The Effect of Locomotion Modes on Spatial Memory and Learning in Large Immersive Virtual Environments: A Comparison of Walking with Gain to Continuous Motion Control

Xianshi Xie[1], Richard A. Paris[2], Timothy P. McNamara[3], and Bobby Bodenheimer[2](✉)

[1] Facebook Inc., 1 Facebook Way, Menlo Park, CA, USA
`xianshi.matt.xie@gmail.com`
[2] Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA
`{richard.a.paris,bobby.bodenheimer}@vanderbilt.edu`
[3] Department of Psychology, Vanderbilt University, Nashville, TN, USA
`t.mcnamara@vanderbilt.edu`

Abstract. This paper examines locomotion methods for large virtual environments presented through head-mounted displays and involving complex navigation. Our interest is on comparing methods involving bipedal locomotion to those involving motion controlled continuously through a joystick. In the first of two experiments 36 participants performed a navigational search task where they either walked, their translation/rotations were controlled by joystick, or their translations were controlled by joystick but their rotations were controlled by body-based turning. In addition, the optic flow rate, or translational gain of movement, was varied. In the second experiment, 24 participants performed a complex search involving only walking and joystick translation with body-based rotation, followed by a task in which they were asked to recall their search from novel viewpoints. Our findings suggest that for such complex tasks walking is preferable if space and equipment allow for such a locomotion mode.

1 Introduction

Large immersive virtual environments (IVEs) provide simulations that are useful in a variety of contexts and situations, such as education (Angulo and de Velasco 2014) and entertainment (Tan et al. 2015). An immediate choice confronting the designer of such large IVEs is what locomotion method to use for moving users through the environment. From the perspective of maintaining spatial orientation, bipedal locomotion is often desirable (Chance et al. 1998; Ruddle and Lessels 2009) because it is naturalistic and able to provide proprioceptive and vestibular cues that are important. However, there are situations where bipedal

S. Creem-Regehr et al. (Eds.): Spatial Cognition 2018, LNAI 11034, pp. 58–73, 2018.
https://doi.org/10.1007/978-3-319-96385-3_5

locomotion as a method of moving through an IVE is not viable, e.g., where the IVEs do not support a tracking interface, or where the physical space housing the IVE is so limited that it does not afford walking as a viable interface. This paper studies alternative locomotion interfaces and compares them 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, as there are a plethora of locomotion methods for IVEs, we limit our investigation of these interfaces. In particular, we consider locomotion with joysticks as an alternative method. Note that joysticks provide continuous translation and (potentially) rotation through a motion control device. There are many such motion control devices in the gaming community, e.g., the Oculus Touch, the Xbox controller, and the Razer Hydra (Young et al. 2014).

Many methods have been developed for navigation in large virtual environments (Bowman et al. 1997; Razzaque et al. 2001; Interrante et al. 2007; Williams et al. 2007; Engel et al. 2008; Paris et al. 2017), but the optimal method of doing so depends on many factors. These factors include the performance metric by which the method will be judged, but also such factors as room size and layout (Azmandian et al. 2017), and the virtual environment (Langbehn et al. 2017). For example, if the room size is sufficiently large, then a redirected walking technique might be employed (Steinicke et al. 2008, 2010); if the room size is small, however, some of these techniques can induce simulator sickness, and methods involving only a joystick or teleportation might be needed, e.g., Langbehn et al. (2018). If the environment is very large, then some method of locomotion beyond normal bipedal locomotion may be appropriate, e.g., the methods of Williams et al. (2007) or Interrante et al. (2007). This paper focuses its attention on metrics associated with spatial memory, as well as comparing the specific locomotion methods of bipedal locomotion to joystick-controlled movement.

2 Background

2.1 Walking in Virtual Environments

Several techniques that permit exploration of large virtual environments have been developed (Razzaque et al. 2001; Interrante et al. 2007; Williams et al. 2007; Engel et al. 2008). These techniques generally revolve around manipulating some form of visual information, which allows for a much larger virtual environment to be explored. One such technique is to add a translational gain factor (Interrante et al. 2007; Williams et al. 2006) that multiplies the virtual space by altering the optic flow experienced by users as they move through the environment. These methods have the advantage of granting proprioceptive and vestibular motion cues that are useful in integrating spatial knowledge (Ruddle 2013; Riecke et al. 2010). This result taken with prior work (Ruddle et al. 2011; Chance et al. 1998; Ruddle and Lessels 2006) suggests the benefits of walking over simple joystick motion.

2.2 Body-Based Cues

Body-based cues generated from natural walking are typically divided into two functions: these cues grant rotation information or translation information. Full rotation-based information can be provided without requiring subjects to physically translate. This allows for the separation of these two streams of information. Rotational-based cues, which provide information on the angle an individual has turned, can be useful when keeping oneself oriented in the virtual environment (Grechkin and Riecke 2014). Providing translation-based cues requires both tracking equipment and physical space. Information provided by translation-based cues are useful for estimating the distances between two objects and gauging the scale of the environment. In this work we separate these two streams of information in that we are able to provide both streams, only rotation-based information, or neither. Ruddle et al. (2011) and Riecke et al. (2010)

Development of spatial knowledge requires the integration of many smaller frames of reference (called local frames) into one larger global frame of reference (Zhang et al. 2014; Kelly et al. 2010; Greenauer and Waller 2010; Meilinger et al. 2014). The integration of these frames of reference require that there be some way of tying that information together. Mou et al. (2007) looked at using geometry as the integration point, whereas Roskos-Ewoldsen et al. (1998) looked at how orientation played a role in this combination of spatial memory. In addition to how one integrates these frames, research has been conducted into how these smaller frames are created. Prior work (McNamara et al. 2003; Marchette et al. 2011) examined the difference between egocentric and exocentric frames of reference and how they affected spatial memory. Looking further into egocentric and exocentric frames of reference, Xiao et al. (2009) analyzed how subjects built these frames relative to themselves or some external object. In the present work, subjects had to integrate spatial information after traveling from station to station. Yamamoto and Shelton (2007) also examined this type of integration and the effect of proprioceptive information.

3 Experiment 1

3.1 Materials

The experiments were conducted in a 7.3 m by 8.5 m laboratory. The virtual environment was presented by a full color stereo NVIS nVisor SX Head Mounted Display (HMD) with 1280 × 1024 resolution per eye, a nominal FOV of 60° diagonally, and a frame rate of 60 Hz. An interSense IS-900 precision motion tracker was used to update the participant's rotational movement around all three axes. Position was updated using four optical tracking cameras that operated with two LED lights. The virtual environment displayed in the HMD was rendered in Vizard (Worldviz, Santa Barbara, CA).

3.2 Participants

Thirty-six students from Vanderbilt University, 18 males and 18 females aged 18–32, participated in this experiment. Each was compensated $10 for participation in the experiment.

3.3 Experimental Setup and Procedure

The experimental setup was based on 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, participants used a wireless joystick to achieve translation and rotation in a virtual environment. In the JR condition, participants physically turned while using the joystick to translate in the virtual environment. In the walking condition, participants 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 participants were not able to get any orienting cues from the surrounding environment (Fig. 1). The sixteen possible locations for targets were randomly distributed according to a Poisson Disk distribution (Fig. 2) so that participants were not able to get cues from intrinsic reference frames based on the layout of the targets (Mou and McNamara 2002).

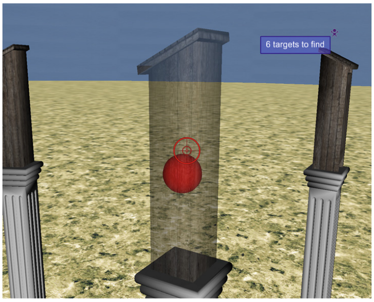

Fig. 1. Experiment 1: screen shot of the bird houses and red ball. (Color figure online)

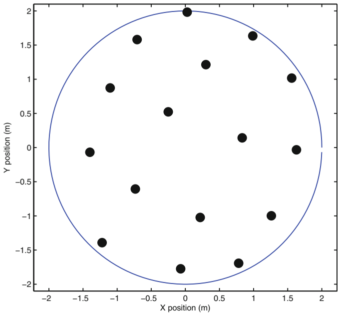

Fig. 2. Experiment 1: example of how birdhouses were distributed in an environment—a Poisson disk distribution.

There were three different optical 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 participants, 12 for each group, were assigned to the three different translational gain conditions. All participants performed the three locomotion methods, i.e., joystick, joystick plus rotation, and walking condition. In this experiment, each participant did one training trial before each locomotion mode condition that contained three experimental trials; in total, there were twelve trials for all three locomotion modes. The locomotion mode order was balanced; the experiment was gender balanced overall.

During the experiment, participants started in the center of the virtual space. When they approached a birdhouse, they clicked the joystick (participants carried a joystick in the Walking condition also), and the birdhouse became transparent, so that they 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, participants could tell they were revisiting the birdhouse from the presence of the blue ball and a revisit audio cue. The task ended when all eight targets had 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 participants 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.

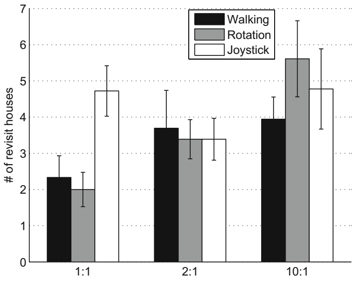

Fig. 3. Mean number of revisits across conditions. Error bars show SEs.

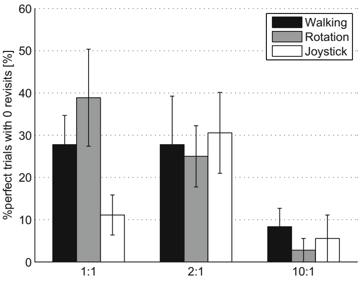

Fig. 4. Mean percentage of perfect search trials across conditions. Error bars show SEs.

3.4 Results and Discussion

We measured the number of times that participants revisited the same birdhouses that had been visited before, the number of targets found by participants, 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.

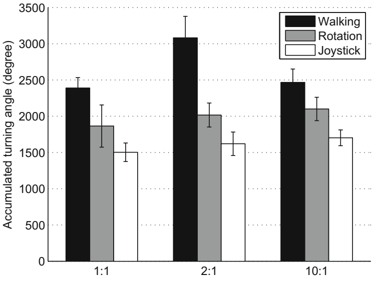

Fig. 5. Mean accumulated turning angle during the search. Error bars show SEs.

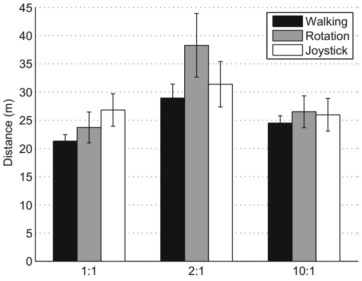

Fig. 6. Mean normalized travel distance across conditions. Error bars show SEs.

Number of Revisits. The number of revisits in the 1:1 gain condition is highest in the pure joystick condition, which is consistent with the findings of Riecke et al. (2010), while in a larger virtual space there is no difference across locomotion methods. Regarding the number of revisits, a 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 Fig. 3 for the interaction, we can see different patterns under different translational gains. In the 1:1 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-subjects ANOVA on the participants 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 findings of Riecke et al. (2010), who showed rotation may suffice for such complex spatial orientation task. However, in the 2:1 gain and 10:1 gains, a within-subjects ANOVA showed no main effect on locomotion methods. Participants 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 the JR condition, but the performance was not detectably changed under the J and W conditions.

Gain. 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, $p = 0.01$ (see Fig. 4). A post-hoc unpaired-sample t-test with Bonferroni correction on gain showed a 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 should not be used in more challenging tasks.

Orienting Motions. 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, $p < 0.0001$ (see Fig. 5). 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 participants looked around more during search under walking and joystick rotation conditions indicates that they may employ a qualitatively different navigation strategy: by looking around more participants 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.

Time and Travel Distance. There was 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 in the two conditions involving the joystick. A mixed model ANOVA showed a main effect of gain condition, F(2,33) = 3.556, $p = 0.04$ (see Fig. 6). A post-hoc unpaired-sample t-test on gain showed difference between the 1:1 (mean = 23.9 m, SD = 6.1) and 2:1 (mean = 32.9 m, SD = 12.1) gains, t(16.153) = 2.265, $p = 0.038$, and a marginal difference between the 2:1 and 10:1 (mean = 25.7 m, SD = 6.4) gains, t(16.59) = 1.8, $p = 0.09$. Our best interpretation of the higher travel distance in the 2:1 gain condition is that it results from the mixture of navigation strategies being employed in solving the problem, as explained next.

Navigation Strategy. Navigation strategy may rely on the size of virtual space. According to the answers to the post-task survey, most participants typically employed two distinct navigation strategies to complete the task. In this analysis, we adopt the terminology used by Ruddle and Lessels (2009): (a) perimeter (participants 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 participants into a perimeter or lawnmower strategy based on their travel paths. In the 1:1 gain, 58% of participants 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 participants in each group.

Interface Preference. The walking interface was preferred by most participants 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.

To summarize, where participants 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, walking is better than joystick rotation, and joystick rotation is better than the pure joystick, but it is not clear how this affects the task. There are suggestions that navigation strategy changes with size of environment, but nothing conclusive.

4 Experiment 2

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 1, 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 more orientation dependent (Shelton and McNamara 2001; McNamara 2003).

4.1 Materials and Participants

The equipment and setting for this experiment was identical to that of Experiment 1. Eighteen subjects, nine male and nine female, aged 18–30 from Vanderbilt University, participated in this experiment and were paid $15 for their participation.

4.2 Experimental Setup and Procedure

In this experiment, participants 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 participants was to memorize the locations of the suspicious trashcans among all trashcans. In particular, they searched a few (eight) of the trashcans. Balls were located in some number of these. After searching all eight of the indicated trashcans, they were asked to indicate where the suspicious trashcans were. Participants searched the trashcans sequentially, that is, a trashcan to be searched was indicated to each of them, and after that trashcan was searched the next was indicated sequentially until all eight had been searched. Thus the order in which the trashcans were searched was controlled.

More specifically, participants started from home position (position 1 in Fig. 7), and the task started when they 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. They then approached it. When they were close, participants 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 participant's working memory.

After the search phase was completed, participants 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 orientation dependent (McNamara 2003; Shelton and McNamara 2001), and even view dependent under some circumstances (Diwadkar and McNamara 1997; Waller 2006). 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). These positions are illustrated in Fig. 7. Participants used the joystick to select the trashcans that they thought contained balls.

We used a within-subject design for this study. Thus, each participant completed both the walking condition and the joystick rotation condition. Before the actual experiment began, they completed a few practice trials to make sure they were familiar with the basic environment and procedures. Within either condition, they did nine trials composed of 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 50 m by 40 m, which is much larger than the size of our physical lab. We used a translational gain of 10:1 in this experiment. We wanted 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).

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 participants' 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, 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 participants 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.

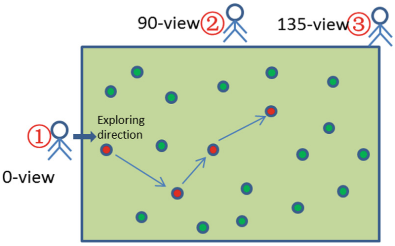

Fig. 7. Experiment 2: participants search in a near-to-far manner. Position 1, 2, and 3 indicate final viewing positions. Position 1 is the original start position.

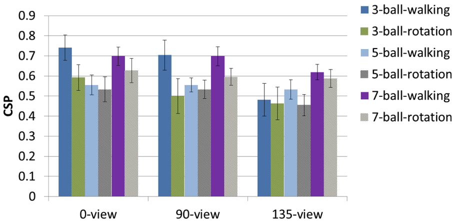

Fig. 8. Mean CSP across all conditions. Error bars show SEs.

---

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

4.3 Results and Discussion

Correct Selection Percentage. 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 = 0.53 m SD = 0.11) and 7-ball (mean = 0.64, SD = 0.12) conditions, t(17) = 5.7, p < 0.001 (see Fig. 8).

Latency. 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 participants have to use a longer time to choose more targets. The latency is 26.7 s, 36.2 s, and 45.4 s for the 3-ball, 5-ball, and 7-ball sets, respectively. For view-angle, the collapsed mean latency is 32.2 s (SD = 7.2) in the 0-view condition, 36.4 s (SD = 10.7) in the 90-view condition, 40.0 s (SD = 10.2) in the 135-view condition. A post-hoc paired sample t-test with Bonferroni correction shows a difference between the 0-view and 135-view, t(17) = 6.0, p < 0.001 (see Fig. 9).

Bidimensional Regression. 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,

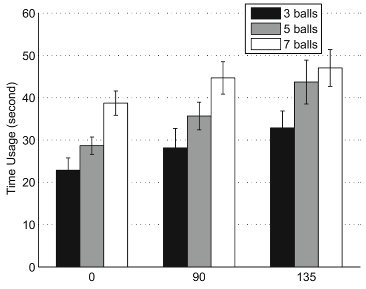

Fig. 9. Mean latency across the conditions. Error bars show SEs.

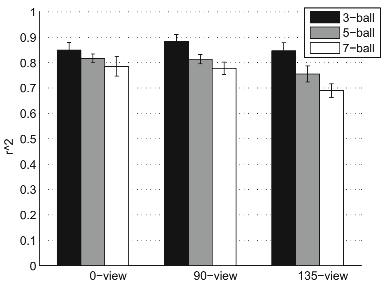

Fig. 10. Mean r^2 across the conditions. Error bars show SEs.

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, $p < 0.001$) (see Fig. 10).

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 (see Fig. 11). 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.

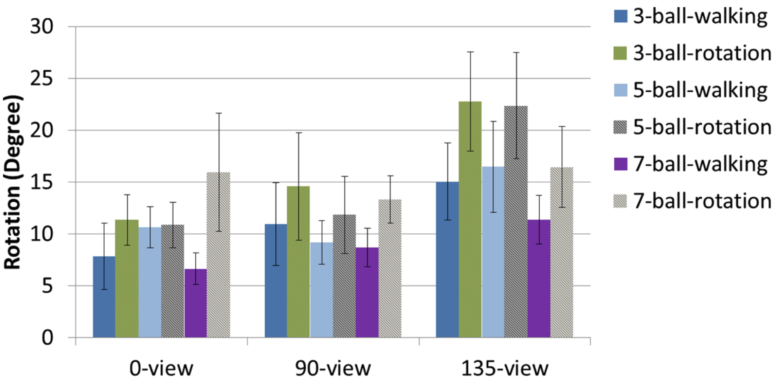

Fig. 11. Mean rotation across the conditions. Error bars show SEs.

From these results, we may conclude participants performed better with the walking interface than with the joystick interface, which indicates the importance of physical translation in spatial navigation, especially in complex memory and search tasks. Participants 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 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 participants 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. We hypothesize that there is interference between strategies that participants are using for remembering layouts between set sizes, but future work will be needed to verify this.

Some participants 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 prior research that indicates physical translation is critical to path integration (Ruddle and Lessels 2009) and spatial navigation. In the first experiment, the traveled path and orientation data of participants were recorded; the data showed participants had more orienting motion and less collision with objects when they walked, consistent with both Riecke et al. (2010) and Ruddle and Lessels (2009).

5 Conclusion

This paper presented two experiments that attempted to better understand 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 2. 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 requires only a small space, no need to walk, and no position tracking, users' spatial performance may suffer worse learning experiences in IVEs due to worse navigation ability. Recent work has shown that walking-in-place methods, while they have some navigational disadvantages, have some potential as locomotion methods if physical space is a concern (Paris et al. 2017). Future work will investigate this question more fully.

Acknowledgement. The authors would like to thank Qiufeng Lin for help throughout, and the reviewers for constructive comments. This material is based upon work supported by the National Science Foundation under grants 1116988 and 1526448.

References

Angulo, A., de Velasco, G.V.: Immersive simulation of architectural spatial experiences. Blucher Des. Proc. **1**(7), 495–499 (2014)

Azmandian, M., Grechkin, T., Rosenberg, E.S.: An evaluation of strategies for two-user redirected walking in shared physical spaces. In: 2017 IEEE Conference on Virtual Reality (VR), pp. 91–98. IEEE (2017)

Bowman, D.A., Koller, D., Hodges, L.F.: Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques. In: 1997 IEEE Conference on Virtual Reality Annual International Symposium, pp. 45–52. IEEE (1997)

Carbon, C.-C.: BiDimRegression: bidimensional regression modeling using R. J. Stat. Softw. Code Snippets **52**(3), 1–11 (2013)

Chance, S.S., Gaunet, F., Beall, A.C., Loomis, J.M.: Locomotion mode affects updating of objects encountered during travel: the contribution of vestibular and proprioceptive inputs to path integration. Presence **7**(2), 168–178 (1998)

Diwadkar, V.A., McNamara, T.P.: Viewpoint dependence in scene recognition. Psychol. Sci. **8**(4), 302–307 (1997)

Engel, D., Curio, C., Tcheang, L., Mohler, B., Bülthoff, H.H.: A psychophysically calibrated controller for navigating through large environments in a limited free-walking space. In: VRST 2008: Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology, pp. 157–164. ACM, New York (2008)

Grechkin, T.Y., Riecke, B.E.: Re-evaluating benefits of body-based rotational cues for maintaining orientation in virtual environments: men benefit from real rotations, women don't. In: Proceedings of the ACM Symposium on Applied Perception, SAP 2014, pp. 99–102. ACM, New York (2014)

Greenauer, N., Waller, D.: Micro-and macroreference frames: specifying the relations between spatial categories in memory. J. Exp. Psychol.: Learn. Mem. Cogn. **36**(4), 938 (2010)

Interrante, V., Ries, B., Anderson, L.: Seven league boots: a new metaphor for augmented locomotion through moderately large scale immersive virtual environments. In: IEEE Symposium on 3D User Interfaces, pp. 167–170 (2007)

Kelly, J.W., Avraamides, M.N., McNamara, T.P.: Reference frames influence spatial memory development within and across sensory modalities. In: Hölscher, C., Shipley, T.F., Olivetti Belardinelli, M., Bateman, J.A., Newcombe, N.S. (eds.) Spatial Cognition 2010. LNCS (LNAI), vol. 6222, pp. 222–233. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14749-4_20

Langbehn, E., Lubos, P., Bruder, G., Steinicke, F.: Application of redirected walking in room-scale VR. In: 2017 IEEE conference on Virtual Reality (VR), pp. 449–450. IEEE (2017)

Langbehn, E., Lubos, P., Steinicke, F.: Evaluation of locomotion techniques for room-scale VR: joystick, teleportation, and redirected walking. In: Proceedings of the Virtual Reality International Conference (VRIC) (2018). (accepted)

Marchette, S.A., Yerramsetti, A., Burns, T.J., Shelton, A.L.: Spatial memory in the real world: long-term representations of everyday environments. Mem. Cogn. **39**(8), 1401–1408 (2011)

McNamara, T.P.: How are the locations of objects in the environment represented in memory? In: Freksa, C., Brauer, W., Habel, C., Wender, K.F. (eds.) Spatial Cognition 2002. LNCS, vol. 2685, pp. 174–191. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45004-1_11

McNamara, T.P., Rump, B., Werner, S.: Egocentric and geocentric frames of reference in memory of large-scale space. Psychon. Bullet. Rev. **10**(3), 589–595 (2003)

Meilinger, T., Riecke, B.E., Bülthoff, H.H.: Local and global reference frames for environmental spaces. Q. J. Exp. Psychol. **67**(3), 542–569 (2014). PMID: 23972144

Mou, W., McNamara, T.P.: Intrinsic frames of reference in spatial memory. J. Exp. Psychol.: Learn. Mem. Cogn. **28**(1), 162–170 (2002)

Mou, W., Zhao, M., McNamara, T.P.: Layout geometry in the selection of intrinsic frames of reference from multiple viewpoints. J. Exp. Psychol.: Learn. Mem. Cogn. **33**(1), 145–154 (2007)

Paris, R., Joshi, M., He, Q., Narasimham, G., McNamara, T.P., Bodenheimer, B.: Acquisition of survey knowledge using walking in place and resetting methods in immersive virtual environments. In: Proceedings of the ACM Symposium on Applied Perception, SAP 2017, pp. 7:1–7:8. ACM, New York (2017)

Razzaque, S., Kohn, Z., Whitton, M.C.: Redirected walking. In: Eurographics Short Presentation (2001)

Riecke, B.E., Bodenheimer, B., McNamara, T.P., Williams, B., Peng, P., Feuereissen, D.: Do we need to walk for effective virtual reality navigation? Physical rotations alone may suffice. In: Spatial Cognition, pp. 234–247 (2010)

Roskos-Ewoldsen, B., McNamara, T.P., Shelton, A.L., Carr, W.: Mental representations of large and small spatial layouts are orientation dependent. J. Exp. Psychol.: Learn. Mem. Cog. **24**(1), 215–226 (1998)

Ruddle, R.A.: The effect of translational and rotational body-based information on navigation. In: Steinicke, F., Visell, Y., Campos, J., Lécuyer, A. (eds.) Human Walking in Virtual Environments, pp. 99–112. Springer, New York (2013). https://doi.org/10.1007/978-1-4419-8432-6_5

Ruddle, R.A., Lessels, S.: For efficient navigation search, humans require full physical movement but not a rich visual scene. Psychol. Sci. **17**(6), 460–465 (2006)

Ruddle, R.A., Lessels, S.: The benefits of using a walking interface to navigate virtual environments. ACM Trans. Comput.-Hum. Inter. **16**(1), 1–18 (2009)

Ruddle, R.A., Volkova, E., Bülthoff, H.H.: Walking improves your cognitive map in environments that are large-scale and large in extent. ACM Trans. Comput.-Hum. Interact. (TOCHI) **18**(2), 10 (2011)

Shelton, A.L., McNamara, T.P.: Systems of spatial reference in human memory. Cogn. Psychol. **43**(4), 274–310 (2001)

Steinicke, F., Bruder, G., Jerald, J., Frenz, H., Lappe, M.: Analyses of human sensitivity to redirected walking. In: Proceedings of the 2008 ACM symposium on Virtual Reality Software and Technology, pp. 149–156. ACM, New York (2008)

Steinicke, F., Bruder, G., Jerald, J., Frenz, H., Lappe, M.: Estimation of detection thresholds for redirected walking techniques. IEEE Trans. Vis. Comput. Graph. **16**, 17–27 (2010)

Tan, C.T., Leong, T.W., Shen, S., Dubravs, C., Si, C.: Exploring gameplay experiences on the oculus rift. In: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY 2015, pp. 253–263. ACM, New York (2015)

Tobler, W.R.: Bidimensional regression. Geogr. Anal. **26**(3), 187–212 (1994)

Waller, D.: Egocentric and nonegocentric coding in memory for spatial layout: evidence from scene recognition. Mem. Cogn. **34**(3), 491–504 (2006)

Williams, B., Narasimham, G., McNamara, T.P., Carr, T.H., Rieser, J.J., Bodenheimer, B.: Updating orientation in large virtual environments using scaled translational gain. In: Symposium on Applied Perception in Graphics and Visualization, Boston, MA, pp. 21–28 (2006)

Williams, B., Narasimham, G., Rump, B., McNamara, T.P., Carr, T.H., Rieser, J.J., Bodenheimer, B.: Exploring large virtual environments with an HMD when physical space is limited. In: Symposium on Applied Perception in Graphics and Visualization, Tübingen, Germany, pp. 41–48 (2007)

Williams, L.E.: Exploring large virtual environments with an hmd on foot when physical space is limited. Ph.D. thesis, Vanderbilt University, Nashville, TN (2007)

Xiao, C., Mou, W., McNamara, T.P.: Use of self-to-object and object-to-object spatial relations in locomotion. J. Exp. Psychol.: Learn. Mem. Cogn. **35**(5), 1137 (2009)

Xie, X., Lin, Q., Wu, H., Narasimham, G., McNamara, T.P., Rieser, J., Bodenheimer, B.: 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 2010, pp. 65–72. ACM, New York (2010)

Yamamoto, N., Shelton, A.L.: Path information effects in visual and proprioceptive spatial learning. Acta Psychol. **125**(3), 346–360 (2007)

Young, M.K., Gaylor, G.B., Andrus, S.M., Bodenheimer, B.: A comparison of two cost-differentiated virtual reality systems for perception and action tasks. In: Proceedings of the ACM Symposium on Applied Perception, pp. 83–90. ACM (2014)

Zhang, H., Mou, W., McNamara, T.P., Wang, L.: Connecting spatial memories of two nested spaces. J. Exp. Psychol.: Learn. Mem. Cogn. **40**(1), 191 (2014)