Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game

Ryan P. McMahan, Doug A. Bowman, David J. Zielinski, and Rachael B. Brady

Abstract—In recent years, consumers have witnessed a technological revolution that has delivered more-realistic experiences in their own homes through high-definition, stereoscopic televisions and natural, gesture-based video game consoles. Although these experiences are more realistic, offering higher levels of fidelity, it is not clear how the increased display and interaction aspects of fidelity impact the user experience. Since immersive virtual reality (VR) allows us to achieve very high levels of fidelity, we designed and conducted a study that used a six-sided CAVE to evaluate display fidelity and interaction fidelity independently, at extremely high and low levels, for a VR first-person shooter (FPS) game. Our goal was to gain a better understanding of the effects of fidelity on the user in a complex, performance-intensive context. The results of our study indicate that both display and interaction fidelity significantly affect strategy and performance, as well as subjective judgments of presence, engagement, and usability. In particular, performance results were strongly in favor of two conditions: low-display, low-interaction fidelity (representative of traditional FPS games) and high-display, high-interaction fidelity (similar to the real world).

Index Terms—Virtual reality, display fidelity, interaction fidelity, presence, engagement.

1 Introduction

Over the past decade, advancing technologies have introduced consumers to more-realistic experiences through higher levels of fidelity (the objective degree of exactness with which real-world experiences and effects are reproduced by a computing system [1]). Larger displays, higher resolutions, faster refresh rates, and stereoscopic capabilities have increased the *display fidelity* (the objective degree of exactness with which real-world sensory stimuli are reproduced) of home televisions. Similarly, the latest generation of video game systems (e.g., Nintendo Wii and Xbox Kinect) with their natural, gesture-based interactions have delivered increased levels of *interaction fidelity* (the objective degree of exactness with which real-world interactions can be reproduced). Despite the popularity of these technological advances, it is not completely clear how increased display fidelity and interaction fidelity impact the user experience of consumers.

In the virtual reality (VR) community, researchers have explored the effects of increasing fidelity by studying *immersion*. Slater et al. defined immersion as "a description of a technology" and discussed how display qualities and interaction mappings affect a system's fidelity [2]. In contrast, Bowman and McMahan defined immersion as "the objective level of sensory fidelity a VR system provides" and excluded interactions from their definition [3]. Additionally, some researchers have used the term "immersion" synonymously with "presence" (the psychological sense of "being there" [2]). To avoid confusion due to these incompatible uses of the term "immersion," and because our work focuses on the effects of a system's level of fidelity (similar to the first two definitions above), we choose to use the term "fidelity" instead of "immersion." Furthermore, we claim that the overall level of fidelity comes from a variety of system characteristics, and that a deep understanding of fidelity requires controlled evaluation of the effects of those different aspects of fidelity. Toward that end, we distinguish between display fidelity (sensory realism) and interaction fidelity (action realism).

- *Ryan P. McMahan is with Virginia Tech, E-Mail: rymcmaha@vt.edu.*
- *Doug A. Bowman is with Virginia Tech, E-Mail: bowman@vt.edu.*
- *David J. Zielinski is with Duke University, E-Mail: djzielin@duke.edu.*
- *Rachael B. Brady is with Duke University, E-Mail: rbrady@duke.edu.*

Manuscript received 15 September 2011; accepted 3 January 2012; posted online 4 March 2012; mailed on 27 February 2012.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

Some VR researchers have evaluated fidelity by comparing high-fidelity VR systems to low-fidelity desktop systems. For example, Gruchalla compared CAVE [4] and desktop versions of a well-path planning application and found that users performed significantly faster with the increased fidelity of the CAVE system [5]. Similarly, Arns *et al*. found that a CAVE version of a statistical data application significantly improved accuracy when compared to a desktop version [6]. Ruddle *et al*. compared a head-mounted display (HMD) to a desktop for navigating large-scale virtual environments (VEs) and found that the increased fidelity of the HMD allowed users to navigate the VEs significantly faster [7]. However, in these and many other prior studies on the effects of fidelity, display fidelity and interaction fidelity were confounded, making it difficult to distinguish the components contributing to any significant effects.

Considering the importance of continuing to explore the effects of increasing fidelity, we designed and conducted a study to independently evaluate both display and interaction fidelity at extremely high and low levels, which yielded four experimental conditions. We used a six-sided CAVE with wireless tracking capabilities to provide the extremely high levels of display and interaction fidelity. To gain a better understanding of the effects of fidelity on the user experience, particularly performance, we chose to use a virtual reality first-person shooter (FPS) game as a complex, performance-intensive context for our study. We also evaluated the effects of fidelity on subjective responses, such as presence [8], engagement [9], and usability.

After describing the details of our experimental design, we provide analyses of the results of our study, which indicate that both display fidelity and interaction fidelity have significant effects on the user experience. We discuss how performance results strongly favor two of the four conditions – low-display, low-interaction fidelity and high-display, high-interaction fidelity – both of which leverage familiar experiences (i.e., a traditional FPS game and the real world, respectively). Overall, our study contributes to a better scientific understanding of the effects of display fidelity and interaction fidelity while addressing some practical concerns of choosing appropriate displays and interaction techniques.

2 Related Work

In addition to the previously mentioned practical evaluations of fidelity, which compared contrasting systems [5-7], researchers have also evaluated specific components of fidelity. Stereoscopy has been demonstrated to significantly reduce time and errors for path-tracing

1077-2626/12/$31.00 © 2012 IEEE Published by the IEEE Computer Society

tasks [10], in addition to improving user performance for spatial understanding tasks [11]. Increased field of view (FOV) has been shown to significantly improve user performance for search tasks [12, 13], comparison tasks [13], and walking tasks [12]. Larger display sizes [14] and faster frame rates [15] have also resulted in better user performance for certain types of tasks.

In addition to components of display fidelity, researchers have also evaluated specific interaction aspects. Pausch *et al*. compared natural head tracking to hand-based viewpoint control and determined that the higher-fidelity interaction improved user performance when searching for non-present targets [16]. Similarly, interaction techniques with higher degrees-of-freedom (DOF) have been shown to outperform techniques based on 2-DOF input (e.g., a mouse) for 3D object manipulations [17] and rotations [18]. On the other hand, high-fidelity steering techniques have been shown to significantly reduce user performance when compared to low-fidelity, joystick-based techniques [19].

Going beyond evaluating single aspects of fidelity, some researchers have conducted systematic, multivariate evaluations of fidelity. Stereoscopy (a component of display fidelity) and head tracking (a component of interaction fidelity) have been evaluated together in several studies, with results indicating that both have significant effects on spatial understanding tasks [20, 21] but not necessarily object manipulation tasks [22, 23]. Field of regard (FOR), the total size of the visual field (in degrees of visual angle) surrounding the user [3], has been systematically evaluated together with various interaction aspects, including head tracking [24] and 3D manipulation techniques [25], but has not been shown to be a statistically significant factor in these studies.

Most of the prior multivariate evaluations of fidelity were limited by the low- and mid-range VR systems used, which offered less than a full 360-degree FOR. This could be the reason FOR has not been shown to be a significant factor. In addition, most of these experiments studied only a single task in isolation and only gathered performance metrics such as speed and accuracy.

3 Experiment

The goal of our experiment was to evaluate the independent and combined effects of display fidelity and interaction fidelity. To address the limitations of the prior work, we decided to conduct a systematic, multivariate evaluation using a VR system that offered a full 360-degree FOR. We also chose to study fidelity in the context of a performance-intensive application with a diversity of user tasks, and to gather data on many aspects of the user experience.

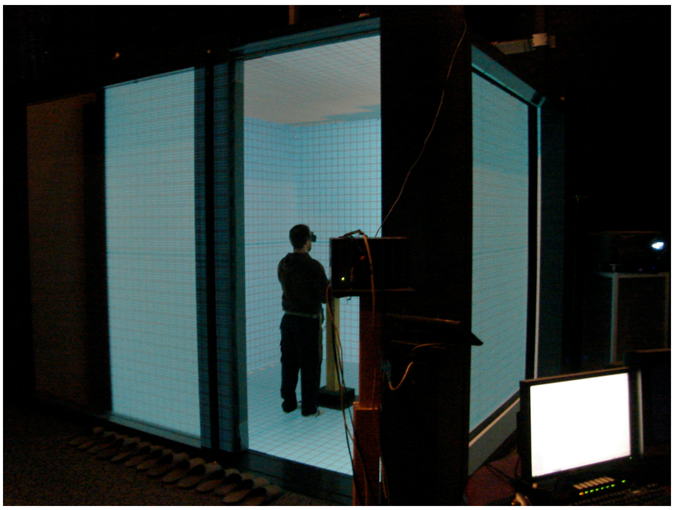

Fig. 1. We used a six-sided CAVE to evaluate the independent and combined effects of display fidelity and interaction fidelity.

3.1 Apparatus

In order to evaluate extremely high and low levels of display fidelity, we used a six-sided CAVE (seen in Fig. 1), which offered a full 360-degree FOR. This rear-projected, cube-shaped display system measured 3m x 3m x 3m and had a display resolution of 1050 x 1050 pixels on each screen. CrystalEyes technology was used for active stereoscopy. We used Syzygy [26] to control the system's master/slave framework, which consisted of a master-node computer and six supporting computers (one per screen).

We used the CAVE's built-in InterSense IS-900 Wireless tracking system with a 6-DOF head tracker and a 6-DOF wand to provide input for high levels of interaction fidelity. We also used a Bluetooth three-button mouse and a Bluetooth standard keyboard for our low level of interaction fidelity. We set both of these atop a 106cm tall podium equipped with a turntable, 46cm in diameter, to afford physical turning with these low-fidelity input devices. We maintained a constant, moderate mouse sensitivity to avoid variability between our participants. Mouse latency was likely lower than wand latency, but no participants commented on this difference, and we do not believe it affected the results.

3.2 Experimental Design

We wished to evaluate very high and very low levels of both display and interaction fidelity. To increase experimental control and reduce confounds between the high and low levels, we adopted a systematic approach that utilizes a CAVE system to control for confounds while investigating specific components of fidelity [3]. Both independent variables – display fidelity and interaction fidelity – had two levels and were varied within subjects. The presentation order of the four conditions was counterbalanced between subjects.

3.2.1 Components of Interest

For display fidelity, we decided to evaluate the components of stereoscopy and FOR while controlling other components such as FOV, resolution, and frame rate. We chose to evaluate stereoscopy because it has been shown to be a significant factor in prior research [10, 11]. Since we were using a six-sided CAVE for our experiment, we also chose to evaluate FOR, which had not often demonstrated significant effects in VR systems with lower FOR [24, 25]. Hence, our high level of display fidelity used stereoscopic graphics and a full 360-degree FOR (i.e., all six sides of the CAVE). In contrast, our low level of display fidelity involved non-stereoscopic graphics and only a 90-degree FOR (i.e., a single wall of the CAVE).

For interaction fidelity, we decided to focus on the two most important FPS interactions – aiming and locomotion, removing other interactions such as crouching and picking up objects. For aiming, we chose to compare a traditional FPS mouse technique to natural 3D pointing using the 6-DOF handheld wand with its ergonomic trigger button. For locomotion, we wanted to compare a traditional FPS keyboard technique to a more natural, high-fidelity locomotion technique. We designed and implemented a new technique for this purpose (see section 3.2.2). Therefore, for our high level of interaction fidelity, users would aim and fire with the handheld controller while physically moving to virtually travel. In contrast, for our low level of interaction fidelity, users would use the mouse to turn, aim, and fire while using the keyboard to travel through the virtual world.

3.2.2 The Human Joystick Technique

As mentioned in the previous section, we wanted to compare a keyboard technique to a more natural, high-fidelity locomotion technique. Unfortunately, real walking [27] – the highest fidelity locomotion technique – was not feasible since users would be limited to the space provided by our CAVE system and our FPS game would involve a much larger virtual space. For similar reasons, redirected walking [28] was not practical to implement for our CAVE-based FPS. We considered various walking-in-place techniques ranging from head tracking [2] to leg tracking [29] to the shadow walking technique [30]. But we decided against walking-in-place for our high-fidelity locomotion technique to avoid fatigue and Type I errors (when the system judges users to be walking in place when they are not [27]) due to the performance-intensive nature of our FPS context.

Thus, we designed a technique similar to Bourdot and Touraine's navigation paradigm [31] called the *human joystick*. By capturing the 2D horizontal vector from the center of the CAVE to the user's tracked head position, and utilizing it as a joystick's 2D vector would be used for locomotion, we essentially turned the user into a giant, human joystick (see Fig. 2). Since only the user's head position is used to calculate the vector, the direction the user is facing has no influence on the locomotion, making the human joystick an omni-directional locomotion technique that allows movement in any horizontal direction.

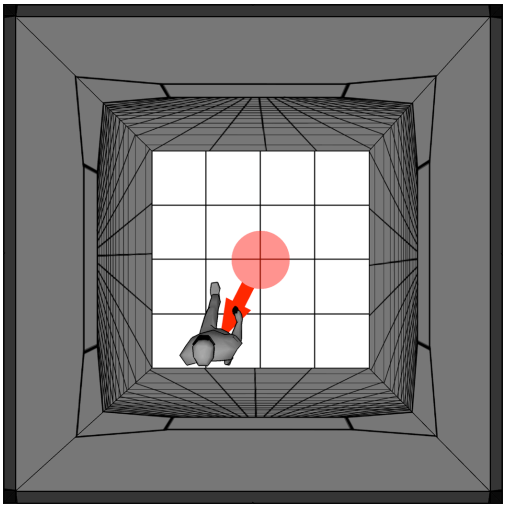

Fig. 2. Top-down illustration of the CAVE displaying a large-scale virtual environment. The human joystick technique utilizes the user's tracked head position from the center of the CAVE for virtual locomotion, as a joystick's 2D vector would be used.

To avoid constant virtual locomotion due to small distances between the user's head position and the center of the CAVE, we also included a "neutral zone" with a 20cm radius at the center of the CAVE. While in the neutral zone, one-to-one head tracking provides the user the ability to make minor changes to the viewpoint, such as peering around a corner. Once outside of the neutral zone, the human joystick technique is activated, and the user is virtually translated. The speed of virtual locomotion is linearly related to the user's distance from the neutral zone with a maximum speed of 5cm per frame near the walls of the CAVE, which is equivalent to our keyboard technique's maximum speed.

After an informal usability study of the human joystick, we adopted it as our high-fidelity locomotion technique for our experiment. Additionally, we decided to position a small floor mat in the center of the CAVE to provide a haptic representation of the neutral zone, which would be particularly important in our low-display-fidelity conditions without floor projection.

3.2.3 Details of Experimental Conditions

In this section, we discuss the details of our four within-subjects conditions and note issues inherent to simultaneously evaluating display and interaction fidelity at high and low levels.

High-display, high-interaction (HDHI): This condition was the most straightforward combination of a level of display fidelity with a level of interaction fidelity. Surrounded by six stereoscopic CAVE sides, the user uses the human joystick technique to move virtually through the world while using the 6-DOF wand to point the weapon crosshair into the 3D environment (see Fig. 3). Due to the 360-degree FOR, the user simply made physical body rotations to turn (i.e., rotate the viewpoint).

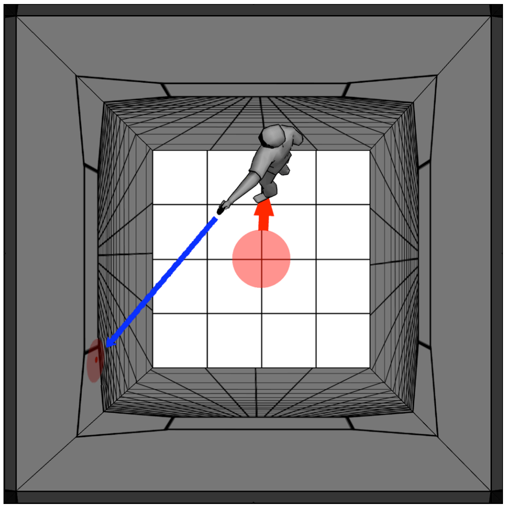

Fig. 3. In the high-display, high-interaction (HDHI) condition, the user used the human joystick technique for virtual locomotion and a 6-DOF wand to control the crosshair for 3D aiming and firing at all six sides of the CAVE.

High-display, low-interaction (HDLI): In most FPS games, the crosshair remains centered on the display screen while the mouse rotates the player's viewpoint to aim in different directions, including upward and downward. This was not a feasible solution for this condition for two reasons. First, the ability to pitch the VE upward and downward would become disorienting since the user would be physically standing in the space and at times could appear to be levitating parallel to the virtual ground. Second, virtual rotations are known to cause simulator sickness for some people [32], and constant virtual rotations would likely cause many users to become ill. To remedy this, we designed the mouse to control the movement of the crosshair across the surrounding display screens, similar to a cursor on a multi-monitor desktop, except with no boundaries and continuous capability.

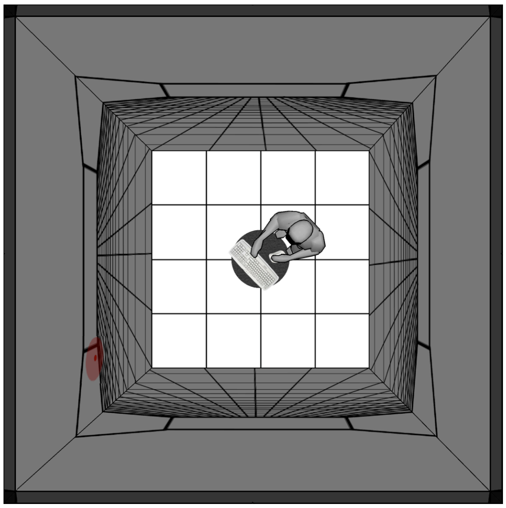

Fig. 4. In the high-display, low-interaction (HDLI) condition, the user used the keyboard for virtual locomotion and the mouse for aiming and firing at all six sides of the CAVE.

A question that arose from this choice was what direction the keyboard locomotion keys would align to. With a single screen, it is intuitive that the up arrow or W key would map to a motion toward the screen, but with surrounding screens, it makes little sense to always map these commands to motion toward the "front" wall of the CAVE, since users are constantly turning to face different screens. Instead, we interpreted keyboard commands relative to the 2D horizontal direction of the crosshair from the center of the

CAVE. Hence, if the player positions the crosshair on the right wall of the CAVE, the up arrow and W key now activate motion towards the right side instead of the front. We equipped our podium with the turntable to allow the user to face the crosshair at all times.

Hence, as depicted in Fig. 4 of the HDLI condition, the user used the mouse to move the crosshair among the six stereoscopic CAVE sides, presumably physically turning with the turntable at the same time to face the crosshair, while using the arrow or WASD keys to move relative to it.

Low-display, high-interaction (LDHI): This condition was very similar to the HDHI condition except that physical body rotations no longer sufficed for turning to see the rest of the environment due to the reduced FOR. Instead, we had to provide a technique for *virtual turning*. In many CAVE applications, the wand's joystick is used to enable virtual turning, but for this experiment, we did not consider this a suitable technique for two reasons. First, it increased the number of physical actions required by the user to fully interact, hence increasing the cognitive load on the user. Second, the low-display, low-interaction condition would not be using the wand device, and, therefore, virtual turning would be confounded between these two conditions.

With this in mind, we decided to activate virtual turning when the crosshair moves within five degrees of the left and right edges of the single CAVE wall. The activating edge determined the direction of the rotation while the rotation speed was linearly related to the distance between the crosshair and the edge, with a maximum rotation speed of 2.5 degrees once the crosshair was at the edge or off-screen.

As seen in Fig. 5, the user still used the human joystick technique to move through the world in any direction while pointing the wand toward the front CAVE wall. Virtual turning was provided by pointing the crosshair near or off the edges of the display screen. The graphics were non-stereoscopic, and the projectors for the other CAVE sides were shuttered to create the reduced FOR.

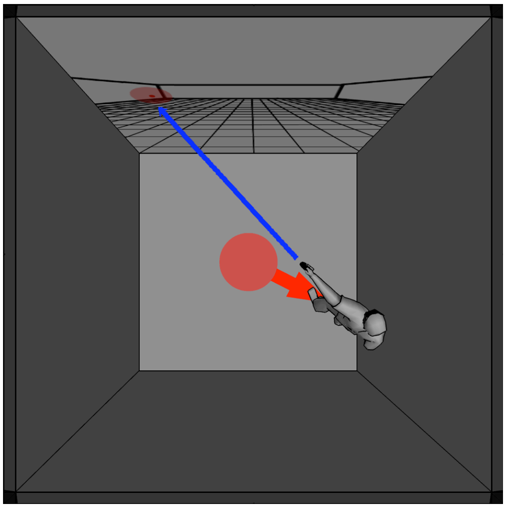

Fig. 5. In the low-display, high-interaction (LDHI) condition, the user used the human joystick technique for virtual locomotion and a 6-DOF wand for 3D aiming and firing at the front wall of the CAVE.

Low-display, low-interaction (LDLI): As in the HDLI condition, we designed the mouse in the LDLI condition to control the movement of the crosshair except that, like the LDHI condition, the crosshair also activates virtual turning when near to or past the left and right edges of the front CAVE wall. As depicted in Fig. 6, the user moved the crosshair around on the front CAVE wall with the mouse, virtually turning by moving the crosshair off-screen, and moved relative to the crosshair by using the arrow or WASD keys. For this condition, the graphics were non-stereoscopic, and the other projectors were shuttered.

3.3 First-Person Shooter Task

As previously mentioned, we chose a first-person shooter game for the context of our experiment. Most FPS games require complex interactions, such as maneuvering around obstacles while shooting at enemies. These complex interactions are performance intensive and require high levels of peripheral awareness and spatial understanding to avoid enemies and obstacles. Hence, we expected that our FPS context would provide greater potential for significant differences, especially with regard to the level of interaction fidelity.

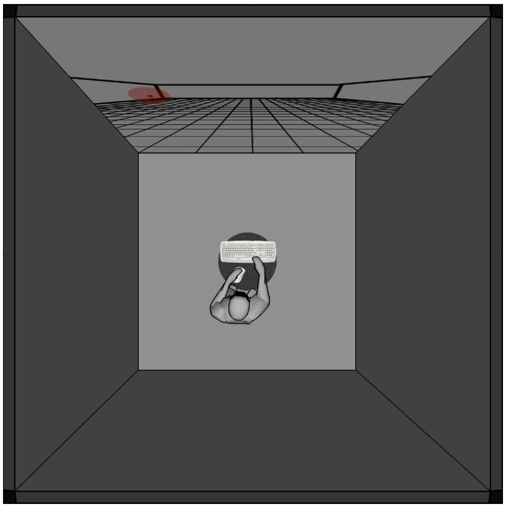

Fig. 6. In the low-display, low-interaction (LDLI) condition, the user used the keyboard for virtual locomotion and the mouse for aiming and firing at the front wall of the CAVE.

After investigating various open-source FPS engines, we eventually used Syzygy's Quake III Arena map viewer called "cubecake". In order to maintain experimental control, we developed our own artificial intelligence (AI) routines, designed our own maps, and tracked various player statistics. Since FPS games are often regarded as inherently violent and believed by some to invoke hostility and aggressiveness [33], we decided to use inanimate "androids" as bots (i.e., enemies) and avoided special effects like blood, to reduce the emotional risks of participating. In all conditions, our FPS game ran at an average of 55 frames per second.

To maintain more control during our experiment, we decided to use simple maps with only a single path from start to finish, instead of using sprawling maps with multiple routes to choose from. In particular, we wanted to use these linear layouts to learn more about our components of interest (stereoscopy, FOR, aiming, and locomotion) by designing map sections that would exercise these components individually and in combination. To exercise stereoscopy, we chose to have bots appear (or "teleport") within three meters of the player since stereoscopy is more effective at close viewing distances [34]. To exercise increased FOR, we decided to have bots teleport around the player's location in many different directions. For aiming, we had several bots teleport in at once to emphasize the ability to quickly change targets. Finally, for locomotion, we used a retreating AI behavior to force the player to chase the bot by moving.

There were ten total map sections, each with a distinct purpose. In order to maintain the purpose of each section despite player movements and actions, we designed "computer station" game elements that players were required to "hack" (i.e., stand near). These elements allowed us to control player locations when bots appeared. In each section, players were required to eliminate eight bots before the entrance to the next section opened. We also used section entrances as respawn locations if players died within the section. The list below details the purpose of each map section with a brief description.

1. ***None:*** Bots teleport in one at a time, more than 6m away in a single direction.
2. ***Stereoscopy:*** Bots teleport in one at a time, within 3m in a single direction.
3. ***FOR:*** Bots teleport in one at a time, more than 6m away in a surrounding fashion.
4. ***Aiming:*** All eight bots teleport in at once, more than 6m away in a single direction.
5. ***Locomotion:*** Bots teleport in one at a time, more than 6m away in a single direction, and retreat when hit.
6. ***Stereoscopy + Locomotion:*** Bots teleport in one at a time, within 3m in a single direction, and retreat when hit.
7. ***FOR + Locomotion:*** Bots teleport in one at a time, more than 6m away in a surrounding fashion, and retreat when hit.
8. ***Stereoscopy + Aiming:*** All eight bots teleport in at once, within 3m in a single direction.
9. ***FOR + Aiming:*** All eight bots teleport in at once, more than 6m away in a surrounding fashion.
10. ***Stereoscopy + FOR + Aiming + Locomotion:*** All eight bots teleport in at once, within 3m in a surrounding fashion, and retreat when hit.

3.4 User Experience Metrics

Concerned with the effects of display and interaction fidelity on the user experience, we gathered a broad range of metrics related to objective user performance and subjective judgments of presence, engagement, and usability. For user performance, we measured several objective metrics per section: completion time, damage taken, accuracy, and headshot count. We did not track the number of enemy deaths as each section involved eliminating eight bots.

To measure perceptions of presence, we administered the Slater-Usoh-Steed (SUS) Presence Questionnaire [8] after each condition. Similarly, to measure engagement, we used a modified Game Engagement Questionnaire (GEQ) [9] after each condition (see Appendix A). For usability and preferences, we developed our own usability questionnaire consisting of seven-point Likert-scale items, which we also administered after each condition (see Appendix B).

3.5 Procedure

Once recruited, each participant was required to sign an informed consent form and fill out a background survey, which collected data about their gaming and firearm experiences in addition to general demographic information. After the background survey, we administered a spatial orientation test [35] to analyze the spatial abilities of our participants.

In the next phase of the procedure, each participant experienced a high-fidelity VR simulation of a kitchen for five minutes. In prior VR studies involving performance tasks, we had observed participants performing poorly in initial tasks due to being engrossed by the high-fidelity VR. Hence, we hoped this VR exposure would eliminate those "wow-factor" situations.

After the VR exposure phase, participants proceeded through the four experimental conditions, using the order of their assigned permutation. We began each condition with a training session, in which the experimenter would explain how to interact in the given condition and allow the participant to practice with a small five-section map with bots. After the practice session, participants were instructed to play through a ten-section map as quickly as possible while avoiding damage and maintaining high accuracy. Afterwards, we gave the participant the presence, engagement, and usability questionnaires. Participation concluded after the fourth condition and lasted for approximately 120 minutes, including scheduled breaks.

3.6 Participants

To balance ordering effects, we recruited 24 unpaid participants (23 males, 1 female), one for each permutation of the four conditions. A 25th participant was recruited but quit during the experiment due to simulator sickness, which was the only case of any degree of simulator sickness observed during the study. The age range of the 24 participants was 18 to 26 years old with a mean age of 20. Using their background survey data, we calculated an FPS-expertise score for each participant by adding the number of hours they played FPS games in the week prior to participation, the average number of hours they played FPS games per week, and the number of FPS games they had ever completed or "beaten". These calculations yielded expertise scores ranging from 0 to 33, with a mean of 12.69. We used these scores during analysis to determine if FPS expertise had a significant effect on our results.

4 Results

4.1 Objective Metrics

For overall completion times, we performed a two-way, repeated-measures ANOVA and determined that neither display fidelity nor interaction fidelity had a significant main effect, but that there was a significant interaction between the two ($F(1, 23) = 82.3503$, $p < 0.0001$). Post hoc comparison using the Tukey HSD test indicated that the LDLI and HDHI conditions were significantly faster than the LDHI and HDLI conditions (see Fig. 7).

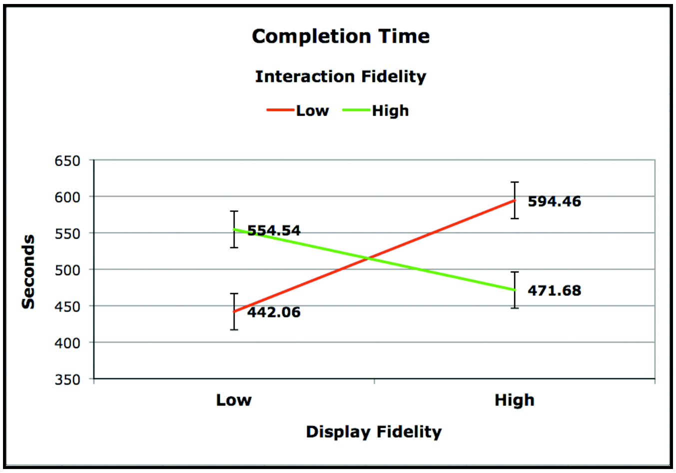

Fig. 7. The LDLI and HDHI conditions were significantly faster than the two "mixed fidelity" conditions for completion time.

For the total damage taken, we performed another two-way, repeated-measures ANOVA and found that the level of interaction fidelity had a significant effect ($F(1, 23) = 71.2675$, $p < 0.0001$), with high interaction performing better than low interaction. Display fidelity did not have a significant effect though there was a significant interaction between the two variables ($F(1, 23) = 62.0083$, $p < 0.0001$). Based on a Tukey HSD post-hoc test, the HDHI condition was significantly the best and the HDLI condition was significantly the worst, while the two low-display conditions were not significantly different from each other (see Fig. 8).

For overall accuracy, another two-way, repeated-measures ANOVA indicated that both display fidelity and interaction fidelity had significant effects. For display fidelity ($F(1, 23) = 10.0048$, $p = 0.0043$), low display ($M = 48.18\%$) provided significantly better accuracy than high display ($M = 45.13\%$). Similarly, for interaction fidelity ($F(1, 23) = 14.3572$, $p = 0.0009$), low interaction ($M = 49.60\%$) was significantly more accurate than high interaction ($M = 43.72\%$). For accuracy, there was no significant interaction between the two aspects of fidelity.

With regard to the total number of headshots, we did not find a significant effect of either display fidelity or interaction fidelity, based on another two-factor ANOVA. Additionally, there was not a significant interaction between the two.

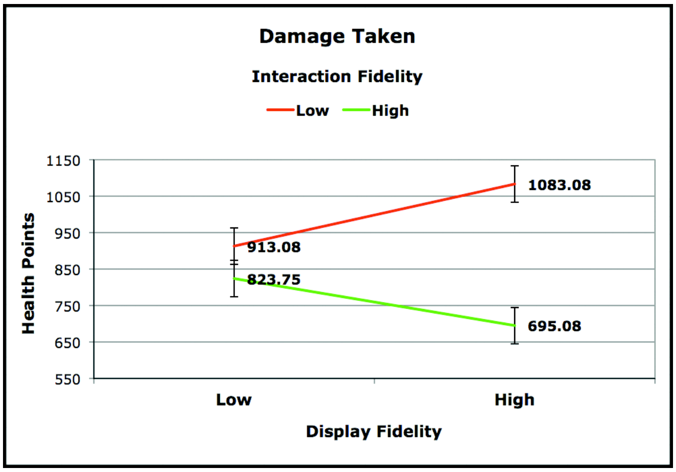

Fig. 8. The HDHI condition was significantly the best for avoiding damage while the HDLI condition was significantly the worst.

To determine if stereoscopy was likely to have had a significant effect within our experiment, we conducted one-way, repeated-measures ANOVAs (display fidelity) on the objective metrics collected from the map sections (2, 6, 8, and 10) designed to exercise stereoscopy. Display fidelity did not have a significant effect for any of the four metrics. Contrastingly, we conducted one-way, repeated-measures ANOVAs (display fidelity) on the FOR sections (3, 7, 9, and 10) and found that display fidelity had significant effects on completion times ($F(1, 23) = 34.3228$, $p < 0.0001$), damage taken ($F(1, 23) = 6.5804$, $p = 0.0173$), and accuracy ($F(1, 23) = 21.8265$, $p = 0.0001$), with high display performing significantly worse than low display in all cases.

Similarly, we conducted one-way, repeated-measures ANOVAs (interaction fidelity) on the objective metrics collected from the sections designed to exercise aiming (4, 8, 9, and 10) and found that interaction fidelity had significant effects on completion times ($F(1, 23) = 4.7341$, $p = 0.0401$) and damage taken ($F(1, 23) = 33.8057$, $p < 0.0001$), with high interaction performing significantly better than low interaction in both cases. In the locomotion sections (5, 6, 7, and 10), we determined that interaction fidelity had significant effects on damage taken ($F(1, 23) = 48.7039$, $p < 0.0001$) and accuracy ($F(1, 23) = 5.8290$, $p = 0.0241$), with high interaction performing better for avoiding damage but worse for accurate firing.

To determine if participants' backgrounds had any significant effects on our results, we computed Pearson correlation coefficients to assess the relationships between our objective metrics and participants' spatial abilities, FPS expertise, and firearm expertise. We found a positive correlation between spatial ability and the number of headshots ($r = 0.3750$, $p = 0.0002$), but a negative correlation between spatial ability and overall accuracy ($r = -0.2241$, $p = 0.0282$). We also found a positive correlation between firearm expertise and total headshots ($r = 0.3050$, $p = 0.0025$). Interestingly, we did not find any correlations between the FPS expertise of participants and our objective metrics.

4.2 Subjective Metrics

For presence, we performed a two-way, repeated-measures ANOVA (display fidelity and interaction fidelity) and determined that both variables had a significant effect on SUS presence scores. For display fidelity ($F(1, 23) = 27.4669$, $p < 0.0001$), high display ($M = 2.7500$) provided significantly more presence than low display ($M = 1.3333$). Similarly for interaction fidelity ($F(1, 23) = 41.9552$, $p < 0.0001$), the high level of interaction ($M = 3.0000$) provided significantly more presence than the low level ($M = 1.0833$). There was a significant interaction between our two variables for presence ($F(1, 23) = 28.2273$, $p < 0.0001$). Based on a Tukey HSD post-hoc test, the HDHI condition provided significantly more presence than the other three conditions.

For engagement, we performed another two-way, repeated-measures ANOVA and determined that both display fidelity ($F(1, 23) = 16.4229$, $p = 0.0005$) and interaction fidelity ($F(1, 23) = 37.8723$, $p < 0.0001$) had significant effects on the modified-GEQ scores. High display ($M = 47.7917$) engaged the participants significantly more than low display ($M = 43.3958$), and the high level of interaction ($M = 49.7708$) provided significantly more engagement than the low level ($M = 41.4167$). Again, there was a significant interaction between our two variables for engagement ($F(1, 23) = 9.5299$, $p = 0.0052$). Post hoc comparison using the Tukey HSD test indicated that the HDHI condition was judged as significantly more engaging than the other three conditions.

For the usability judgments, another two-way, repeated-measures ANOVA indicated that both display fidelity ($F(1, 23) = 4.4756$, $p = 0.0454$) and interaction fidelity ($F(1, 23) = 38.0305$, $p < 0.0001$) had significant effects on total usability scores. Participants judged high display ($M = 59.5000$) as significantly more usable than low display ($M = 56.2292$), and the high interaction ($M = 62.9583$) was judged as significantly more usable than low interaction ($M = 52.7708$). There was a significant interaction between the two variables ($F(1, 23) = 70.3930$, $p < 0.0001$). Post hoc comparison using the Tukey HSD test indicated that the HDHI condition was judged as significantly more usable than the other three conditions.

5 Discussion

Based on the results, our observations, and the comments of participants, we have drawn four inferences from our systematic evaluation of display and interaction fidelity.

5.1 Level of Fidelity Impacted Strategy

During our study, we observed that both display fidelity and interaction fidelity seemed to affect the strategies of our participants. Depending on the level of display fidelity, users switched between two strategies for firing at targets. In the high-display conditions, most participants would "spray" gunfire while moving the crosshair towards an enemy (participants were given an unlimited amount of ammunition). In the low-display conditions, however, participants usually lined up the crosshair with the enemy before firing. The significant effect display fidelity had on overall accuracy supports these observations, as low display fidelity was more accurate than high display fidelity. One important consideration about these observed strategies is that in the high-display conditions, participants were able to see their enemies using peripheral vision before moving the crosshair, while in the low-display conditions, participants often would not see their enemies until after virtually turning.

We observed a similar difference in strategies related to interaction fidelity. In the high-interaction conditions, participants tended to move more using the human joystick technique than they moved with the keyboard technique in the low-interaction conditions. This difference in strategies resulted in participants taking less damage and being less accurate with high interaction fidelity while taking more damage and being more accurate with the low level of interaction fidelity. Our analyses concerning damage taken and overall accuracy support both of these observations. One possible explanation for these choices in strategies is that the human joystick technique requires physical movement back to the neutral zone to stop any current virtual movement while the keyboard technique only requires the user to stop pressing the arrow or WASD keys.

5.2 Familiarity Improved Performance

A key lesson we took from our systematic evaluation of fidelity is that familiarity improved user performance. Despite being the two extreme combinations of display and interaction fidelity, and despite affording contrasting user strategies, the LDLI and HDHI conditions outperformed the other two "mixed fidelity" conditions with regard to our objective metrics. For completion times, both of these conditions were significantly faster than the LDHI and HDLI conditions. For damage taken, the HDHI condition significantly outperformed the others. Similarly, the LDLI condition provided the best accuracy due to its combination of low display and low interaction, both of which were significantly better than their higher counterparts for accuracy. We saw these results despite the fact that participants were trained on each condition and practiced before completing the actual trials.

Based on our observations and comments from participants, we attribute the excellent performance of the LDLI and HDHI conditions to familiarity. Conceptually, the LDLI condition was very similar to a standard, desktop FPS game with the participants using a mouse and keyboard to interact with a low FOR (though desktop games normally use software FOVs different from the physical FOV of the display). Several participants commented that this condition reminded them of playing a desktop FPS. The HDHI condition was similar to interaction in the real world. Participants were expected to hold and point the 6-DOF wand like they would a real weapon, they turned physically to face different directions in the virtual world, and they were afforded some physical locomotion through the use of the human joystick technique. Several participants also commented on how "realistic" the HDHI condition was for them.

In contrast, it is difficult to draw analogies between the two mixed conditions and anything users are familiar with. For instance, the HDLI condition is uncommon because most higher-end display systems come with capabilities to provide higher-fidelity interactions through tracking systems. There might be slight similarities between the LDHI condition and some current motion-controlled video game systems, which would explain why it fared better than the HDLI condition. In particular, the ability to physically point at a target can be found in many newer video games, but the ability to physically move at the same time is less common.

5.3 High Fidelity Increased Presence, Engagement, and Usability

Our study also showed that high levels of either display fidelity or interaction fidelity increased the positive subjective responses of users to the VR system. For presence, engagement, and usability, we found that both display and interaction fidelity had significant positive effects. More importantly, we found that users had the greatest senses of presence, engagement, and usability with the high-display, high-interaction condition. Designers concerned with achieving high levels of presence, engagement, or usability should consider that higher levels of fidelity may be most suitable. In particular, high levels of both display fidelity and interaction fidelity appear to combine for the best results.

5.4 Effects of Fidelity Components Were Nuanced

Despite prior studies demonstrating the significant effects of stereoscopy [10, 11], the results of our evaluation imply that stereoscopy did not have a significant effect on our FPS task. Even though we specifically designed map sections to exercise this display-fidelity component, we found no significant effects of display fidelity on performance in these sections. The most likely reason for this lies in the nature of the FPS aiming task. Regardless of the level of interaction fidelity, a user positions an always-visible crosshair at a human-sized target and then proceeds to fire. Obviously, when the target is far away from the user, stereoscopy is not going to have a significant effect, but even when the target is within 3m, the target appears much larger for the purpose of aiming and stereoscopy is not necessary.

On the other hand, we did find evidence that a component of display fidelity (FOR) and components of interaction fidelity (aiming and locomotion realism) had significant effects on performance. Although we need further evaluation to understand these effects in detail, we can infer that the influences of display and interaction fidelity components are not universal, but have subtle nuances. They are likely to be dependent on the specific application context, tasks, user strategies, and levels of fidelity.

6 Conclusions and Future Work

Although consumers have witnessed a technological revolution geared towards more-realistic experiences in recent years, we have demonstrated that there is still much to be learned about the effects of increasing a system's fidelity to the real world. Using a six-sided CAVE and a performance-intensive FPS game, we systematically evaluated extremely high and low levels of display and interaction fidelity to gain a better understanding of their effects on the user experience.

The results of our study show that the levels of display and interaction fidelity can be significant factors in determining performance, presence, engagement, and usability. Combined with existing results in the literature, we have contributed to the overall understanding of the effects of fidelity in two important ways. First, we have shown that increased display fidelity can often have positive effects on the user experience (the negative effect of display fidelity on accuracy in this experiment appears to be due to the display's influence on user strategy). Second, the combination of display fidelity and interaction fidelity can determine the familiarity of the overall system, and it is this familiarity that seems to determine overall performance in many cases.

For future work, we plan to conduct additional systematic evaluations of display and interaction fidelity. In particular, we plan to study FOR, aiming, and locomotion independently in the same FPS context, since those components of interest had potential significant effects on the results of this evaluation. Additionally, considering our discovery that fidelity influences user strategy, we plan to conduct some studies of individual FPS subtasks to reduce the variance caused by differing strategies, in order to learn more about the effects of fidelity on user performance. We will use these future studies and the results of this study to produce guidelines helping designers to choose displays and interaction techniques with appropriate levels of fidelity based on the desired user experience.

Appendix A: Modified Engagement Questionnaire

Directions: For each of the following, please rate your experience of the sensation while playing the game, on the following scale from 1 (did not experience) to 5 (definitely experienced).

1. I lost track of time.
2. Things seemed to happen automatically.
3. I felt different.
4. I felt scared.
5. The game felt real.
6. I felt tense.
7. Time seemed to stand still or stop.
8. I felt unaware of my surroundings.
9. Playing seemed automatic.
10. My thoughts were fast.
11. I forgot I was in a virtual environment.
12. I played without thinking about how to play.
13. Playing made me feel calm.
14. I really got involved with the game.
15. I did not want to stop playing.

Appendix B: Usability Questionnaire

1. Rate how easy it was for you to look for the androids (1=extremely difficult, 7=extremely easy).
2. Rate how easy it was for you to fire and hit the androids.
3. Rate how easy it was for you to move around the space ship.
4. Rate how easy it was for you to distinguish close objects from far objects.
5. Rate how easy it was for you to play the game in general.
6. Rate how natural it was for you to look for the androids (1=extremely unnatural, 7=extremely natural).
7. Rate how natural it was for you to fire and hit the androids.

8. Rate how natural it was for you to move around the space ship.
9. Rate how natural it was for you to distinguish close objects from far objects.
10. Rate how natural it was for you to play the game in general.
11. Rate how much fun you had playing the game (1=extremely frustrating, 7=extremely fun).
12. Rate how tiring it was for you playing the game (1=extremely exhausting, 7=not tiring at all).

ACKNOWLEDGMENTS

The authors wish to thank their participants. This work was supported in part by a grant from the Office of Naval Research.

REFERENCES

[1] S.J. Gerathewohl, "Fidelity of Simulation and Transfer of Training: A Review of the Problem," Federal Aviation Administration Office of Aviation Medicine Report, 1969.
[2] M. Slater, M. Usoh, and A. Steed, "Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 2, no. 3, pp. 201-219, 1995.
[3] D.A. Bowman and R.P. McMahan, "Virtual Reality: How Much Immersion Is Enough?," *IEEE Computer*, vol. 40, no. 7, pp. 36-43, 2007.
[4] C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti, "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE," *Proceedings of Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 135-142, 1993.
[5] K. Gruchalla, "Immersive Well-Path Editing: Investigating the Added Value of Immersion," *Proceedings of IEEE Virtual Reality (VR)*, pp. 157-164, 2004.
[6] L. Arns, D. Cook, and C. Cruz-Neira, "The Benefits of Statistical Visualization in an Immersive Environment," *Proceedings of IEEE Virtual Reality (VR)*, pp. 88-95, 1999.
[7] R.A. Ruddle, S.J. Payne, and D.M. Jones, "Navigating Large-Scale Virtual Environments: What Differences Occur Between Helmet-Mounted and Desk-Top Display?," *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 2, pp. 157-168, 1999.
[8] M. Slater, M. Usoh, and A. Steed, "Depth of Presence in Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 3, no. 2, pp. 130-144, 1994.
[9] J.H. Brockmyer, C.M. Fox, K.A. Curtiss, E. McBroom, K.M. Burkhart, and J.N. Pidruzny, "The Development of the Game Engagement Questionnaire: A Measure of Engagement in Video Game-Playing," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 624-634, 2009.
[10] C. Ware and P. Mitchell, "Reevaluating Stereo and Motion Cues for Visualizing Graphs in Three Dimensions," *Proceedings of Symposium on Applied Perception in Graphics and Visualization (APGV)*, pp. 51-58, 2005.
[11] Y.-Y. Yeh and L.D. Silverstein, "Spatial Judgements with Monoscopic and Stereoscopic Presentation of Perspective Displays," *Human Factors*, vol. 34, no. 5, pp. 583-600, 1992.
[12] K.W. Arthur, "Effects of Field of View on Performance with Head-Mounted Displays," Doctor of Philosophy, Computer Science, The University of North Carolina at Chapel Hill, 2000.
[13] R. Ball and C. North, "Effects of Tiled High-Resolution Display on Basic Visualization and Navigation Tasks," *Proceedings of Extended Abstracts on Human Factors in Computing Systems (CHI EA)*, pp. 1196-1199, 2005.
[14] D.S. Tan, D. Gergle, P. Scupelli, and R. Pausch, "With Similar Visual Angles, Larger Displays Improve Spatial Performance," *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pp. 217-224, 2003.
[15] C. Ware and R. Balakrishnan, "Reaching for Objects in VR Displays: Lag and Frame Rate," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 1, no. 4, pp. 331-356, 1994.
[16] R. Pausch, D. Proffitt, and G. Williams, "Quantifying Immersion in Virtual Reality," *Proceedings of Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 13-18, 1997.
[17] C. Ware and D.R. Jessome, "Using the Bat: A Six-Dimensional Mouse for Object Placement," *IEEE Computer Graphics and Applications*, vol. 8, no. 6, pp. 65-70, 1988.
[18] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell, "Usability Analysis of 3D Rotation Techniques," *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pp. 1-10, 1997.
[19] R.P. McMahan, A.J.D. Alon, S. Lazem, R.J. Beaton, D. Machaj, M. Schaefer, M.G. Silva, A. Leal, R. Hagan, and D.A. Bowman, "Evaluating Natural Interaction Techniques in Video Games," *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 11-14, 2010.
[20] K.W. Arthur, K.S. Booth, and C. Ware, "Evaluating 3D Task Performance for Fish Tank Virtual Worlds," *ACM Transactions on Information Systems (TOIS)*, vol. 11, no. 3, pp. 239-265, 1993.
[21] C. Ware and G. Franck, "Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions," *ACM Transactions on Graphics (TOG)*, vol. 15, no. 2, pp. 121-140, 1996.
[22] M. Narayan, L. Waugh, X. Zhang, P. Bafna, and D.A. Bowman, "Quantifying the Benefits of Immersion for Collaboration in Virtual Environments," *Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST)*, pp. 78-81, 2005.
[23] W. Barfield, C. Hendrix, and K.-E. Bystrom, "Visualizing the Structure of Virtual Objects Using Head Tracked Stereoscopic Displays," *Proceedings of Virtual Reality Annual International Symposium (VRAIS)*, pp. 114-120, 1997.
[24] D. Raja, D.A. Bowman, J. Lucas, and C. North, "Exploring the Benefits of Immersion in Abstract Information Visualization", in *8th International Immersive Projection Technology Workshop*, 2004.
[25] R.P. McMahan, D. Gorton, J. Gresock, W. McConnell, and D.A. Bowman, "Separating the Effects of Level of Immersion and 3D Interaction Techniques," *Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST)*, pp. 108-111, 2006.
[26] B. Schaeffer and C. Goudeseune, "Syzygy: Native PC Cluster VR," *Proceedings of IEEE Virtual Reality (VR)*, pp. 15-22, 2003.
[27] M. Usoh, K.W. Arthur, M.C. Whitton, R. Bastos, A. Steed, M. Slater, and F.P. Brooks, "Walking > Walking-in-Place > Flying, in Virtual Environments," *Proceedings of Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 359-364, 1999.
[28] S. Razzaque, "Redirected Walking," Doctor of Philosophy, University of North Carolina at Chapel Hill, 2005.
[29] J.N. Templeman, P.S. Denbrook, and L.E. Sibert, "Virtual Locomotion: Walking in Place through Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 6, pp. 598-617, 1999.
[30] D.J. Zielinski, R.P. McMahan, and R.B. Brady, "Shadow Walking: an Unencumbered Locomotion Technique for Systems with Under-floor Projection," *Proceedings of IEEE Virtual Reality (VR)*, pp. 167-170, 2011.
[31] P. Bourdot and D. Touraine, "Polyvalent display framework to control virtual navigations by 6DOF tracking," *Proceedings of IEEE Virtual Reality (VR)*, pp. 277-278, 2002.
[32] R.H.Y. So and W.T. Lo, "Cybersickness: An Experimental Study to Isolate The Effects of Rotational Scene Oscillations," *Proceedings of IEEE Virtual Reality (VR)*, pp. 237-241, 1999.
[33] K.L. Nowak, M. Krcmar, and K.M. Farrar, "The Causes and Consequences of Presence: Considering the Influence of Violent Video Games on Presence and Aggression," *Presence: Teleoperators and Virtual Environments*, vol. 17, no. 3, pp. 256-268, 2008.
[34] R.T. Surdick, E.T. Davis, R.A. King, and L.F. Hodges, "The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 5, pp. 513-531, 1997.
[35] R.B. Ekstrom, J.W. French, H.H. Harman, and D. Dermen, *Manual for Kit of Factor-Referenced Cognitive Tests*, Educational Testing Service, Princeton, NJ, 1976.