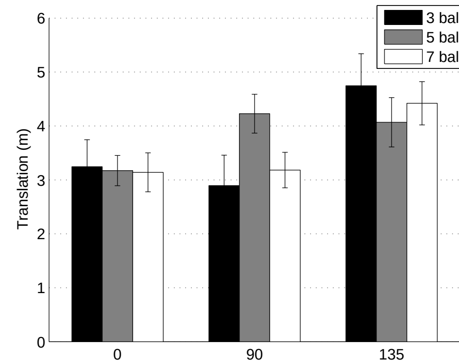

Figure VI.8: Experiment 9: Translation from the bi-dimensional regression across conditions. Error bars show standard errors of the mean.

ANOVA showed a main effect of set, $F(2,34)=30.4$, $p<0.0001$. The average time taken for a trial within a set is 34.9 s (SD=7.4), 29.7 s (SD=5.4), and 26.3 s (SD=4.9), respectively. This result shows that people completed the task faster as they gained experience.

In post-task surveys, subjects reported using either an egocentric system or allocentric system to memorize target locations. Some subjects (7 people) reported that they kept track of the target positions with respect to the start position (i.e., the red pole position) and integrated the path. In this way, they selected the trashcans that matched best to the segment points along their mental path. Other subjects (11 people) reported that they memorized the target locations with respect to the locations of other trashcans, such as grouping close targets together, dividing the search area into four sub-areas: near-left, near-right, far-left, and far right, counting the rough row and column number of the targets in the array. The first group thus seemed to be doing things egocentrically, and the second, allocentrically. The egocentric group seemed more view-dependent than the allocentric group. However, we did not control for this variable in our study and the survey correlation was too weak to be anything other than suggestive. Therefore, our next experiment is a controlled study to determine if navigation strategy is an important factor in a task of this type.

VI.3 Experiment 10

Experiment 9 showed that spatial memory is both view-dependent and set-size dependent. We would like to see whether this result still holds in the scenario where a human supervisor is embedded with a large robot team. We chose human robot teaming as the scenario, so that in this experiment there will be active robots.

VI.3.1 Method

In this experiment, eight robots move side by side and search for suspicious targets, in a near-to-far manner, among many objects in a large field. A human supervisor follows the robots and examines the objects flagged by the robots indicated as suspicious targets. When the search is finished, the human supervisor needs to recall the locations of those suspicious targets (the ones containing balls inside). Please refer Figure VI.9 and Figure VI.10.

The details of the procedure are as follows. Subjects followed a robot team that is searching suspicious trashcans over a large virtual space. The robots were searching the trashcans to find out which ones contained balls; the others were empty. When a robot found a suspicious target (one that might contain a ball), it would stop and emit a beeping sound. Subjects would also see the trashcan turned red. At this time subjects would approach the trashcan and check whether the trashcan had a ball inside by clicking the trigger of a joystick. When they clicked the trigger, the trashcan would disappear and the subjects would be able to see if there was a ball inside or not.

Since there was tentative evidence showing that subjects used two distinct navigation strategies in Experiment 9, in Experiment 10 we controlled for navigation strategy, as we would like to see if this individual difference would impact performance. This experiment consisted of two sections, a pre-screening and the main study. We first hired subjects to do pre-screening using the on-line test provided by Gramann (2012), which categorizes categorized them as egocentric and allocentric based on their performance after a few minutes of stimulus. We then employed those who were qualified for our study (who used a stable strategy in pre-screening) to do the main study. For the main study, we used 18 egocentric and 18 allocentric, 22 males and 14 females, aged 18 to 32. For the pre-screening, we tested a total of 140 subjects, including 27 egocentric and 90 allocentric, who used a stable strategy during the whole pre-screening test. We were only able to use 18 of the 27 egocentric subjects because nine of them either chose not to participate in the full experiment or quit during it. We note that the ratio of egocentric to allocentric subjects that we found is quite different from that reported in Goeke *et al.* (2013).

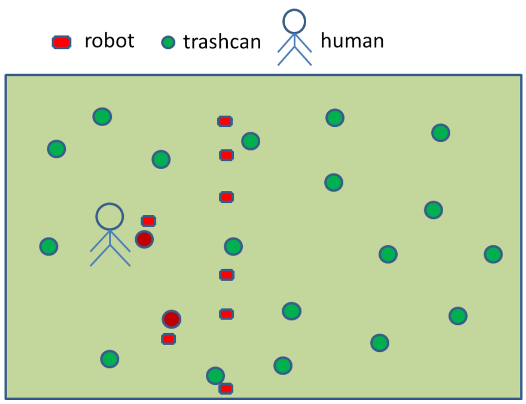

Figure VI.9: Experiment 10: A human supervisor following a robot team (consisting of 8 robots) to search suspicious trashcans.

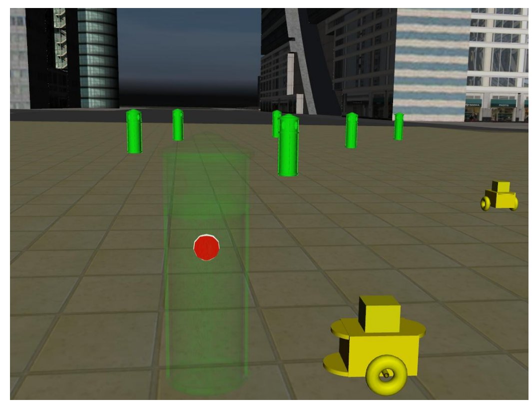

Figure VI.10: Experiment 10: A snapshot of the virtual environment from human supervisor's perspective.

The protocol of Experiment 10 is similar to that of Experiment 9. Each subject, egocentric or allocentric, completed three view-angles (the 0-view, 90-view, and 135-view) and three set-sizes (the 3-ball, 5-ball, and 7-ball), and three trials for each of the nine conditions. The measures are similar to those of Experiment 9. We measured the CSP, latency, and bi-dimensional regression metrics. Each subject was paid $20 in the full study.

VI.3.2 Results

For CSP, a three-way mixed ANOVA shows main effects of view-angle ($F(2,68)=42.4$, $p<0.001$) and set-size ($F(2,68)=3.7$, $p=0.03$), an interaction between set-size and view-angle, $F(4,136)=4.6$, $p=0.002$, and a three-way interaction, $F(4,136)=2.6$, $p=0.038$. We will analyze the three-way interaction. To understand this interaction we performed a contrast analysis. We are interested in the contrast between allocentric group and egocentric group, so their weights were set to -1 and 1, respectively. CSP data shows a linear effect across the view-angles, therefore, the weights were set to -1, 1/5, and 4/5, for 0-view, 90-view, and 135-view, respectively. The 5-ball and 7-ball conditions for the allocentric group look different from the 3-ball condition, therefore, the weights were set to -1/2, -1/2, and 1, respectively. The three-way contrast interaction analysis shows significance, $F(1,1)=4.1$, $p=0.04$. The results show above contrast holds. Therefore, the egocentric indeed performed differently from the allocentric, in a specific way, that egocentric people perform view dependently in all ball conditions, but allocentric people do not in the 5-ball and 7-ball conditions.

We next analyzed egocentric people and allocentric people separately. For egocentric people, a two way repeated measures ANOVA shows a main effect of view-angle, $F(2,34)=24.8$, $p<0.0001$. Please refer Figure VI.11 for details. For allocentric people, a two-way repeated measures ANOVA shows an interaction between set-size and view-angle, $F(4,68)=5.9$, $p<0.001$. Next, a one-way repeated measures ANOVA shows a significant effect of view-angle only in the 3-ball condition (in the allocentric condition), but no effect in either the 5-ball or 7-ball conditions, which indicates the source of the two-way interaction: allocentric people performed in a non-view-dependent manner in the 5-ball and 7-ball conditions but in a view-dependent manner in the 3-ball condition. The above results are consistent with the contrast analysis. Please refer Figure VI.12 for illustration.

For latency, a three-way mixed ANOVA shows main effects of both set-size, $F(2,68)=135$, $p<0.0001$, and view-angle, $F(2,68)=21.4$, $p<0.001$. The results make sense for set-size because subjects need a longer time to choose more targets. The latencies were 24.7s, 34.3s, and 47.3s for the 3-ball, 5-ball, and 7-ball conditions, respectively. For view-angle, the collapsed mean latency is 32.2s (SD=7.6) in the 0-view condition, 35.9s (SD=8.5) in the 90-view condition, and 38.1s (SD=9.5) in the 135-view condition. A post-hoc paired sample t-test with Bonferroni correction shows a difference between the 0-view and 90-view,

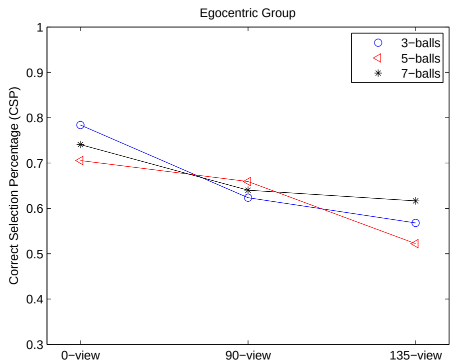

Figure VI.11: Experiment 10: CSP for the egocentric group.

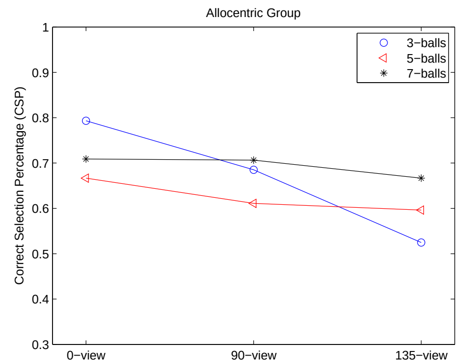

Figure VI.12: Experiment 10: CSP for the allocentric group.

t(35)=4.1, p<0.001, between the 0-view and 135-view, t(35)=6.0, p<0.001, and between the 90-view and 135-view, t(35)=2.7, p=0.01. The above results show that subjects needed a longer time when the view angle differed from original start point. Please refer Figure VI.13 for illustration.

For r^2, a three-way repeated measures ANOVA shows main effects of both set-size (F(2,68)=55, p<0.00001), and view-angle (F(2,68)=30, p<0.0001). For view-angle, the collapsed mean r^2 is 0.86 (SD=0.08) in the 0-view, 0.86 (SD=0.06) in the 90-view, 0.79 (SD=0.07) in the 135-view. A post-hoc paired sample t-test with Bonferroni correction shows a significant difference between the 90-view and 135-view, t(17)=7.6, p<0.001, and a significant difference between the 0-view and 135-view, t(17)=5.5, p<0.001. For set-size, the collapsed mean r^2 is 0.89 (SD=0.07) in the 3-ball, 0.83 (SD=0.07) in the 5-ball, and 0.78 (SD=0.07) in the 7-ball. A post-hoc paired sample t-test with Bonferroni correction shows a significant difference between the 3-ball and 5-ball (t(35)=4.8, p<0.001), the 5-ball and 7-ball (t(35)=5.9, p<0.001), and the 3-ball and 7-ball (t(35)=11.1, p<0.00001). Please refer to Figure VI.14 for illustration.

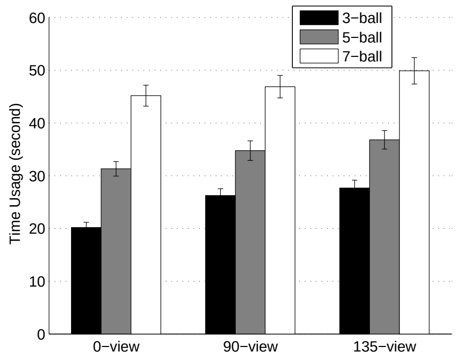

Figure VI.13: Experiment 10: Latency across conditions. Error bar shows one standard error of the mean.

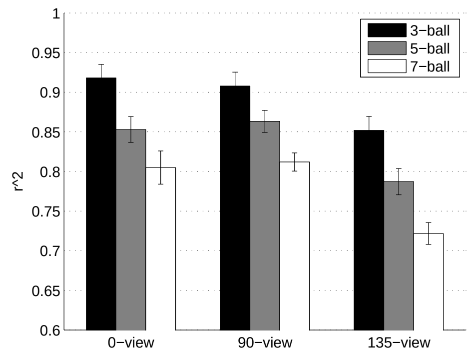

Figure VI.14: Experiment 10: r^2 across conditions. Error bar shows one standard error of the mean.

VI.4 Discussion

In Experiment 9, in terms of CSP, subjects' performance is view-dependent. They performed best in the 0-view condition and the performance decreased when the viewing angle differed from the 0-view angle. In particular, they started with the 0-view angle, searched in a near-to-far manner, roughly maintained forward direction during the search, and this resulted the best performance when they were teleported back to the start position (the 0-view condition).

Subjects' performance also varied by set-size. In particular, subjects performed better in the 3-ball and 7-ball conditions and worse in the 5-ball condition. This result is interesting and surprising. We conjecture that there was interference between strategies used for remembering layouts that occurred in the 5-ball condition. Specifically, when subjects searched the eight trashcans in the 5-ball condition, subjects were affected by the three trashcans that were empty (i.e., they were false positive trashcans). They had to memorize the locations of the five balls but also needed to handle the interruption of the three false-positive trashcans. However, they could easily ignore the only empty trashcan in the 7-ball condition, which might allow them to devote more attention and focus to the seven balls; as well, subjects found memorizing three balls in the 3-ball condition relatively simple.

Another possible reason as to why subjects performed better in the 7-ball condition over the 5-ball condition might be that they tended to have more relative location cues in the 7-ball condition. Some subjects reported they did better in 7-ball condition because they could group some of the targets together due to their proximity and that this strategy helped them remember the targets. Possibly it is more likely that seven balls out of twenty could be grouped better by proximity than five. However, when analyzing the data with the bi-dimensional regression, data from the 5-ball condition has a significantly higher fit to the correct configuration than data from the 7-ball condition. We conjecture that even though subjects missed relatively more balls, i.e., the exact locations of the balls as measured by CSP, in the 5-ball condition compared to the 7-ball condition, subjects remembered better the shape of ball configurations in the 5-ball condition.

We applied a bi-dimensional regression to these data sets to provide a quantitative look at the geometric nature of the data, for view-angle as well. Our results are broadly consistent with the CSP in that subjects' performance as measured by correlation of the selected configuration to the correct configuration is view dependent, highest at the 0-view condition, and lowest at the 135-view condition.

When we broke down the overall transformation in the bi-dimensional regression, we found scaling, rotation, and translation were all view dependent. In particular, the performance got worse when the viewing angle differed from original viewing angle. The rotation is interesting; when the viewing angle of subjects were rotated in the 90-view and 135-view conditions, their selections rotated as well, which implies that subjects could not hold the array independent of final viewing angle. As discussed in Section II, this result is consistent with prior work implicating that spatial memory is view dependent and deteriorates when the final view differs from the experienced viewing perspective (Diwadkar and McNamara, 1997; Shelton and Mcnamara, 1997; Wang *et al.*, 2006).

Based on some suggestive analysis from this experiment, we planned Experiment 10, where we pre-screened and categorized subjects into two groups, e.g., egocentric group and allocentric group, using the on-line test provided by Gramann (2012). We tested 140 subjects, and identified 27 egocentric and 90 allocentric. The ratio between the two groups are quite different from that reported in Goeke *et al.* (2013), where the two groups are quite even. We consulted the authors about this discrepancy; they hypothesized that the difference may be due to the cultural differences between Europe, Asia, and the United States.

A team of eight moving robots were also added to the scene in Experiment 10, as we wanted to see how moving robots would impact performance. The results showed that the dynamic robot team did not affect people's overall performance (with overall CSP above 60%). This finding is good because the large robot team did not distract users' attention.

We performed a three-way mixed ANOVA for this experiment and the results showed a three-way interaction. Our contrast analysis showed that the 5-ball and 7-ball conditions had different linear rates across view-angles between the egocentric and the allocentric. It also showed that the egocentric group indeed performed differently from the allocentric group, in a specific way: the allocentric group was not view dependent in the 5-ball and 7-ball conditions while the egocentric were view dependent.

Our analysis further indicates that the two groups performed in distinct patterns. The egocentric were view dependent in all three ball conditions, but the allocentric was only view dependent in the 3-ball condition. In other words, the allocentric were not view dependent in the 5-ball and 7-ball conditions. The above result suggest that the allocentric group might use a different strategy from the egocentric group in the 5-ball and 7-ball conditions. This result is consistent with our hypotheses derived from the previous experiment, that allocentric people are more likely to employ inter-object reference system to form the spatial mental representation, which results in a less view dependent spatial memory than egocentric people. Apparently, there are more inter-object relations in the 5-balls and 7-balls conditions, therefore, the allocentric group may take advantage of that and thus perform equally well in all three view-angle conditions. Prior research showed people's spatial memory is view dependent, but did not consider this individual factor.

The results also showed that latency was view dependent. People needed a longer time to complete the task when final view angle differed from the original viewpoint, which is consistent with prior research (Diwadkar and McNamara, 1997). The authors argued that people may need to form spatial representation of a novel perspective by normalizing it to the closest familiar learned spatial representation. That may be the reason why people needed a longer time in our experiment when they were teleported to a novel viewpoint (i.e., the 90-view and 135-view conditions). In particular, the 90-view condition took a longer time then the 0-view, and the 135-view condition took a longer time than the 90-view. In terms of the geometric nature of the results, with our r^2 measure, we did not find any difference between the egocentric and the allocentric, but we found main effects of both view angle and set size. The r^2 indicates the shape similarity of two 2D configurations after scaling, rotating, and translating. Specifically, when they were in the 135-view condition, the subjects could not hold the shape of the trashcan array. In terms of set size, the larger of set-size, the more challenging for subjects to hold the geometric shape of the array. This result is consistent with that of Experiment 9.

VI.5 Conclusion

This chapter studied how people navigate in the immersive virtual environment scenario where they have to search a subset of a set of objects among many objects in a sequential near-to-far manner, with and without the presence of arrays of moving indicators (e.g., robots). The results showed spatial memory was both view dependent and set-size dependent, which is consistent with previous study (Shelton and Mcnamara, 1997; Diwadkar and McNamara, 1997; Wang *et al.*, 2006). The results also showed individual difference of the navigation strategy. Particularly, in Experiment 10, the users were pre-screened and categorized to two groups by their preferential strategy: egocentric group and allocentric group. The results showed allocentric group were less view dependent than egocentric group, especially when they need to memorize more targets, which indicated that people indeed employed different strategies, and thus resulted in different performance patterns. The findings of Experiments 9 and 10 advance cognitive knowledge by understanding better how people's spatial memory is affected by view angle and set size in demanding scenarios, and by determining how individual navigation strategy impacts people's performance, with or without the presence of moving indicators. These findings are important, because we have a better understanding of how people navigate, form spatial mental representation, and use spatial inter-object cues. Given that egocentric people and allocentric people perform in two distinct patterns, this study suggests first steps for future research toward a more thorough understanding of individual difference and individual training in demanding spatial navigation scenarios, with dynamic proxies.

In particular, Experiment 10 studied a demanding human-robot teaming task. The results showed people's performance did not deteriorate due to the presence of a dynamic robot team. In this experiment, there were eight robots in total. Therefore, it seems that such a number of robots will not be a distractor to the supervisor or affect performance in demanding human-robot teaming systems, which is good. As shown above, this study can also provide some guidelines to the pre-training of the supervisor of large robot team. For example, a supervisor can take a simple pre-screening test (e.g., the online test provide by Gramann (2012), which only takes less than 30 mins). Then we can identify whether the supervisor belongs to which group, egocentric, allocentric, or a mixed. Most people are either egocentric or allocentric. If he or she is egocentric, we may be able to provide the supervisor some form of training so that they can take advantage of inter-object cues, like the allocentric people do. In this way, in demanding navigation tasks, their performance may be less view dependent. This changing of preferential navigation strategy is feasible. Gramann *et al.* (2005) showed that when virtually navigating through a tunnel with turns (i.e., no locomotion, only watching optical flow on a monitor), allocentric subjects can perform as egocentric if they are instructed to travel through the tunnel like a cyclist, and egocentric subjects can perform like allocentric if they are instructed to maintain a mental birds-eye view of the tunnel while traveling. Therefore, while our study may provide suggestions for pre-training, how to conduct such training is still an open question.

CHAPTER VII

CONCLUSION AND FUTURE DIRECTIONS

VII.1 Conclusion

This thesis developed an optimized locomotion system for large immersive virtual environments, investigated the trade-off between various locomotion interfaces, and studied people's navigation and spatial memory in demanding human-robot teaming scenarios. This thesis had two primary themes: (1) it leveraged psychological measures to develop methods for building IVE systems that allow effective navigation; (2) it used large IVEs as a platform to study people's perception, i.e., attention and spatial memory.

Virtual environment systems have increasing applications in many areas, such as training, education, and physical therapy. In order to use an IVE to its fullest potential, people have to be able to explore IVEs effectively. Thus, in Chapter II this thesis discussed a number of different ways to navigate through virtual environments and the trade-offs of these methods. In addition, Chapter II discussed the state-of-art research in spatial memory, navigation, and human-robot interaction, because this thesis is interested in people's navigation and attention in demanding human-robot teaming scenarios.

Chapter III built an optimized locomotion system that allows people to freely and effectively explore large immersive virtual environments within a limited physically tracked space. This chapter continued and extended previous work done by Williams (2007) in combining their translational gain method and resetting method. The results showed users are able to successfully update their location and orientation in our combined system, even when users were experiencing a high translational gain of 50:1 and four resets along the path, although there are some cognitive costs of resets. Given that two resets had the highest turning errors overall, especially at 50:1 gain, and recalling that four resets conditions aligned subjects' orientation at the end of each path segment (i.e., the locations of the poles) with that of the 0-resets condition, we conjectured that subjects might use sound cues from the experimenter and so we designed Experiment 2 in which subjects heard white noise and voice commands from a noise-masking headphone. In Experiment 2, we compared the 0, 2, and 4 resets conditions at 50:1 gain. The results showed two and four resets had significantly higher turning errors than zero resets, but no difference between the former two conditions. A mixed ANOVA analysis showed there was a significant difference between Experiments 1 and 2. It seemed that subjects indeed may use sound cues from the experimenter in Experiment 1 and that there are cognitive costs of resets.

Noticing that people were not sensitive to the rotational gain in the resetting method, e.g., the rotational gain does not have to be exactly 2:1, we designed a method, called the varying gain method, to optimize the rotational gain in such a way that users are reset to the most open space of the room, e.g., the center of the tracked room, when they reach the boundary of the room. The results of our evaluation of this method (Experiment 3) again showed the cost of resets, but did not find a difference between the varying gain method and the 2:1 gain method, and showed that users are able to maintain their spatial orientation. However, since the varying gain method has the potential to reduce the number of resets, thus reducing the cost introduced by resets, we believe varying the gain is superior to a 2:1 gain.

Our locomotion system is different from other redirected methods in several ways. First, some of the main redirected methods (Razzaque, 2005; Steinicke *et al.*, 2010) seek to manipulate the virtual camera of IVEs, affecting the translational gain and rotational gain in an imperceptible way, so that users are able to navigate large IVEs in a limited physical space. Thus, those methods typically require a large physical space — Razzaque (2005) used a tracked space with a 15m radius, and Steinicke *et al.* (2008a) used a tracked space with a 24m radius — which is outside the realm of many virtual environment facilities. However, our method seeks to maintain users' spatial orientation, and manipulates translational gain and rotational gain overtly. Thus our method does not require a large tracked space. For example, our tracked space is around 4m by 5m. Secondly, most of those main redirected methods need to know targets beforehand and thus do not allow free navigation, which brings up some limitations of IVE exploration. Our method enables free walking and thus allows free exploration of large IVEs.

However, a locomotion interface may not be suitable for all IVE systems, e.g., those system that do not support a tracking interface. In Chapter III, we studied alternative locomotion interfaces and investigated the trade-offs compared to a bipedal locomotion interface. Experiment 4 compared three different interfaces, i.e., a pure joystick (the joystick condition), joystick translation plus physical rotation (the joystick rotation condition), and walking (the walking condition). This experiment replicated the experiment done by Riecke *et al.* (2010), but extended the study to large IVEs. Thus this experiment also compared three translational gains for walking, 1:1, 2:1, and 10:1. The results showed that in 1:1 gain the pure joystick interface is worst, but we did not find any difference between the joystick rotation condition and the walking condition, which is consistent with Riecke *et al.* (2010). The results also showed that people had fewer perfect searches in 10:1 gain, which tends to indicate that for more complicated tasks higher gains are not as good.

Given the somewhat ambivalent results of Experiment 4, we further examined the relationship between joystick rotation and walking in Experiment 5. Experiment 5 is another demanding spatial search scenario, where subjects searched for various numbers of suspicious targets from a subset of randomly distributed objects and had to recall the locations of those suspicious targets. The results showed that the walking condition is superior to the joystick rotation condition. Thus, from Experiment 4 and 5, the effectiveness of locomotion interfaces depends on the nature of navigation tasks. In more demanding scenarios, the walking interface is desirable; but in less demanding scenarios, the joystick rotation interface may be equivalent (Klatzky *et al.*, 1998). In a highly demanding spatial navigation scenario, the user employing a joystick rotation interface may suffer more spatial disorientation. Thus, our results may give guidance to both designers and users of IVEs on locomotion interfaces that are both efficient and economic. Recently, Young *et al.* (2014) also replicated Riecke's experiment except in their experiment the subjects only did the joystick rotation condition, but in both a commodity-level HMD system (i.e., Oculus Rift) and a standard HMD system (i.e., nVisor SX60, the same as the system used in this thesis). The authors found that the commodity-level system outperformed the standard system, in terms of task completion time and total object visits, but did not find any difference in other measures, such as the total number of targets found or the number of revisits. Their results suggest that the Oculus Rift might be a good alternative to high-cost HMD systems, although users may suffer more simulator sickness.

Most of the experiments in this thesis involve high translational gains, e.g., a 50:1 gain in Experiments 1 and 2, a 10:1 gain in Experiments 1, 3, and 4. Williams (2007) suggested an upper bound for translational gain is 50:1, under which people are able to maintain their spatial orientation. Consistent with the author's result, subjects are able to do so at 50:1 gain in our Experiments 1 and 2. However, in Experiment 4, people performed worse in 10:1 gain than in 1:1 gain or 2:1 gain, in terms of perfect search, which indicates high gains may not be as good for such complicated tasks. Therefore, the effectiveness of high translational gain may depend on the nature of the task.

Having developed a system capable of allowing navigation in IVEs, this thesis turned its focus to an area where those systems might be employed. Chapters V and VI focused on problems motivated by fundamental issues involved in simulating human-robot teams in a large IVE, where a human acting in a supervisor's role may have to oversee a large robot team that is potentially distributed geographically. Chapters V and VI looked at issues of attention and spatial navigation that arose from a consideration of this problem.

In particular, Chapter V discussed how a human supervisor attends to a large robot team, potentially separated geographically into multiple groups. In such situations, the human supervisor may have to divide his or her attention between multiple robot groups. Chapter V investigated how this division would affect human performance in a typical robot-team searching scenario. In Experiment 6, eight robots swept a large virtual plaza and conducted an exhaustive search over twenty randomly distributed trashcans. A human supervisor had to recall the locations of suspicious targets indicated by the robots. Experiment 6 examined three conditions: robots acting as a single unit, robots split into two groups with a standing supervisor, and robots split into two groups with a moving supervisor. The results showed the separation of the robot team significantly affected attention and performance, but locomotion did not further decrease performance. From Experiment 6, we have a greater understanding of the cognitive limitations of a human supervisor in such human-robot teaming tasks. In addition, the results suggest a primary cognitive cost of robot-team division. To remove a potential confound factor of head moving, we next compared two conditions, together and separated, similar to those in Experiment 6, but the field of view (FOV) of the field and head moving angle were the same (Experiment 7). The results showed that performance was significantly worse in the separated condition, which indicates it was the separation of robot team that affected attention and performance.

Since locomotion significantly increased perceived workload in Experiment 6, we next examined how locomotion affected attention in another demanding task (Experiment 8). To verify whether it was a locomotion effect or an occlusion effect, an occlusion factor was also added. The results showed that only locomotion had an effect. The above three experiments give us a better idea of the cognitive limitations of a human supervisor in a robot team and of the limits when attention is divided in such scenarios. Chapter V further provided suggestions on how to design human-robot teaming systems and provided implications for the command and control structures of such a team.

In some situations a human supervisor may have to embed with the robot team and follow the robots in a search task. Chapter VI examined how a supervisor's ability to navigate and maintain spatial orientation will be affected in such scenarios with moving robots. First, Experiment 9 studied a simplified scenario where users have to search suspicious targets in a subset of randomly distributed objects, in a near-to-far manner, without robots.

Experiment 9 showed spatial memory is both view dependent and set-size dependent, consistent with prior research (Diwadkar and McNamara, 1997; Wang *et al.*, 2006).

There was some evidence from this experiment indicating that users employed two distinct navigation strategies resulting in two distinct performance patterns. Therefore, we explored this hypothesis by designing a controlled study for individual navigation strategy, and added moving robots into the scenario. This study consisted of two parts: a pre-screening and a main study. During pre-screening, we selected subjects who employed a stable navigation strategy, either egocentric or allocentric, to participate in the main study. Results from the main study showed people's performance did not deteriorate due to the presence of moving robots. Results also showed the 5-ball and 7-ball conditions of the allocentric had different linear rate across view-angles from those of the egocentric. Finally, our results indicated that egocentric and allocentric subjects performed in distinct patterns; in particular, allocentric people performed in a less view dependent manner than egocentric people, especially when the set-size increased.

These last two studies advance our knowledge of how people form their spatial memories in demanding navigation tasks, and by determining how individual navigation strategy plays a role in their performance, with and without the presence of an array of dynamic proxies. With a better understanding of such individual differences, we may be able to provide guidelines for individualized training for people. While our results are not yet strong enough to justify such a thing, one can imagine taking a simple test similar to the online test provided by Gramann (2012), which takes less than 30 minutes. The results of such a test would determine the type and course of further training. In this example, training might be used to reinforce or change the preferential navigation strategy, something that has been shown to be feasible (Gramann *et al.*, 2005).

This thesis primarily builds virtual environment systems that support effective interaction, learning, and exploration. It focuses on a fundamental function of IVE systems: navigation. Chapter III designed and optimized a locomotion system that allowed people to freely and effectively navigate through large immersive virtual environments within a limited physically tracked space, and Chapter IV determined the trade-offs between a walking interface and other non-locomoting alternatives (i.e., joystick) in demanding navigation tasks in large IVEs. Secondly, having developed systems that allow navigation in large IVEs, in Chapters V and VI we studied fundamental issues in human-robot teaming systems, such as navigation, spatial attention, and spatial memory, all of which involve locomotion and navigation in large IVEs. These latter two chapters advance our knowledge of spatial cognition, e.g., how people divide attention in demanding scenarios, how people form spatial memory, and how individual navigation strategy impacts performance. Therefore, this work can form the basis for navigation training, either in human-robot teaming scenario or other navigation scenarios.

VII.2 Future Directions

We would like to revisit many of the topics discussed in this thesis in future work. This thesis is only the beginning of a complete solution for large IVE navigation and a first step toward a complete human-robot teaming system in large IVEs. Therefore, future work can involve better locomotion interfaces for large IVEs, a better understanding of cognitive limitations and capabilities in demanding human-robot teaming systems, and better suggestions for the training of human supervisor of human-robot teaming systems. Future work can also advance our knowledge in several cognitive areas, such as spatial memory, navigation, and spatial attention.

First, Chapter III designed and optimized a locomotion system for large IVEs within a limited physically tracked space, but there are still cognitive costs to resets. People may get disoriented after walking a long distance and experiencing many resets. How to remedy this issue is an important issue.

Second, Chapter IV investigated trade-offs between a joystick interface and a walking interface. In Experiment 4, subjects appeared to have two different strategies, i.e., the lawnmower and perimeter, similar to Ruddle and Lessels (2009), and seemed to perform differently, especially in the 2:1 gain condition. As we saw in Chapter VI, in a similar study (Experiments 9, with only a walking interface), subjects report using an egocentric strategy (e.g., memorizing the targets with respect to the starting position, or memorizing the path) or an allocentric strategy (e.g., memorizing the targets with respect to other objects or the surrounding environment), and the two groups seem to perform differently. Experiment 10 had a controlled study and the results showed some difference between the two groups. Other work has presented similar individual differences as well, e.g., people have a preferential strategy, egocentric or allocentric, as they navigate through a simulated tunnel or star field (Gramann *et al.*, 2005; Goeke *et al.*, 2013). Therefore, a better understanding of the individual differences, in a scenario similar to that of Experiment 4, may provide more ideas as to how navigation strategy may impact performance, and may provide suggestions for methods to train and help users gain better navigation ability in large IVEs.

Third, Chapter V determined how humans supervising a large robot team divide their attention in demanding tasks. We discovered that attention division affects performance and locomotion further affects performance in more demanding tasks. The result is consistent with other prior research: most of the evidence suggests a unified model of spatial attention, that attention is confined to a single, indivisible focal region in the visual field (McCormick *et al.*, 1998). Jans *et al.* (2010) suggested attention division may not be easily achieved by a naive observer, but, instead, it can be achieved through training. Therefore, one direction of the future work could be supervisor training of attention division, which may be helpful for the design and training of such human-robot teaming system. Similarly, some evidence shows locomotion, actively or passively, impairs the ability to keep track of multiple moving objects (Thomas and Seiffert, 2010). The authors conjectured that the reason for this is that people may use a common strategy to track their own location during locomotion, which requires attentional resources. Thus one future direction could be to study how the amount of locomotion and the demands of the task affect performance.

Fourth, Chapter VI focuses on people's individual navigation strategy in a demanding human-robot teaming scenario. We found that the allocentric group indeed performed differently from the egocentric group: the egocentric group was view dependent, but the allocentric group was not as view dependent, especially when the set-size increased. Most people seem to have a preferred navigation strategy, but prior work has shown that this preferred strategy can be changed through training (Gramann *et al.*, 2005). Therefore, one future direction would lie in the training of a supervisor of large robot-teams: in a view-critical, demanding task we may be able to identify the preferential navigation strategy of a supervisor through straightforward testing, and devise appropriate training to improve navigation results, potentially by modifying the preferred navigation strategy. Significant work remains to determine if this idea is reasonable and bring it to fruition.

Another promising avenue opened by this research is in the area of locomotion modes. From Experiments 4 and 5, we understand better the trade-offs between a walking interface and a joystick rotation interface: walking is desirable for more complicated tasks, while the joystick rotation may be equivalent in less demanding tasks. But how to relate this relationship to tasks demands is unclear and could potentially be quantified. In the task of Klatzky *et al.* (1998), where subjects traveled a two-segment path and were asked to point back to the start locations, physical rotation was shown to be critical and important. In a more demanding task of Riecke *et al.* (2010), rotation was shown to be sufficient. We have the similar results in 1:1 gain condition of Experiment 4, a replicated study of Riecke *et al.* (2010). But in another similar task of Ruddle and Lessels (2009), rotation is important but not sufficient. In our Experiment 5, rotation is not sufficient either. Therefore, it seems that physical rotation is critical in most navigation scenarios, but to determine situations in which physical translation is critical is still an open question. There are other interfaces besides a joystick, such as gaming controllers and omni-directional treadmills. Omni-directional treadmills have been prohibitively expensive, but recently gaming oriented omni-directional treadmills, such as the Virtuix Omni, promise to bring the price down to the commodity level. It will be necessary to compare such interfaces to a walking interface, not only because of the price, but also because omni-treadmills provide more motion cues than joystick. Another direction is to compare various locomotion interfaces in commodity-level HMD systems, e.g., the Oculus Rift. Young *et al.* (2014) showed Oculus Rift outperformed the standard high-cost HMD system, i.e., nVisor SX60. Therefore, to determine the trade-offs between various locomotion interfaces in Oculus Rift-based systems looks promising.

BIBLIOGRAPHY

ADAMO, M., PUN, C., PRATT, J. and FERBER, S. (2008). Your divided attention, please! the maintenance of multiple attentional control sets over distinct regions in space. *Cognition*, 107(1):295–303.

ADAMS, J. A., COOPER, J. L., GOODRICH, M. A., HUMPHREY, C., QUIGLEY, M., G, B. and MORSE, B. S. (2007). Byuhcmi technical report: Camera-equipped mini uavs for wilderness search support: Task analysis and lessons from field trials.

ALLEN, R. C. and SINGER, M. J. (1997). Landmark direction and distance estimation in large scale. *In Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, pages 1213–1217.

ANDERSON, J. and BALTES, J. (2007). A mixed reality approach to undergraduate robotics education. *In AAAI'07: Proceedings of the 22nd national conference on Artificial intelligence*, pages 1979–1980. AAAI Press.

AZAMASAB, E. and HU, X. (2007). An integrated multi-robot test bed to support incremental simulation-based design. *In System of Systems Engineering, 2007. SoSE '07. IEEE International Conference on*, pages 1 –7.

BOUDOIN, P., DOMINGUES, C., OTMANE, S., OURAMDANE, N. and MALLEM, M. (2008). Towards multimodal human-robot interaction in large scale virtual environment. *In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, HRI '08, pages 359–366, New York, NY, USA. ACM.

BOWMAN, D. A., DAVIS, E. T., HODGES, L. F. and BADRE, A. N. (1999). Maintaining spatial orientation during travel in an immersive virtual environment. *Presence*, 8(6): 618–631.

BRUDER, G., STEINICKE, F. and HINRICHS, K. H. (2009). Arch-explore: A natural user interface for immersive architectural walkthroughs. *3D User Interfaces*, 0:75–82.

BURGESS, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends in Cognitive Sciences*, 10(12):551 – 557.

CARBON, C.-C. (2013). Bidimregression: Bidimensional regression modeling using r. *Journal of Statistical Software, Code Snippets*, 52(1):1–11.

CASPER, J. and MURPHY, R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 33(3):367 – 385.

CHANCE, S. S., GAUNET, F., BEALL, A. C. and LOOMIS, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence: Teleoperators and Virtual Environments*, 7(2):168–178.

CHEN, H., WULF, O. and WAGNER, B. (2006). Object detection for a mobile robot using mixed reality. *In* ZHA, H., PAN, Z., THWAITES, H., ADDISON, A. and FORTE, M., éditeurs : *Interactive Technologies and Sociotechnical Systems*, volume 4270 de *Lecture Notes in Computer Science*, pages 466–475. Springer Berlin / Heidelberg.

CHEN, I. Y.-H., MACDONALD, B. and WÜNSCHE, B. (2009). Mixed reality simulation for mobile robots. *In ICRA'09: Proceedings of the 2009 IEEE international conference on Robotics and Automation*, pages 922–927, Piscataway, NJ, USA. IEEE Press.

CHENG, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, 23:149–178.

COLLETT, T. H., MACDONALD, B. A. and GERKEY, B. P. (2005). Player 2.0: Toward a practical robot programming framework. *In Proceedings of the Australasian Conference on Robotics and Automation (ACRA 2005)*, page 145.

DIWADKAR, V. A. and MCNAMARA, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, 8(4):302–307.

DRURY, J., SCHOLTZ, J. and YANCO, H. (2003). Awareness in human-robot interactions. *In Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 1, pages 912 – 918 vol.1.

ENDSLEY, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):65–84.

ENDSLEY, M. R. (2000). Direct measurement of situation awareness:validity and use of SAGAT. *In* ENDSLEY, M. R. and GARLAND, D. J., éditeurs : *Situation awareness analysis and measurement*. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

ENGEL, D., CURIO, C., TCHEANG, L., MOHLER, B. and BÜLTHOFF, H. H. (2008). A psychophysically calibrated controller for navigating through large environments in a limited free-walking space. *In Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, VRST '08, pages 157–164, New York, NY, USA. ACM.

FRANKENSTEIN, J., MEILINGER, T., MOHLER, B. J. and BÜLTHOFF, H. H. (2009). Spatial memory for highly familiar environments. *In Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 2650–2655. Curran Red Hook, NY.

GALLISTEL, C. R. (1990). *The Organization of Learning*. MIT Press, Cambridge, MA.

GERKEY, B., VAUGHAN, R. T. and HOWARD, A. (2003). The player/stage project: Tools for multi-robot and distributed sensor systems. *In Proceedings of the 11th international conference on advanced robotics*, volume 1, pages 317–323.

GOEKE, C. M., KÖNIG, P. and GRAMANN, K. (2013). Different strategies for spatial updating in yaw and pitch path integration. *Frontiers in behavioral neuroscience*, 7.

GOODRICH, M., COOPER, J., ADAMS, J., HUMPHREY, C., ZEEMAN, R. and BUSS, B. (2007). Using a mini-uav to support wilderness search and rescue: Practices for human-robot teaming. *In Safety, Security and Rescue Robotics, 2007. SSRR 2007. IEEE International Workshop on*, pages 1–6.

GOODRICH, M. A., MORSE, B. S., ENGH, C., COOPER, J. L. and ADAMS, J. A. (2009). Towards using unmanned aerial vehicles (uavs) in wilderness search and rescue: Lessons from field trials. *Interaction Studies*, 10(3):453–478.

GOODRICH, M. A., MORSE, B. S., GERHARDT, D., COOPER, J. L., QUIGLEY, M., ADAMS, J. A. and HUMPHREY, C. (2008). Supporting wilderness search and rescue using a camera-equipped mini uav. *Journal of Field Robotics*, 25(1-2):89–110.

GRAMANN, K. (2012). A test for your spatial navigation skills. http://www.navigationexperiments.com/TurningStudy.html.

GRAMANN, K., MULLER, H. J., EICK, E.-M. and SCHONEBECK, B. (2005). Evidence of separable spatial representations in a virtual navigation task. *Journal of Experimental Psychology-Human Perception and Performance*, 31(6):1199–1223.

GREENAUER, N. and WALLER, D. (2008). Intrinsic array structure is neither necessary nor sufficient for nonegocentric coding of spatial layouts. *Psychonomic Bulletin and Review*, 15:1015–1021.

HARRIOTT, C. E., ZHANG, T. and ADAMS, J. A. (2011). Evaluating the applicability of current models of workload to peer-based human-robot teams. *In Proceedings of the 6th international conference on Human-robot interaction*, HRI '11, pages 45–52, New York, NY, USA. ACM.

HART, S. G. and STAVENLAND, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *In* HANCOCK, P. A. and MESHKATI, N., éditeurs : *Human Mental Workload*, chapitre 7, pages 139–183. Elsevier.

HODGSON, E., BACHMANN, E. and THRASH, T. (2014). Performance of redirected walking algorithms in a constrained virtual world. *Visualization and Computer Graphics, IEEE Transactions on*, 20(4):579–587.

HODGSON, E., BACHMANN, E. and WALLER, D. (2008). Redirected walking to explore virtual environments: Assessing the potential for spatial interference. *ACM Trans. Appl. Percept.*, 8(4):22:1–22:22.

HOLLERBACH, J. M., CHECCACCI, D., NOMA, H., YANAGIDA, Y. and TETSUTANI, N. (2003). Simulating side slopes on locomotion interfaces using torso forces. *International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 0: 91.

HUMPHREY, C. M., HENK, C., SEWELL, G., WILLIAMS, B. W. and ADAMS, J. A. (2007). Assessing the scalability of a multiple robot interface. *In Proceedings of the*

2^{nd} *ACM/IEEE International Conference on Human-Robot Interaction*, pages 239–246, New York, NY, USA.

HUMPHREY, Curtis M.; Adams, J. A. (2009). Robotic tasks for chemical, biological, radiological, nuclear and explosive incident response. *Advanced Robotics*, 23:1217–1232.

INTERRANTE, V., RIES, B. and ANDERSON, L. (2007). Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments. *3D User Interfaces*, 0:167–170.

IWAMURA, Y., SHIOMI, M., KANDA, T., ISHIGURO, H. and HAGITA, N. (2011). Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets? *In Proceedings of the 6th international conference on Human-robot interaction*, HRI '11, pages 449–456, New York, NY, USA. ACM.

JANS, B., PETERS, J. C. and DE WEERD, P. (2010). Visual spatial attention to multiple locations at once: the jury is still out. *Psychological review*, 117(2):637.

JERALD, J., PECK, T., STEINICKE, F. and WHITTON, M. (2008). Sensitivity to scene motion for phases of head yaws. *In APGV '08: Proceedings of the 5th symposium on Applied perception in graphics and visualization*, pages 155–162, New York, NY, USA. ACM.

KELLY, J. and MCNAMARA, T. (2008a). Spatial memory and spatial orientation. *In Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, volume 5248 de *Lecture Notes in Computer Science*, pages 22–38. Springer Berlin Heidelberg.

KELLY, J. W. and MCNAMARA, T. P. (2008b). Spatial memories of virtual environments: How egocentric experience, intrinsic structure, and extrinsic structure interact. *Psychonomic Bulletin and Review*, 15(2):322–327.

KELLY, J. W., MCNAMARA, T. P., BODENHEIMER, B., CARR, T. H. and RIESER, J. J. (2008). The shape of human navigation: How environmental geometry is used in maintenance of spatial orientation. *Cognition*, 109(2):281 – 286.

KELLY, J. W., MCNAMARA, T. P., BODENHEIMER, B., CARR, T. H. and RIESER, J. J. (2009). Individual differences in using geometric and featural cues to maintain spatial orientation: Cue quantity and cue ambiguity are more important than cue type. *Psychonomic Bulletin & Review*, 16(1):176–181.

KLATZKY, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. *In Spatial cognition*, pages 1–17. Springer.

KLATZKY, R. L., LOOMIS, J. M., BEALL, A. C., CHANCE, S. S., G., R. and GOLLEDGE, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science*, 9:293–298.

KUHL, S. A., CREEM-REGEHR, S. H. and THOMPSON, W. B. (2008). Recalibration of rotational locomotion in immersive virtual environments. *ACM Trans. Appl. Percept.*, 5(3):1–11.

LATHROP, W. B. and KAISER, M. K. (2002). Perceived orientation in physical and virtual environments: Changes in perceived orientation as a function of idiothetic information available. *Presence*, 11(1):19–32.

LATIMER, D., I., SRINIVASA, S., LEE-SHUE, V., SONNE, S., CHOSET, H. and HURST, A. (2002). Towards sensor based coverage with robot teams. *In Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, volume 1, pages 961 – 967 vol.1.

LIVATINO, S., MUSCATO, G. and PRIVITERA, F. (2009). Stereo viewing and virtual reality technologies in mobile robot teleguide. *Robotics, IEEE Transactions on*, 25(6):1343 – 1355.

LOOMIS, J. M., BLASCOVICH, J. J. and BEALL, A. C. (1999a). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Reseach Methods, Instruments, and Computers*, 31:557–564.

LOOMIS, J. M., KLATZKY, R. L., GOLLEDGE, R. G. and PHILBECK, J. W. (1999b). *Wayfinding: Cognitive mapping and other spatial processes*, chapitre Human navigation by path integration, pages 125–151. Johns Hopkins University Press, Baltimore, MD.

MCCORMICK, P., KLEIN, R. and JOHNSTON, S. (1998). Splitting versus sharing focal attention: comment on castiello and umiltà (1992). *Journal of experimental psychology: Human perception and performance*, 24(1):350–357.

MCNAMARA, T. P. (2003). How are the locations of objects in the environment represented in memory? *In* FREKSA, C., BRAUER, W., HABEL, C. and WENDER, K. F., éditeurs : *Spatial cognition III: Routes and navigation, human memory and learning, spatial representation and spatial learning*, pages 174–191. Springer Berlin / Heidelberg.

MCNAMARA, T. P., RUMP, B. and WERNER, S. (2003). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin & Review*, 10(3): 589–595.

MICHAEL, N., FINK, J. and KUMAR, V. (2008). Experimental testbed for large multirobot teams. *Robotics Automation Magazine, IEEE*, 15(1):53 –61.

MOU, W. and MCNAMARA, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28(1):162–170.

MOU, W., MCNAMARA, T. P., VALIQUETTE, C. M. and RUMP, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1):142–157.

NBJ (2006). Local homeland security agency to stage disaster exercise. Nashville Business Journal, on-line, http://www.bizjournals.com/nashville/stories/2006/02/27/daily16.html.

NETH, C. T., SOUMAN, J. L., ENGEL, D., KLOOS, U., BULTHOFF, H. H. and MOHLER, B. J. (2012). Velocity-dependent dynamic curvature gain for redirected walking. *Visualization and Computer Graphics, IEEE Transactions on*, 18(7):1041–1052.

NILSSON, N. C., SERAFIN, S. and NORDAHL, R. (2014). Establishing the range of perceptually natural visual walking speeds for virtual walking-in-place locomotion. *Visualization and Computer Graphics, IEEE Transactions on*, 20(4):569–578.

NISHIWAKI, K., KOBAYASHI, K., UCHIYAMA, S., YAMAMOTO, H. and KAGAMI, S. (2008). Mixed reality environment for autonomous robot development. *In Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 2211–2212. IEEE.

NITZSCHE, N., HANEBECK, U. D. and SCHMIDT, G. (2004). Motion compression for telepresent walking in large target environments. *Presence*, pages 44–60.

OCULUS (2014). Oculus rift. www.oculus.com. Accessed: 2014-11-12.

ODASHIMA, T., ONISHI, M., LUO, Z. and HOSOE, S. (2003). Immersion type virtual environment for human-robot interaction. *In Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 4, pages 3651 – 3656 vol.4.

PECK, T. C., FUCHS, H. and WHITTON, M. C. (2009). Evaluation of reorientation techniques and distractors for walking in large virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 15:383–394.

PECK, T. C., FUCHS, H. and WHITTON, M. C. (2010). Improved redirection with distractors: A large-scale-real-walking locomotion interface and its effect on navigation in virtual environments. *Virtual Reality Conference, IEEE*, 0:35–38.

PÉRUCH, P., BELINGARD, L. and THINUS-BLANC, C. (2000). Transfer of spatial knowledge from virtual to real environments. *In* FRESKA, C., éditeur : *Spatial Cognition II*. Springer, Berlin.

PYLYSHYN, Z. W. and STORM, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3):179–197.

RAZZAQUE, S. (2005). *Redirected walking*. Thèse de doctorat, Chapel Hill, NC, USA. Adviser-Brooks,Jr., Fredrick P.

RAZZAQUE, S., KOHN, Z. and WHITTON, M. C. (2001). Redirected walking. *In Proceedings of EUROGRAPHICS*, volume 9, pages 105–106. Citeseer.

REKLEITIS, I., LEE-SHUE, V., NEW, A. P. and CHOSET, H. (2004). Limited communication, multi-robot team based coverage. *In Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, volume 4, pages 3462 – 3468 Vol.4.

RIECKE, B. E. (2008). Consistent left-right reversals for visual path integration in virtual reality: More than a failure to update one's heading? *Presence: Teleoperators and Virtual Environments*, 17(2):143–175.

RIECKE, B. E., BODENHEIMER, B., MCNAMARA, T. P., WILLIAMS, B., PENG, P. and FEUEREISSEN, D. (2010). Do we need to walk for effective virtual reality navigation? physical rotations alone may suffice. *In Spatial Cognition*, pages 234–247.

RIESER, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology*, 15:1157C1165.

RIESER, J. J., PICK, H. L., ASHMEAD, D. H. and GARING, A. E. (1995). Calibration of human locomotion and models of perceptual-motor organization. *Journal of Experimental Psychology-Human Perception and Performance*, 21(3):480–497.

RUDDLE, R. A. (2001). Navigation: Am i really lost or virtually there? *Engineering Psychology and Cognitive Ergonomics*, 6:135–142.

RUDDLE, R. A. and LESSELS, S. (2006). For efficient navigation search, humans require full physical movement but not a rich visual scene. *Psychological Science*, 17(6):460–465.

RUDDLE, R. A. and LESSELS, S. (2009). The benefits of using a walking interface to navigate virtual environments. *ACM Transactions on Computer-Human Interaction*, 16(1):1–18.

RUDDLE, R. A., PAYNE, S. J. and JONES, D. M. (1999). Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays? *Presence*, 8(2):157–168.

RUDDLE, R. A., VOLKOVA, E. and BÜLTHOFF, H. H. (2013). Learning to walk in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 10(2):11.

SHELTON, A. and MCNAMARA, T. (1997). Multiple views of spatial memory. *Psychonomic Bulletin and Review*, 4:102–106. 10.3758/BF03210780.

SHELTON, A. L. and MCNAMARA, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology*, 43(4):274 – 310.

SLATER, M., USOH, M. and STEED, A. (1995). Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 2:201–219.

SOUMAN, J. L., GIORDANO, P. R., SCHWAIGER, M., FRISSEN, I., THÜMMEL, T., ULBRICH, H., LUCA, A. D., BÜLTHOFF, H. H. and ERNST, M. O. (2008). Cyberwalk: Enabling unconstrained omnidirectional walking through virtual environments. *ACM Trans. Appl. Percept.*, 8(4):25:1–25:22.

STEINFELD, A., FONG, T., KABER, D., LEWIS, M., SCHOLTZ, J., SCHULTZ, A. and GOODRICH, M. (2006). Common metrics for human-robot interaction. *In Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, HRI '06, pages 33–40, New York, NY, USA. ACM.

STEINICKE, F., BRUDER, G., JERALD, J., FRENZ, H. and LAPPE, M. (2008a). Analyses of human sensitivity to redirected walking. *In Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, VRST '08, pages 149–156, New York, NY, USA. ACM.

STEINICKE, F., BRUDER, G., JERALD, J., FRENZ, H. and LAPPE, M. (2010). Estimation of detection thresholds for redirected walking techniques. *IEEE Transactions on Visualization and Computer Graphics*, 16:17–27.

STEINICKE, F., BRUDER, G., KOHLI, L., JERALD, J. and HINRICHS, K. (2008b). Taxonomy and implementation of redirection techniques for ubiquitous passive haptic feedback. *Cyberworlds, International Conference on*, 0:217–223.

SUMA, E., CLARK, S., KRUM, D., FINKELSTEIN, S., BOLAS, M. and WARTE, Z. (2011). Leveraging change blindness for redirection in virtual environments. *In Virtual Reality Conference (VR), 2011 IEEE*, pages 159 –166.

SUMA, E., LIPPS, Z., FINKELSTEIN, S., KRUM, D. and BOLAS, M. (2012). Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):555 –564.

TEMPLEMAN, J. N., DENBROOK, P. S. and SIBERT, L. E. (1999). Virtual locomotion: Walking in place through virtual environments. *Presence*, pages 598–617.

THOMAS, L. E. and SEIFFERT, A. E. (2010). Self-motion impairs multiple-object tracking. *Cognition*, 117(1):80–86.

TOBLER, W. R. (1994). Bidimensional regression. *Geographical Analysis*, 26(3):187–212.

TOMBU, M. and SEIFFERT, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, 108(1):1–25.

USOH, M., ARTHUR, K., WHITTON, M. C., BASTOS, R., STEED, A., SLATER, M. and BROOKS, Jr., F. P. (1999). Walking > walking-in-place > flying, in virtual environments. *In Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pages 359–364, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

VIRTUIX (2014). Virtuix omni. http://www.virtuix.com/. Accessed: 2015-1-3.

WALLER, D. (2006). Egocentric and nonegocentric coding in memory for spatial layout: Evidence from scene recognition. *Memory & cognition*, 34(3):491–504.

WALLER, D. and HODGSON, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of experimental psychology. Learning, memory, and cognition*, 32(4):867.

WALLER, D., HUNT, E. and KNAPP, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence*, 7(2):129–143.

WANG, R. F., CROWELL, J. A., SIMONS, D. J., IRWIN, D. E., KRAMER, A. F., AMBINDER, M. S., THOMAS, L. E., GOSNEY, J. L., LEVINTHAL, B. R. and HSIEH, B. B. (2006). Spatial updating relies on an egocentric representation of space: Effects of the number of objects. *Psychonomic bulletin & review*, 13(2):281–286.

WANG, R. F. and SPELKE, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, 77:215–250.

WANG, R. F. and SPELKE, E. S. (2002). Human spatial representation: insights from animals. *Trends in Cognitive Sciences*, 6(9):376 – 382.

WILLIAMS, B., BAILEY, S., NARASIMHAM, G., LI, M. and BODENHEIMER, B. (2011). Evaluation of walking in place on a wii balance board to explore a virtual environment. *ACM Trans. Appl. Percept.*, 8(3):19:1–19:14.

WILLIAMS, B., NARASIMHAM, G., MCNAMARA, T. P., CARR, T. H., RIESER, J. J. and BODENHEIMER, B. (2006). Updating orientation in large virtual environments using scaled translational gain. *In Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, APGV '06, pages 21–28, New York, NY, USA. ACM.

WILLIAMS, B., NARASIMHAM, G., RUMP, B., MCNAMARA, T. P., CARR, T. H., RIESER, J. and BODENHEIMER, B. (2007). Exploring large virtual environments with an hmd when physical space is limited. *In Proceedings of the 4th symposium on Applied perception in graphics and visualization*, APGV '07, pages 41–48, New York, NY, USA. ACM.

WILLIAMS, L. E. (2007). *Exploring Large Virtual Environments with an HMD on Foot When Physical Space is Limited*. Thèse de doctorat, Vanderbilt University, Nashville, TN.

XIE, X., LIN, Q., WU, H., ADAMS, J. A. and BODENHEIMER, B. E. (2012). Immersion with robots in large virtual environments. *In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, HRI '12, pages 273–274, New York, NY, USA. ACM.

XIE, X., LIN, Q., WU, H., NARASIMHAM, G., MCNAMARA, T. P., RIESER, J. and BODENHEIMER, B. (2010). A system for exploring large virtual environments that combines scaled translational gain and interventions. *In Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV '10, pages 65–72, New York, NY, USA. ACM.

YOUNG, M. K., GAYLOR, G. B., ANDRUS, S. M. and BODENHEIMER, B. (2014). A comparison of two cost-differentiated virtual reality systems for perception and action tasks. *In Proceedings of the ACM Symposium on Applied Perception*, pages 83–90. ACM.