The Effect of Avatar Model in Stepping Off a Ledge in an Immersive Virtual Environment

By

Qiang Fu

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Computer Science

December, 2015

Nashville, Tennessee

Approved:

Robert E. Bodenheimer, Ph.D.

Nilanjan Sarkar, Ph.D.

COMPUTER SCIENCE

The Effect of Avatar Model in Stepping Off a Ledge
in an Immersive Virtual Environment

Qiang Fu

Thesis under the direction of Professor Robert E. Bodenheimer

Animated digital self-representations of the user in immersive virtual environments, called self-avatars, have been shown to be an aid in perceptual judgments in virtual environment and provide critical information for people deciding whether an action can be take or not. In this body of work, the size of the self-avatar is carefully calibrated to match the size of user. However, little attention has been paid to the graphical model used to represent the self-avatar. In this thesis, we further investigate the question of whether the form of the model can affect perceptual judgments in an IVE. We study this question in the context of affordance judgments, that is, properties of the virtual environment that represent possibilities for action. Our specific task concerns the judgment of stepping off a virtual ledge, a task we have studied before. In that work, we showed that the presence of a self-avatar provided important information in making the judgment of whether to step off the virtual ledge or not. In this work, we will again employ that task, but vary the underlying representation of the self-avatar across subjects to see if it affects this judgment. The forms of self-avatars vary between no self-avatar, a simple line-based skeleton avatar, or a full-body, rich polygonal, gender-matched self-avatar. Our results replicate our prior work, and show that presenting a self-avatar significantly affects people's perceptual judgment in virtual environments. However, the form of the self-avatar seem to make no difference in such tasks.

Table of Contents

Acknowledgments

I would like to thank Bobby Bodenheimer, who is my mentor and provided full support throughout the study. I also would like to thank John Rieser, who was a source of insight and support throughout the study. Qiufeng Lin provided technical support for this project, and Lian Zhang helped with the pilot study.

List of Tables

List of Figures

I. Introduction

The immersive virtual environment (IVE) is one of the most widely studied topics in computer graphics. The dramatic evolution of computer hardware has greatly improved the way we interact with computers. A fundamental question that arises during the development of virtual environments is how good a virtual environment can represent the real world; in other words, how good the virtual environments are at conveying the situation they are intended to convey [Lin, Rieser, & Bodenheimer, 2015]. To answer this question, we need to know how we can measure the fidelity of virtual environment. For many situations, people tend to behave similarly in virtual environments, compared to the way they behave in the real world [Lin et al., 2015]. However, in some situations, people will behave differently. For example, a large body of literature has demonstrated that people would make different decisions when judging egocentric distances in head-mounted display (HMD) based virtual environments, compared to their decisions in the real world. This difference has been massively studied by various research groups and several studies shows that in head-mounted display-based virtual environments, distances beyond a few meters are compressed [Bodenheimer et al., 2007; Grechkin, Nguyen, Plumert, Cremer, & Kearney, 2010; Jones, J. Edward Swan, Singh, Kolstad, & Ellis, 2008; Loomis & Knapp, 2003; Ries, Interrante, Kaeding, & Phillips, 2009; Thompson et al., 2004]. The exact reason for this compression remains unknown.

In this article, we examined the fidelity of a virtual environment by analyzing people's affordance judgments. The term affordance was proposed by Gibson in 1979 [Gibson, 2014] to represent a person's perceived ability for an action, and the term is now widely used in both perception [Michaels, 2003] and human-computer interaction [Norman, 2013].

There are some interesting issues in the application of affordance to immersive virtual environments. As many IVEs are artificially constructed to mimic or replicate the real world, one issue is the method to measure the fidelity of IVEs. This fidelity could be measured by comparing the possibilities for action in the virtual environment with those possibilities in the real world, with closely matched possibilities representing higher fidelity. A second issue regards the fit between people's physical body in the real world and the virtual body in IVEs. In the real world, the fit between the body and the environment is critical in motor learning and development [Thelen, 1995]. So, it is reasonable to require a fit between people's physical body and the virtual body. However, a digit representation of a person's body (a self-avatar) is absent from most virtual environments. It is then reasonable to ask if learning is necessarily limited if no self-avatar is provided to explore in a virtual environment. On the other hand, the implementation of a size matched self-avatar in a virtual environment is nontrivial. It requires extra hardware and software, as well as the time of the users of the virtual environment to calibrate their self-avatars. Thus, it is worth investigation whether a self-avatar is needed for fidelity purposes in a virtual environment. And, as different forms of self-avatar require different efforts to implement, it is also necessary to investigate whether the form of the self-avatar could affect such fidelity in virtual environment.

The visual cliff is one of the most extensively used virtual environments [Meehan, Insko, Whitton, & Frederick P. Brooks, 2002; Slater, Khanna, Mortensen, & Yu, 2009; Slater, Usoh, & Steed, 1995; Usoh et al., 1999]. It provides an outstanding sense of immersion, which allows subjects to feel that they are in the environment. In our study, we explored the visual cliff as an affordance, by asking the question of whether subjects are willing to step off a ledge with a given height in our head-mounted display based virtual environment. Previous study has showed that if the drop height of a cliff is small, relative to the subject, the subject will choose to step off the ledge. However, when the drop height is large, relative to the subject, the subject will not choose to step off. So, for any subject, there will be a specific threshold of height that the subject will alter the choice (step off or not). And the threshold must be subject specified, because it is relative to one's capability of stepping off ledges. A person's capability of stepping off is influenced by multiple factors. For example, taller people who have better body strength may be capable of stepping off from higher ledges than others.

Recent studies have shown that presenting a self-avatar, which is a virtual body animated by subject, in a virtual environment can improve the accuracy of subject's egocentric distance estimation [Mohler, Creem-Regehr, Thompson, & Bülthoff, 2010; Ries, Interrante, Kaeding, & Anderson, 2008]. Although some work has not found such effects [Geuss, Stefanucci, Creem-

Regehr, & Thompson, 2010; McManus et al., 2011], we can infer that self-avatars can provide a more realistic way for people to interact with virtual environments. In our work, we further explored the factors that affect the utility of self-avatar affordance judgements in virtual environment. Because the exact effects of self-avatar in those tasks are unclear, we made the hypothesis that, through the presence of the self-avatar, human subjects can get some degree of information about their own body dimension, and make their estimation based on their self-avatar in virtual environments. It has already been demonstrated that a size-matched self-avatar can affect the affordance judgment [Lin et al., 2015], and similar results has been concluded by multiple studies [Lin, Rieser, & Bodenheimer, 2012; Loomis & Knapp, 2003; McManus et al., 2011; Mohler et al., 2010; Thompson et al., 2004]. One assumption for this size matched self-avatar effect is that the present of this avatar can implicitly provide the size information of the subject, because the self-avatar allows the subject to observe their feet, legs, etc. in the virtual environment, which can be used as references for distance (vertical distance, in these cases) judgments.

One of the most interesting questions is what factors can affect the influence of self-avatar in affordance judgements in virtual environment. Since we inferred that people can receive physical information of their own body, it is possible that the similarity between self-avatar and the subject can affect subject's distance estimation in virtual environment. To be specific, the presence of a fully rendered human avatar, which is close to the appearance of the subject, will have a deeper influence on the subject, compared with a simply rendered stickman-style avatar (line-avatar), which can only provide the basic dimension information. And both fully rendered and simply rendered avatars should have positive effects. It is well supported in social and behavioral studies that the form of the self-avatar is important [Aymerich-Franch, Kizilcec, & Bailenson, 2014; Fox, Bailenson, & Tricase, 2013; Yee & Bailenson, 2009]. However, little attention has been paid to the form of the avatar in affordance judgments [Bodenheimer & Fu, 2015].

Since the scale of a subject's body parts in virtual environments may have an impact on a subject's actions and perceptions in an IVE, something that has been found by different groups in different studies [Banakou, Groten, & Slater, 2013; Lin, Rieser, & Bodenheimer, 2012;

Linkenauger, Leyrer, Bülthoff, & Mohler, 2013; van der Hoort, Guterstam, & Ehrsson, 2011], all self-avatars in our study are size matched and fully animated by subject's real-time movement. The ledge virtual environment we used in our experiment is inspired by our previous work [Lin, Rieser, & Bodenheimer, 2013], which primarily focused on whether a self-avatar can affect people's affordance in stepping off judgements. In order to help subjects to have a better perception of their self-avatar, we added a mirror in the virtual environment and asked each of subject to look into the mirror and get familiar with their avatars before the experiments. Because the presence of the mirror, subjects are able to see not only their legs and feet, but also their entire torso and head in virtual environment, which provides even more information about their own body dimension.

In our study, we have explored that if altering the form of a self-avatar in an IVE can greatly affect people's affordance judgements. We designed our experiment with three conditions: no-avatar, line-avatar and full-avatar. In order to eliminate the learning effect of subjects, a between subject experiment was designed, which means each subject only experienced one condition during their experiment.

II. Background

Immersive virtual environments (IVEs) can provides realistic and controllable simulated scenarios for people to interact in. They have a great potential in many industries because people can have the opportunities to experience scenarios which are impractical to experience in the real-world. However, there are some major limits of IVEs for their widespread application. One of these factors involves space perception. Several studies have shown that, in head-mounted display-based virtual environments, distances beyond a few meters appear to be compressed. In typical head-mounted display based virtual environments, the environments are presented through the HMD and users need to perceive the environment as a disembodied camera. Multiple studies have reported that when using a self-avatar, subjects perceive the environment differently in typical environments [F. A. Biocca & Rolland, 1998; Mohler, 2010; McManus, 2011]. An avatar is the digital representation of a person in a virtual environment and it is commonly used in a wide range of applications such as video games and IVEs. A major number of avatars are used in the third person perspective, where they are representing either the user at a distance or other characters in the scene. A large body of research has showed that when avatars respond with appropriate behavior, people will interact with avatars in a similar manner to their interactions with real people [Durlach & Slater, 2000; Slater et al., 2006; Zhang, Yu, & Smith, 2006]. Compared with third person perspective, the use of first person perspective, aka, first person avatar (self-avatar) is less common.

Some work has reported that a self-avatar can improve distance estimation in virtual environment[Mohler et al., 2010; Ries et al., 2008], while some other studies reported differently [McManus et al., 2011]. One possible explanation for the effect of self-avatar is that people can see part of their body in the virtual environment, and those body parts can be used as size references. Since the presence of a self-avatar provides a more realistic way for people to interact with the virtual environment, people may use the body parts from their self-avatar as scales, which can be used to measure the absolute distances in virtual environments because people are familiar with the size of their own body. A majority of studies examining distance compression in virtual environments seek the factors that increase or ameliorate this phenomenon. The most commonly used method to study it is blind walking. It has already been shown by multiple groups that people are accurate at distance judgments in the real world when the distance is less than 25 meters [Loomis, Silva, Philbeck, & Fukusima, 1996; Rieser, Ashmead, Talor, & Youngquist, 1990]. On the other hand, in virtual environments, some individual factors have been rejected, e.g., stereopsis [Willemsen, Gooch, Thompson, & Creem-Regehr, 2008], while some other factors are reported differently, e.g. image quality has been rejected by some groups [Messing & Durgin, 2005; Thompson et al., 2004], while other group who used a different method of distance estimation [Kunz, Creem-Regehr, & Thompson, 2009] did find that images rendered in higher quality can help subjects make more accurate judgment.

Gibson [Gibson, 1979] proposed a term “affordance” to describe the property of environment that represent the possibility for an action. An affordance should be subject specified and it is independent of subject’s ability to perceive it. Warren and Whang [Warren Jr & Whang, 1987] reported that one’s affordance is based on the body dimension while studying people’s capability of passing through apertures. And, Mark et al. [Mark, Balliett, Craver, Douglas, & Fox, 1990] showed further evidence when studying people’s ability of setting and stair climbing that the eye height is a preferred body based measure when determining whether an action is a possible or not. Now, the concept of affordance is widely used in perception [Michaels, 2003] and human-computer interaction [Norman, 2013] studies. There are two main issues regarding the affordance that we are particularly interested in. First, since the virtual environments are artificially constructed, we know all the possibilities of action (a.k.a. affordances) are artificially constructed. As one of the primary usages of immersive virtual environments is to replicate the real world, the possibilities of action in these virtual environments should, as close as possible, match the possibilities they replicated from the real world. Thus, the fidelity of an immersive virtual environment should be able to be measured by the similarity between possibilities in the virtual environment and the corresponding possibilities in the real world. Second, it might be reasonable to study the fit between a subject’s own physical body and their self-avatar in virtual environment.

In the real world, it has already been shown that motor learning and development involves exploring the fit between the person's body and the environment [Thelen, 1995]. However, in most virtual environments, such information is not provided due to the lack of a self-avatar, which means the subjects can only explore the environment through a disembodied camera. Thus, if a virtual environment does not provide a "body", then, it is reasonable to question that whether the learning is limited in these virtual environments. Providing a size matched fully real-time animated self-avatar is not a trivial process, it requires a series of hardware and software and needs extra work for every subject to establish and calibrate the whole virtual environment. So, investigating if a self-avatar can really improve the fidelity of a virtual environment is worthwhile. Affordances, which represent the threshold of every individual that divide a possible action from an impossible action [Warren Jr & Whang, 1987; Warren, 1984], are widely used as the basis for comparing virtual environments to the real world in terms of fidelity. A contemporary view of affordances suggests that in respect of individual differences, affordances can be treated as a probabilistic function to represent an individual's likelihood of successful performance of an action [Franchak & Adolph, 2014]. This view allows researchers to apply standard psychophysical procedures and signal detection [Green & Swets, 1966; Nevin, 1969] when estimating affordance functions when necessary.

The term inverse optics problem refers to the fundamentally ambiguous mapping between sources of retinal stimulation and the retinal images that are caused by those sources [Pizlo, 2001]. In the classic view of visual space perception, geometric analysis is needed to infer the structure of the environment (sources) that is likely generated the sensed image. A new approach has been proposed during the convergence of research in psychology and neuroscience. In this new approach, sometimes called embodied perception, the subject's body is considered central to the act of perceiving [Barsalou, 2008; Proffitt, 2006; Wilson, 2002]. It has been recognized for a long time the body-based perception is important in immersive virtual environments [F. Biocca, 1997; Hillis, 1999; Slater & Usoh, 1994], and the way body is represented in an IVE may have a significant impact on how the subject perceive and act in an IVE. However, only recently has IVE technology evolved sufficiently to allow high fidelity replication of subject's body in virtual environments [Mohler et al., 2010].

Multiple research groups have shown that the visual representation of a user's body parts has an important influence on the viewer's spatial position judgment. It was found five decades ago that vision has a dominating role in body representation [Hay, Pick Jr, & Ikeda, 1965]. They reported that when a viewer face a conflict between visual represented and proprioceptive position of their own arm, they tend to resolve the conflict by being biased in favor of the visual represented arm's position, and feel the arm as where it is seen. In another study [Botvinick & Cohen, 1998], similar visual capture effects have been induced in the rubber hand illusion by using an artificial limb, which is a fake hand visible to the viewer. In this experiment, the viewer is asked to look at the fake hand in front of them while their own hand is unseen, and both fake hand and real hand are stroked simultaneously. The researchers reported that subjects will feel and act as if the stroking is really happening at the location of the fake hand. This illusion has also been demonstrated in virtual environment recently [Slater et al., 2007; Slater, Perez-Marcos, Ehrsson, & Sanchez-Vives, 2008].

Body-based information in IVEs may provide extra useful information in spatial perception tasks. The presence of a body may serve to anchor the body's position in space, while other body based visual information may help subjects to establish references, especially when the cues of location conflict and / or the body's position is ambiguous. Also, the awareness may serve as a scale when measuring the dimensions of space due to the familiarity of one's own body parts and visual feedback of their own motion. It is known that the size of a self-avatar, or components of the self-avatar can have an effect on perceptual judgment tasks in virtual environments. In mixed reality, people's distance judgments changed when they are induced to believe that their body size are changed [van der Hoort et al., 2011]. In virtual environments, people's affordance judgment changed when the leg length of their self-avatar are modified [Lin et al., 2012]. Linkenauger and colleagues found that subject's graspability judgments can be affected when the visual hand of their self-avatar is scaled [Linkenauger et al., 2013]. Banakou et al. [Banakou et al., 2013] reported that subjects make different size judgments between a proportionally scaled self-avatar and a self-avatar that is scaled to a child's proportions, with subjects significantly overestimating sizes when using the child's avatar. However, all these studies did not manopulate the form of the self-avatars.

It has already been shown that viewing a subject's own feet and the immediate surroundings on the ground will not improve the accuracy of the subject's distance estimation [Creem-Regehr, Willemsen, Gooch, & Thompson, 2005]. This experiment was performed in a real world environment which provides a number of depth cues. The difference in perceptual uncertainty between the real world environment and the virtual environment may result in making the subject rely on different body-based or environment-based information in these two environments. Other factors which may result in the variation of distance judgment in immersive virtual environment include body movement tracking, the rendering of different body parts and the realism of rendering. In the real world, the accuracy of space perception is usually measured using visually directed actions [Loomis, Da Silva, Fujita, & Fukusima, 1992; Rieser et al., 1990]. In those studies, subjects are provided with controlled visual stimulus and asked to perform a series of actions based on the visual information, and no visual feedback is provided. Because of the absence of visual feedback, it is arguable that the accuracy of resulted action can represent the accuracy of the corresponding space perception. These visually directed action tasks that are used to exam distance perception include blind walking [Fukusima, Loomis, & Da Silva, 1997; Loomis et al., 1992; Rieser et al., 1990], pointing to a previously seen target [Loomis et al., 1992] and throwing at a previously seen target [Eby & Loomis, 1987; Sahm, Creem-Regehr, Thompson, & Willemsen, 2005]. In the real world, people are accurate at absolute distance perception, which is indicated by their performance of visually directed action [Loomis et al., 1992; Loomis et al., 1996; Rieser et al., 1990]. However, as reported by multiple groups, the accuracy of absolute distance perception in an HMD-based virtual environment is not as good as the accuracy in the real world [Henry & Furness, 1993; Loomis & Knapp, 2003; Richardson & Waller, 2007; Sahm et al., 2005; Thompson et al., 2004; Waller & Richardson, 2008]. The results showed that, the actions in virtual environment are performed as if the distances were perceived 20% - 50% smaller than it should have been. Many studies have been done to find the factors that may cause the distance compression in IVEs and many isolated factors have been rejected, such as stereopsis, field of view [Creem-Regehr et al., 2005; Knapp & Loomis, 2004] and motion parallax [Beall, Loomis, Philbeck, & Fikes, 1995]. Other factors are found may have impact on the compression, such as physical properties of HMD [Willemsen, Colton, Creem-Regehr, & Thompson, 2004] and cognitive effects [Foley, 2007; Interrante, Ries, & Anderson, 2006; Interrante, Ries, Lindquist, Kaeding, & Anderson, 2008; Richardson & Waller, 2005].

III. Experiment & Results

In this thesis, our goal was to exam whether the form of a self-avatar could affect people's interaction with immersive virtual environment. In our approach, we chose the visual cliff with variable heights as the platform for examining people's affordance judgments in this tasks. Thus, people's judgment in these tasks can be represented by their threshold value at which they report changing their decision between willing to step off the cliff and not willing to do so. We grouped our subjects into three different conditions: no-avatar, line-avatar and full-avatar. While the subjects in the no-avatar condition cannot see their self-avatar in virtual environment, the subjects in line-avatar condition and full-avatar condition were able to see a line-based skeleton self-avatar and a full-body self-avatar, respectively.

We also implemented a virtual mirror, which was not included in a previous study [Lin et al., 2013], in order to give subjects awareness of their own avatar and help them differentiate between the models that we provided. Due to the fact that this study was also focused on the effect of different forms of avatars, we felt it was necessary for subjects to receive more information about their self-avatar.

3.1 System setup

To implement our planned functionality, we used a system as described in Figure 1. This system, as a whole, provided subjects with the ability to experience immersive virtual environment with a real-time self-avatar.

3.1.1 Subject

Subjects experienced the virtual environment through a head-mounted display (HMD) and are allowed to physically move around with a designated area (About 4m × 4m). They wore a six-component tracking device, which was tracked by a Vicon tracking system. This method of tracking greatly reduced the preparation cost for tracking since a traditional tracking suit might take hours to calibrate while our method usually takes 10 - 15 minutes.

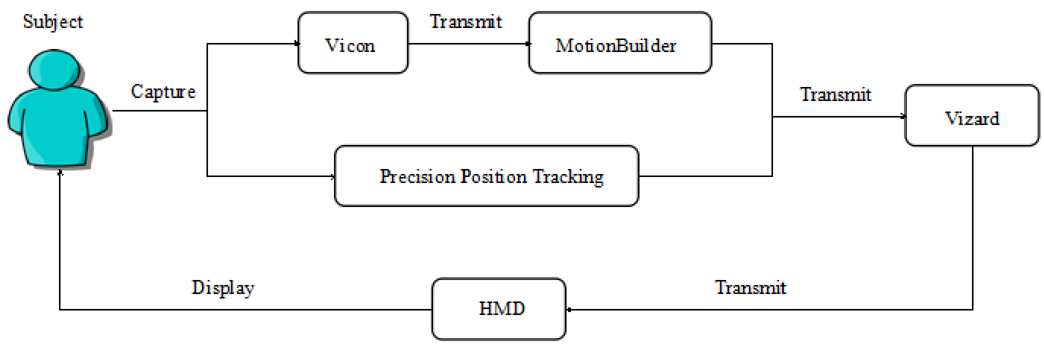

Figure 1. System framework

3.1.2 Vicon

This subsystem consisted of eight cameras and software used for the tracking calculation. This subsystem tracked all tracking components worn by the subject in real-time, while each component worn represented one body part of the subject (Head, waist, right hand, left hand, right foot and left foot). The Vicon tracked both positions and orientations of all six body parts of the subject and sent these data to the MotionBuilder software.

3.1.3 MotionBuilder

This subsystem mapped raw data received from Vicon to a virtual character and used inverse kinematics to generate the animation of the entire character in real-time. The resulting character was then sent to the Vizard rendering system.

Inverse kinematics refers to the process of determining joint parameters that provide a desired position of the parts of the resulting character using a kinematics equation. This process allowed us to generate the motion of self-avatar's entire body using only six tracked objects.

3.1.4 Precision Position Tracking

This subsystem was used to track the head positions and orientations of the subjects, which were used for viewpoint calculation. It provided more stable and accurate results in a greater range than Vicon. Since the Vicon could lose tracking of trackers sometimes, which could result in a very unappealing experience to the subjects if the viewpoint is tracked by Vicon, we felt it was necessary to use a separate tracking system exclusively for viewpoint.

4 cameras were used to track the position of active LED markers, while the orientation was tracked by a built-in gyroscope.

3.1.5 Vizard

This subsystem took avatar character data from MotionBuilder, as well as head positioning data from Precision Position Tracking to render the final virtual environment, which was displayed in the HMD. All the items such as the final visible self-avatar, the room, the mirror, the ledge, and all the logic controls of the experiment were calculated in this subsystem.

3.1.6 HMD

The rendered stereoscopic final images were received by subjects through the HMD. This is the interface for subjects to experience the virtual environment.

3.2 Motion capture pipeline

For each subject in all three conditions, a self-avatar was specifically calibrated to fit the subject, even if the subject cannot see the self-avatar, as in the case of the no-avatar condition. Every self-avatar had the same leg length, arm length and eye height as the subject it represented. Also, for the full-avatar condition, a gender-matched avatar model whose skin

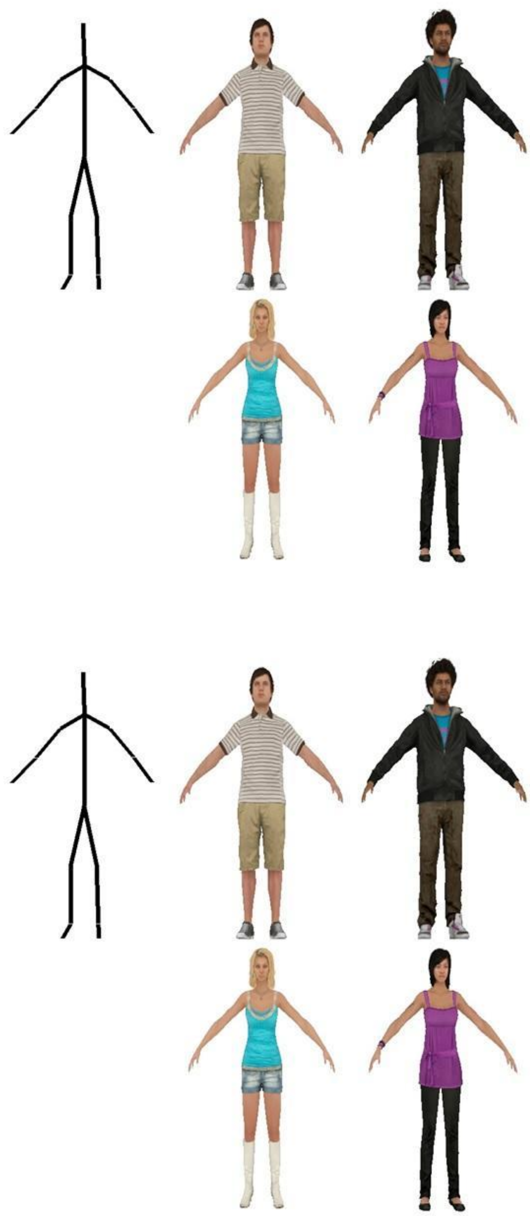

Figure 2. Self-avatars for different conditions: Line-avatar (upper left), No-avatar (lower left), Full-avatar (center and right)

tone most closely matched the subject was chosen from our model library. Figure 2 shows the avatar model possibilities.

3.2.1 Motion tracking mechanism

Our motion capture tracking device consisted of six individual components: one component for each hand, one for each foot, one for the waist and one for the head. Each component contained at least five tracking balls, which were small balls covered by material highly reflective to near infrared light and can be tracked by the eight cameras mounted around the zone. In each component, the relatively position is fixed. As we know, in three-dimensional space, every object's position and orientation can be determined by the position of at least three points, while the position of one point can be determined by at least two cameras. Thus, the positions and orientation of each component can be calculated by the information of its tracking balls captured by the eight Vicon cameras. See Figure 4

3.2.2 Inverse kinematics mapping

The motion of entire avatar in our system was calculated using inverse kinematics, which used the information of two or more joints to calculate and simulate the position and orientation of other parts. For example, if we know the positions and orientations of both the left wrist and left shoulder, and other information such as the lengths of the upper and lower arm, and the restriction of elbow, we could calculate the information of the entire left arm. In our system, MotionBuilder received data from 6 critical joints from the Vicon and these data are mapped to the corresponding joints of a virtual character. The body dimensional information of the virtual character was set manually based on each subject's own body dimension. Thus, the poses and positions of the subjects were calculated, and the resulting data was packaged and streamed to the Vizard machine. See Figure 3.

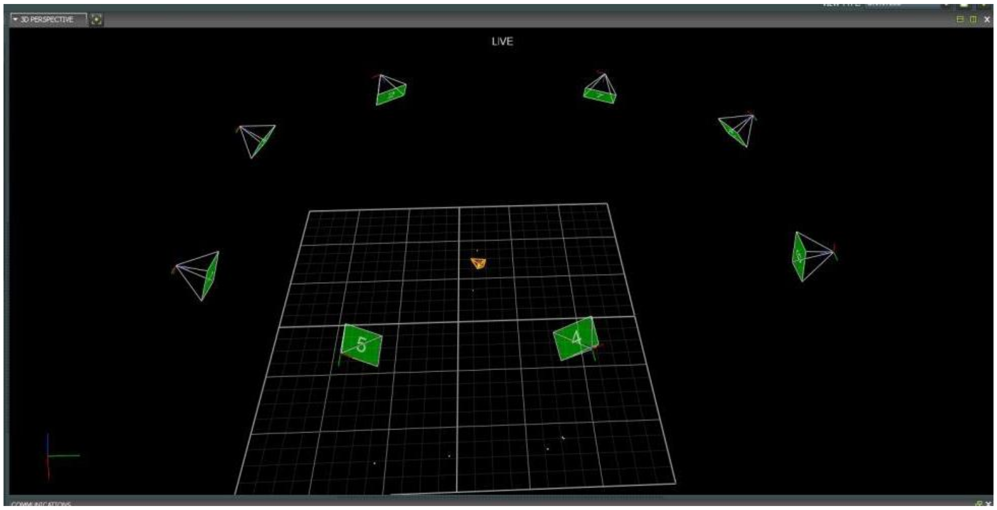

Figure 4.The head mounted display tracked by Vicon

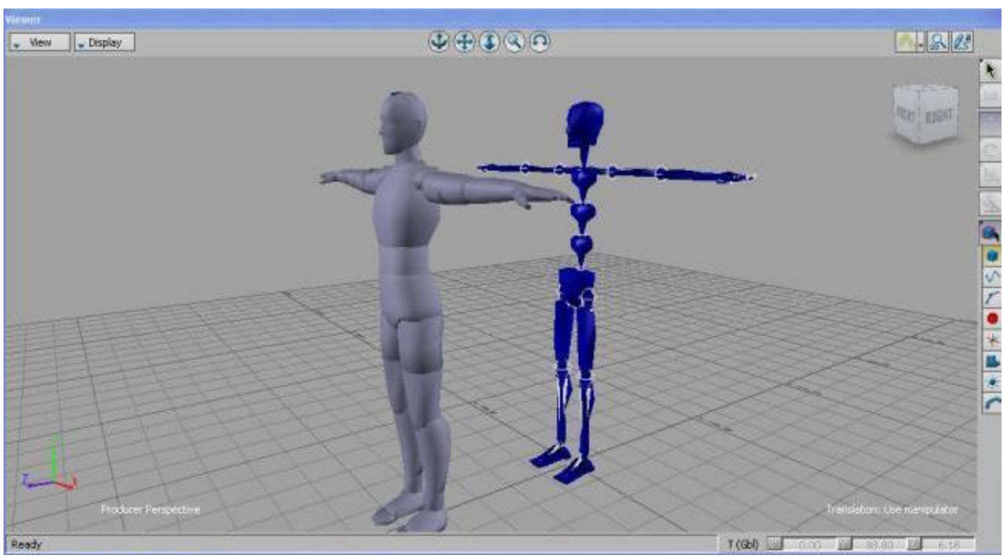

Figure 3. Data received from Vicon (represented by the left avatar) and the avatar generated by the inverse kinematics mapping (the right skeleton). Note: The two character are usually coincident in the real experimental procedure.

3.2.3 Render

When the poses and positions are calculated, the character is ready to be rendered. We used Vizard toolkit to render the entire scene of our virtual environment. Different models were used for different subjects, no matter whether the model was visible to the subject or not. While all models were attached to their character by mapping the joints from the character to the corresponding joints in the model, these models move in the exact way as their corresponding character. Thus, the motion of subjects now could be seen as the motion of avatars in our virtual environment.

3.3 Software design

In our study, we have implemented an immersive virtual environment containing a room, a mirror, a stage and a self-avatar. In the experiment, subjects found themselves standing in a brick-textured square room, with a wooden-textured stage under their feet, and a full length mirror in front of the subjects for them to observe the environment. The avatars were presented to the subjects through a first-person view, which can be observed in the same manner as subjects observe themselves in the real world, with an exception for the no-avatar condition in which no avatar can be seen in the virtual environment. See Figure 5.

The virtual environment was implemented mainly in Python and run in Vizard. All elements in the virtual environment were updated in real-time, while the HMD had a frame rate of 60Hz.

3.3.1 System workflow

The workflow of the entire system is shown in Figure 6. The system contained 3 parallel autonomous subsystems as well as a parallel logic control thread that used to get subjects' feedback and control the experiment progression. Each experiment contained 25 trials, while each trial was consisted of three processes: a maximum-likelihood procedure, scene setup and feedback.

Initialization:

The initialization module was used to start the system. All initial states of every element in the program were set in this stage. As well, some initial user inputs were received and processed here. Some major initialization included: setting the scene, configuring the view point, establishing connections with both MotionBuilder and Precision Position Tracking, gathering initial information of self-avatar and filling initial parameters for the maximum-likelihood procedure.

Maximum-likelihood:

An adaptive maximum-likelihood stimulus procedure [Grassi & Soranzo, 2009] was used in our system to determine the ledge height in each trial. The values of the midpoint, slope and false alarm rate used in this procedure were all chosen from our prior work [Lin, Rieser, & Bodenheimer, 2015]. This algorithm took subjects decisions of stepping off or not as inputs, and converged to the threshold values of the subjects. These values were used to estimate the affordance threshold of the subjects. The stimulus (ledge height) for the next trial was calculated in this process.

Please see section 3.3.2 for detailed description.

Scene setup:

This process was used to modify the scenes based on the desired ledge height. We noticed that when the height is close to people's threshold, if they have a clear visual on the translation of ledge height, they may feel compelled to make a decision based on their experience (the decision of last trial) rather than observation of the environment. To reduce this effect, the images on HMD turned black for two seconds between each two trials, during which the height of ledge was changed. We felt this would help the subjects to be more focus on their observation.

Get feedback:

The decisions of whether stepping off the ledge or not were given by subjects during trials. These decisions were recorded and used in the next iteration of the maximum-likelihood procedure. The whole thread ended and three subsystems terminated when all 25 trials were done.

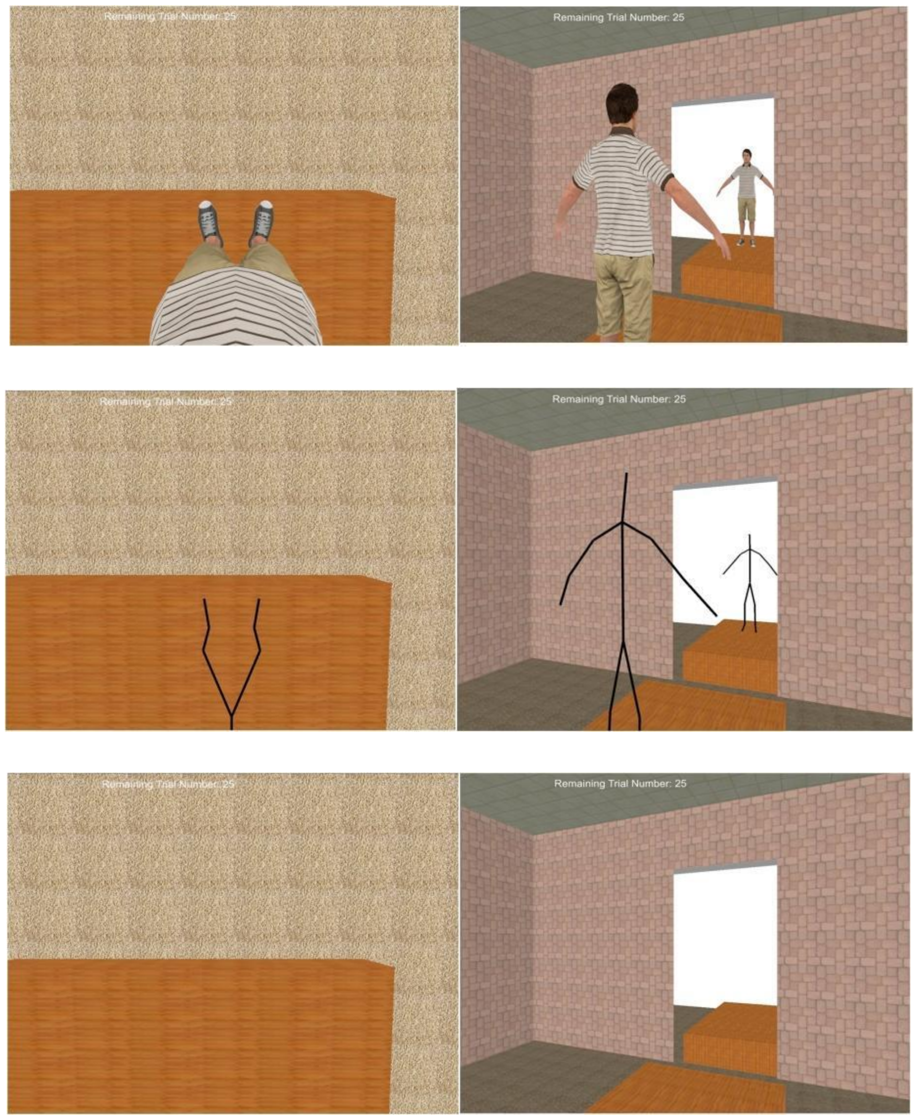

Figure 5. The virtual environment used in our experiment. Left: First-person view, Right: Third-person view; Up: Full-avatar condition, Middle: Line-avatar condition, Bottom: No-avatar conditio

Mirror control:

The mirror in our virtual environment was implemented as a reflection texture. The basic idea was to treat the mirror as a screen while a virtual camera is set on the other side of the mirror. So, as the avatar moved with the subjects, the mirror updated itself continuously.

Please see section 3.3.2 for detailed description.

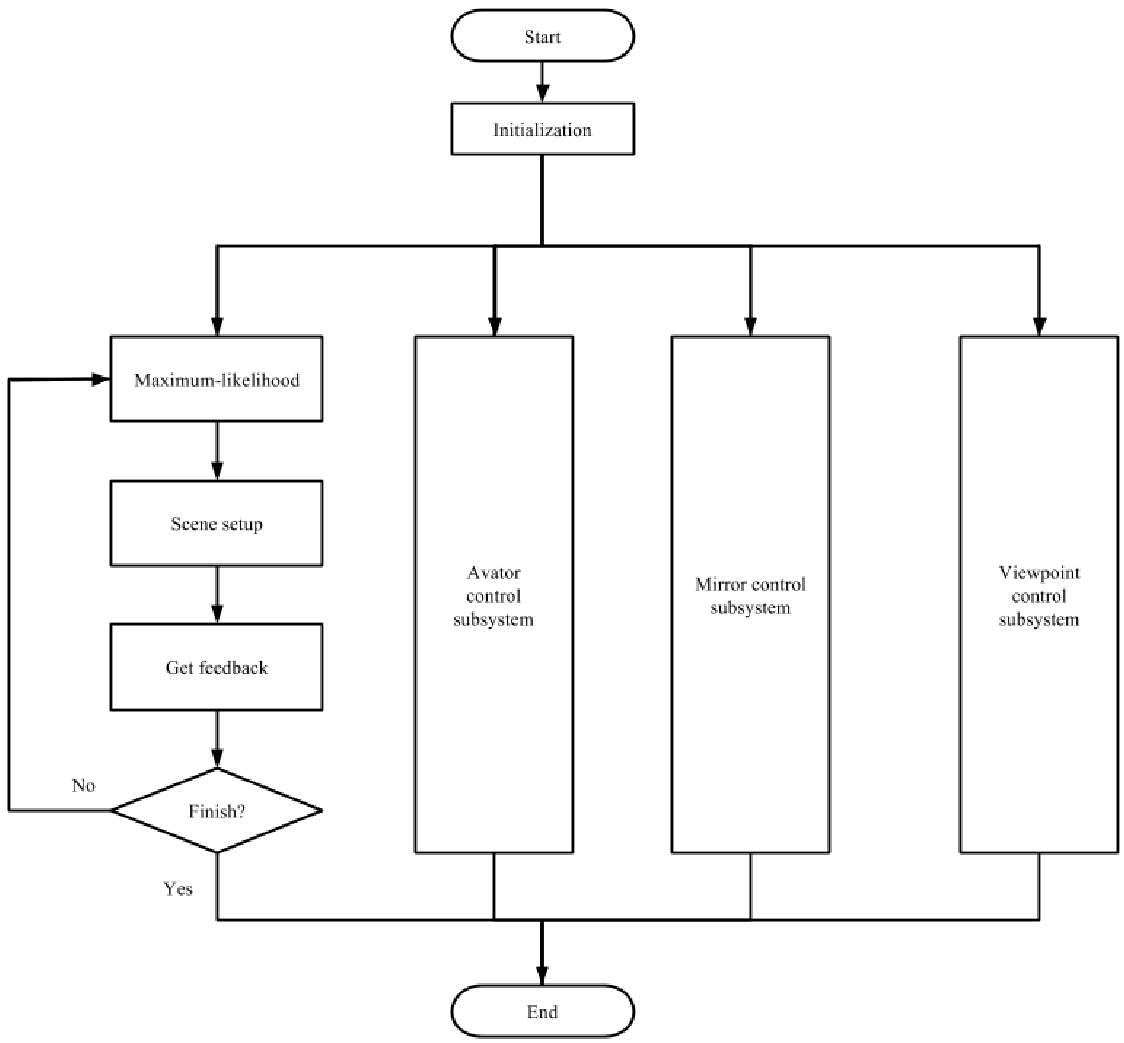

Figure 6. The overall system logic of the Viard. These four subsystems ran simultaneously.

Viewpoint control:

The viewpoint was rendered stereoscopically. Two viewpoints were the virtual eyes of a subject and were the points where subjects perceived the virtual environment. These viewpoints were essentially a pair of viewpoints that were rendered separately so that the subjects experienced stereoscopic images, which, like the real world, provided more distance information than traditional redering. The position and orientation of this viewpoint is tracked by the Precision Position Tracking system, which streamed the tracked data to Vizard.

Avatar control:

This module was used to control the self-avatar in our virtual environment. While the detailed process varied between different conditions, in all condition, this subsystem received streamed avatar data from MotionBuilder and updated its designated avatar in an autonomous manner.

Please see section 3.3.2 for detailed description.

3.3.2 Subsystem design and detailed implementation

Mirror control:

This subsystem continuously created render nodes to render the mirror textures for both eyes. These textures were attached to the mirror object in the scene of our virtual environment, and were rendered with other objects in Vizard to generate the final images for each eye.

According to the law of reflection, the image can be observed from the mirror is the same as be observed from the viewpoint's mirrored position with mirrored gaze direction, or, the same as be observed in the original position with mirrored objects, as shown in Figure 7. Common ways of implementing mirror include either rotating camera or rotating objects to generate the scene people should see in the mirror. In our study, the implementation of the mirror was treated as the subjects observed the entire mirrored environment from their origin position. The workflow of the mirror implementation is shown in Figure 8. We chose this method to implement the mirror due to the traits of Vizard and the fact that the number of objects in our virtual environment is relatively small. We believed that the performance of the program can be optimized by mirror objects with matrix operations.

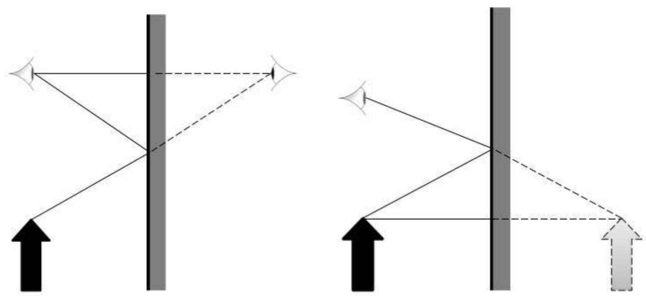

Figure 7. Two common methods of implementing mirror in virtual environment.

Create render texture: Created a new texture for every render process. Every texture in each frame represented one static image that can be seen by subjects. These textures were rendered by Vizard along with other objects in the scene.

Create render node: In every frame, a new render node was registered to the shader and the texture was attached so the texture can be rendered.

Setup reflection matrix: This process calculated the mirrored objects using matrix multiplication. The operation we used is as following:

$$M_{-pos} \times M_{invRot} \times M_{scale} \times M_{rot} \times M_{pos}$$

The first term translated the mirror to the origin of the world coordinate while every other object had added the same offset. The second element rotated the whole scene for X degree to allow the mirror to coincide with the x-y plane in the coordinate system. The third element inverted the z value of all objects, which mirrored the objects by x-y plane. The forth element rotated the whole scene back to the original orientation. Finally, the fifth element translated the whole scene back to the position where the mirror is now at the original position with same original in world coordinate.

Setup reflection clip plane: This process clipped the mirrored scene. Since all objects behind the mirror should not be seen, this process also reduced the cost of mirroring objects.

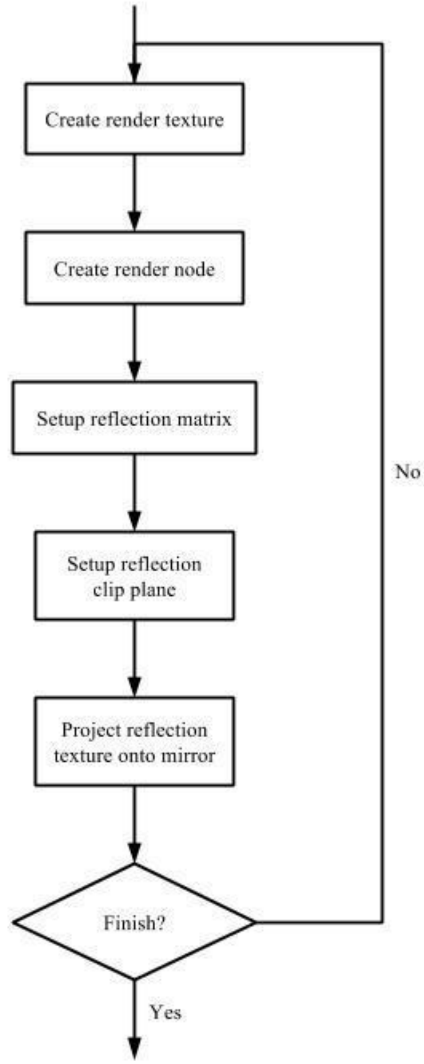

Figure 8. Mirror subsystem workflow

Project reflection texture onto the mirror: This process was used to attach the resulting texture onto the mirror object, which was the visible object in our virtual environment.

Avatar control:

This subsystem took autonomous control of the avatars, and was terminated with the main program. The module was divided into 3 conditions, as show in Figure 9.

In the no-avatar condition, the subsystem remained idle and discarded any data received from MotionBuilder. Since the viewpoint was tracked and controlled separately, the subjects experienced and perceived the environment as a disembodied camera.

In the full-avatar condition, a full-body avatar was loaded in each experiment. The selection of full-body avatar was based on subject's gender and skin tone. The subsystem received avatar data from MotionBuilder and immediately updated the avatar in the scene. The data received from MotionBuilder were packed, each pack contained position and orientation information of 15 joints. These data were then mapped to the 15 respective joints on the avatar in the virtual environment. These 15 joints are: head, left and right calves, left and right upper arms, left and right feet, left and right forearms, neck, pelvis, left and right hands and left and right toes.

One problem we found during the development of this module was the Vicon may lose tracking of a tracking component, which usually last less than one second. This issue was not especially serious when it happened on most joints, but for the head joint, this stability issue could cause position by comparing both position and orientation data received from MotionBuilder and Precision Position Tracking. There are two situations that could help us to detect the issue: The difference between two system's data became too great, or the data from MotionBuilder suddenly changed by a significant amount which could not be the result of a subject's movement. When a glitch was found, the position and orientation of the head joint was bonded to the view point until the difference of two data from two machines became close again.

In the line-avatar condition, a default invisible full-body avatar was used for every subject. The control of this default avatar was similar to the avatars in the full-avatar condition. This default avatar took data from MotionBuilder and was used as the template of the line-avatar. In each iteration, a line was drawn between each pair of connected joints while the old line was removed, so the line-avatar acted exactly the same as the unseen default avatar. As the way of how joints were connected, 14 lines were drawn to represent the skeleton of the avatar, they were: Head (with neck), both shoulders, both big arms, both forearms, spine, both thighs, both calves and both feet.

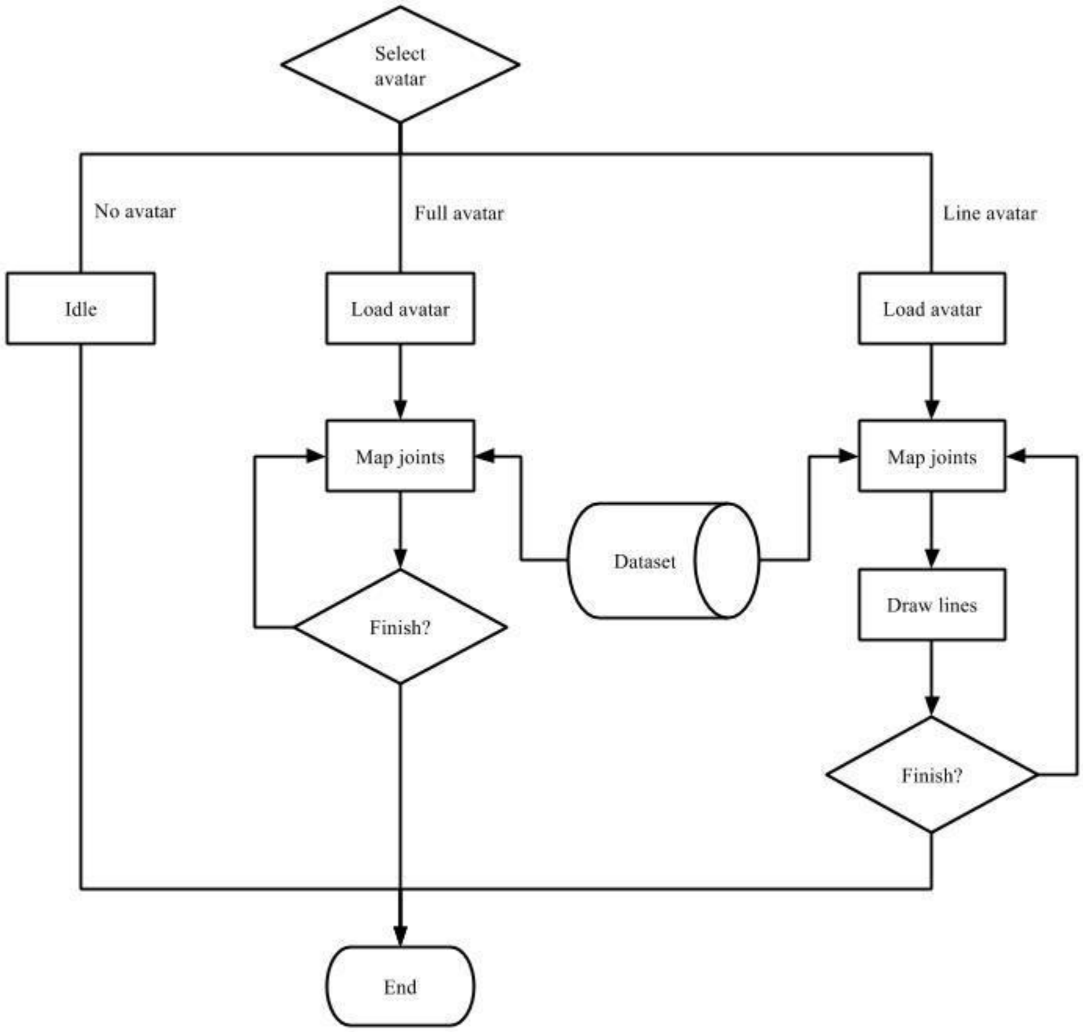

Figure 9. Avatar subsystem workflow

Maximum-likelihood procedure:

The adaptive maximum-likelihood stimulus procedure was used in our system to determine the ledge in each trial. As described in [Grassi & Soranzo, 2009], this procedure iterated using prior responses and prior stimulus. The values of the midpoint, slope and false alarm we used in this study were chosen from our prior work that with similar experiment design [Lin et al., 2015].

According to the Sensory Threshold Estimation Theory, people's performance in yes or no tasks (such as in our experiment) can be represented by a psychometric function, which is a function of the stimulus level. One of the most widely used psychometric functions is logistic psychometric function, which is:

$$\Psi = \gamma + (1 - \gamma - \lambda)\left[\frac{1}{1 + e^{\beta(\alpha - x)}}\right]$$

In this function, α (also referred as the midpoint) was used to enable the function displacement along the stimulus-level axis and it corresponded to the average value of γ and λ. β is the rate of change of subject's performance, which represented by the slope of the function. Also, γ corresponded to the subject's false alarm rate, which was the rate of biased responses and affected the lower bound. These three values were used to calibrate the function to the specific need in each experiment.

The maximum-likelihood procedure consisted of two independent processes, which are the maximum-likelihood estimation and the stimulus selection. Our experiments started by providing the maximum stimuli to the subjects, which stimuli is the height of their own eyes. Then the subjects' responses were recorded to calculate the likelihood of threshold estimation. The likelihood value was calculated by the following function:

$$L(H_j) = \sum_{i=1}^{n} H(x_i)^C [1 - H(x_i)]^W$$

In this function, $L(H_j)$was the likelihood the the j-th hypothesized function while i was the trial number. The hypothesized function $H(\cdot)$, which described in the first formula, was inspired by our prior study [Lin et al., 2015]. The maximum-likelihood procedure selected the highest likelihood hypothesis once all likelihoods were calculated. The hypotheses with the highest likelihood resembled the subjects' actual psychometric functions and were identified by midpoint α. This procedure returned an estimated threshold after every trial, and the accuracy of this estimation increased with the number of trials.

The stimulus selection was used to choose the stimulus for the next trial. One common method to choose the stimulus was to set the stimulus level to the estimated threshold (the p-target). Hence, even at the beginning of the experiment, the stimuli were generated with enough information for a successive trial. As the threshold was the inverse function of the likelihood hypothesis, the next stimulus was calculated by the following Equation:

$$\Psi^{-1}(p_t) = \alpha_j - \frac{1}{\beta}\ln\left(\frac{1-\gamma-\lambda}{p_t-\gamma} - 1\right)$$

The p_t in this equation was the p-target. Green [Green, 1990, 1993] proposed an optimal p-target that could be used to optimize the estimate of the subject's threshold, and this optimal p-target was often referred as the sweet point. This sweet point p-target was calculated by the following equation:

$$p_{SW} = \frac{2\gamma + 1 + \sqrt{1+8\gamma}}{3+8\gamma}$$

With all four equation presented above, our maximum likelihood procedure used in our experiments converged to the threshold of the subjects quickly. A typical result of estimated threshold for a subject is shown in Figure 10. We used the mean of the last four estimated thresholds of each subject to calculate the final result.

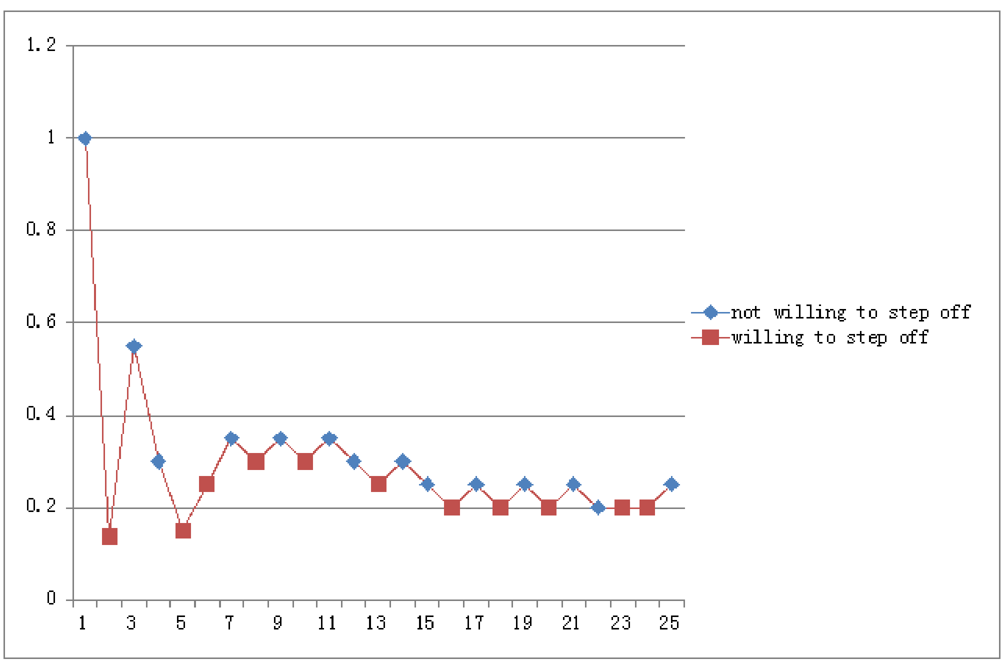

Figure 10. A typical result of estimated threshold

The pseudo-code of maximum-likelihood algorithm used in our system is shown below:

```
Input: Subject's decision

Output: The threshold value and the next stimulus

float[] stimulusArr,

int[] responseArr;

float[] alpha = [midpoint values]

float[] falseAlarm = [false alarm values]

float[] slope = [slope values]


Function maxlike(responseArr)

        Float maxLocalLikelihood

        Float thresholdEst

        Float gamma

        Float beta

        For each f in falseAlarm:

                For each s in slope:

                        For each a in alpha:

                                Float localLikelihood = 0

                                For each r in responseArr:

                                        P = falseAlarm[f] + (1 - falseAlarm[f]) * (1 / (1 + math.exp(slope[s] *
                                                                                        (alpha[a]        -
                                stimulus[r]))))

                                        If p < p_min

                                                p = p_min

                                        If p > p_max

                                                p = p_max

                                        If responses[r] == 1

                                                localLikelihood = localLikelihood + math.log(p)

                                        Else

                                                localLikelihood = localLikelihood + math.log(1 - p)

                                If localLikelihood  > maxLocalLikelihood

                                        maxLocalLikelihood  = localLikelihood
```

```
                    thresholdEst = alpha[a]
                    gamma = falseAlarm[f]
                    beta = slope[s]
float sweetPoint = (2 * gamma + 1 + math.sqrt(1 + 8 * gamma)) / (3 + math.sqrt(1 + 8 * gamma))
float nextStimulus = thresholdEst - (1 / beta) * math.log(((1 - gamma) /  (sweetPoint - gamma)) - 1)
stimulusArr.append(nextStimulus)

Return thresholdEst and nextStimulus
```

3.4 Experiment

The affordance task we used in our IVE is to examine people's perception of visual cliff. In our experiment, subjects are asked to stand beside the edge of a ledge (visual cliff) and report if they are willing to step off from the height of the ledge. Thus, the threshold value is the value of height at which people change their decision from willing to step off the ledge to not willing to step off the ledge.

There are three conditions in our experiment. One is the no-avatar condition, which the subjects have no self-avatar. Another is the line-avatar condition, in which subjects are provided with a line-based skeleton avatar. The last one is the full-avatar condition, in which subjects are provided with a full body, gender-matched avatar. We performed our experiment in a between subject manner, in which one-third of the subjects participated in the no-avatar condition, one-third of them participated in the line-avatar condition, and one-third of them participated in the full-avatar condition. Subjects in this experiment are asked only to report their decision (step off or not); no physical action is required as we deem stepping from ledge while wearing a HMD to be unsafe.

3.4.1 Participant

18 subjects with ages range from 18 to 24 participated in our experiment. All subjects are recruited from Vanderbilt University and all of them had normal or corrected-to-normal vision. Four males and two females experienced the no-avatar condition, four males and two females experienced the line-avatar condition, and three males and three females experienced the full-avatar condition.

3.4.2 Materials and Apparatus

In our experiment, subjects wore six components for tracking, and their motion is captured by an eight-camera Vicon (Los Angeles, CA) MX-F40 optical tracking system and Tracker (v. 1.0) in real-time. The raw motion data were transmitted to a separate machine that runs Motionbuilder (Autodesk, San Rafael, CA), where the motion data were mapped to a calibrated character using inverse kinematics. The data of the resulting character was then transmitted to a third machine which ran Vizard (Worldviz, Santa Barbara, CA) to render the immersive virtual environment and output the rendered image to a full color stereo NVIS (Reston, VA) nVisor SX head mounted display (HMD). The field of view (FOV) of the HMD was 60°, the resolution was 1280 × 1024 pixels per eye, and the frame rate was 60Hz. The subjects' head position and orientation were also been tracked by Precision Position Tracking (Worldviz, Santa Barbara, CA) on another machine for more accurate results.

The experiment for each subject started with the subject reading the experimental instructions, after which, we explained the task procedures again and answered questions orally to insure that the subject understood them. It took 10 - 15 minutes to suit a subject with the six components of the tracking device and fully calibrate the corresponding self-avatar. Each self-avatar was calibrated to the same size as the subject: The leg length and arm length of the avatar were calibrated to the actual leg and arm lengths of the subject, and the eye height of the avatar was calibrated to the eye height of the person. For subjects in the full-avatar condition, a gender-matched avatar model with skin tone that can best represent the subject was chosen from our avatar library. For all subjects in the line-avatar condition, a black line-based model that represented subject's skeleton was used. And, for factor elimination reason, even the subjects in the no-avatar condition were not able to see any avatar, a default invisible avatar was used and subjects were still asked to wear the same tracking device as other conditions. In all three conditions, subjects experienced the IVE in first-person perspective and avatar was collocated with the participant.

3.4.3 Method

After the preparation and calibration were finished, subjects were asked to stand on the floor, in the center of our physical lab while wearing the HMD. When the virtual environment first appeared, subjects found themselves in a virtual room, standing on a wooden stage in the virtual room and facing a virtual mirror. Then, subjects in both line-avatar condition and full-avatar condition were given one minute to move around while observe their self-avatar both in first person and from the mirror, to get familiar with the avatar. Subjects in the no-avatar condition were also asked to move around and observe the mirror even they had no visual feedback about their body motion. After they had done this, subjects were asked to turn around and walk until they stood beside an edge of the wooden stage (the ledge), and the main experiment procedure began.

There are in total 25 trials for each subject. In each trial, subjects experienced different ledge heights, which can be observed by looking downward, and answered the same question: "Are you able to step off the ledge gracefully and comfortably without losing your balance". Subjects were able to move their body to observe and analyze the height, but no real stepping off action was performed. The height of ledge for next trial is selected automatically based on subjects' response, while the height for the first trial for each subject is selected based on subject's eye height. The screens of HMD turned black for two seconds between every two trials to reduce the effect of previous experience on subject's decision.

IV. Results

The maximum likelihood procedure used in our experiment converged quickly. The estimated threshold converged by about 15 trials for most subjects. The final threshold value of each subject was calculated by averaging the estimated threshold of last four trials. The mean threshold values and standard errors of the mean were expressed as a proportion of ledge height to subject's eye height. These threshold values and standard errors for all subjects in each condition are shown in Table 1 and Figure 11.

Table 1. Mean threshold values for all subjects for each avatar condition expressed as a proportion of ledge height to eight. Standard errors of the mean are shown in parentheses.

	No-avatar	Line-avatar	Full-avatar
Mean threshold	0.47 (0.06)	0.25 (0.04)	0.20 (0.03)

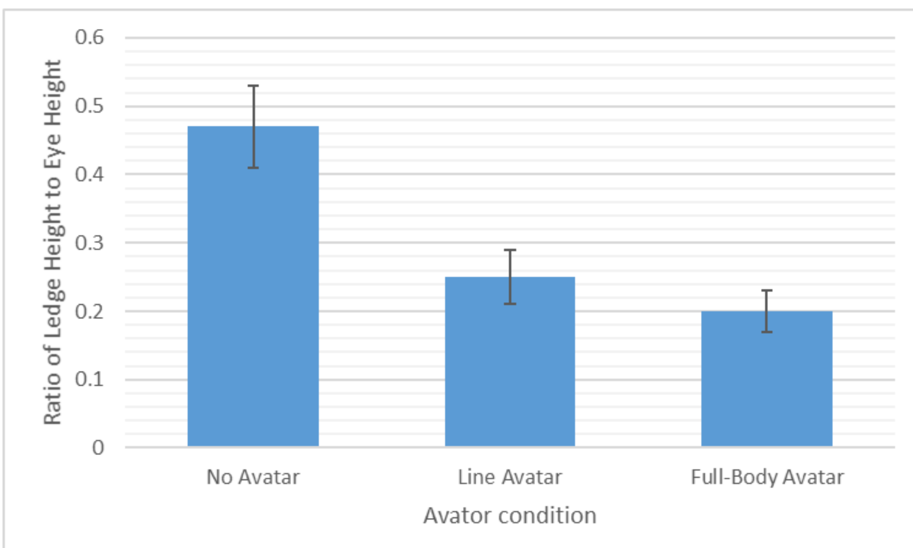

Figure 11. The threshold values involved in stepping off the ledge by avatar condition. The Y-axis is the proportion of the ledge height to the subject's eye height. Error bars indicate standard errors of the mean.

The results on the mean thresholds using analysis of variance (ANOVA) shows that avatar has a significant effect in distance estimation tasks, with $F(2,15) = 10.1$, $p < 0.01$. Post hoc analysis using Tukey's Honestly Significant Difference (HSD) indicates that subjects in the no-avatar condition group had a significantly higher threshold than the other two conditions ($p < 0.01$). However, the results of line-avatar and full-avatar did not differ significantly. Thus, the presence of the self-avatar significantly reduced the magnitude of the threshold, which is consistent with our prior work [Lin et al., 2015]. However, the form of the avatar did not significantly affect these thresholds.

It is useful to compare the mean thresholds in our result with our prior work qualitatively. In our prior work, we used the same method but lacked a virtual mirror in the environment, and the mean threshold of no-avatar condition was 0.54 (0.06) and full-body avatar's mean threshold was 0.27 (0.03). The differences between the two conditions in the two studies are almost identical. However, the thresholds in this study are lower than in prior work. It may be that the virtual mirror biases the threshold in some way.

V. Conclusion and discussion

In this study, we found that the presence of a self-avatar provides important information in the distance estimation task that people report whether they would step off a visual ledge or not. When self-avatars are presented to subjects, the maximum heights of the ledge that they step off are significantly lower than when the self-avatars are absent. This result is consistent with our prior study. Also, we found that the form of avatar (either full-avatar or line-avatar) makes no significant difference in such tasks. This finding could imply that, in these affordance judgments, rich polygonal models are not particularly useful. Thus, our finding is consistent with other studies [Thompson et al., 2004] that perception-action judgments are independent of the quality of computer graphics, while our study was more focused on the form of avatars. Although, additional study is needed to fully understand the relation between these judgments and computer graphics, it is useful for virtual environment designers to know in what situations high fidelity avatars are needed.

As discovered in Lin's study [Lin, Rieser, & Bodenheimer, 2015], people's thresholds in these stepping off ledge tasks in the real world is close to the thresholds of full-avatar condition. Thus, we can explicitly conclude that, for the immersive virtual environments that are primarily focused on people's spatial perception, the presence of a self-avatar can greatly improve the fidelity of the IVE. Although implementing a self-avatar would require extra equipment and time, the improvement of user's performance may be worth the effort. On the other hand, the form of self-avatar is not critical in virtual environment in terms of affecting IVE's fidelity and user's performance. Thus, if the authenticity of the graphic is not necessarily high, designers of these virtual environments may prefer to choose a less polygon-rich self-avatar to reduce the cost of developing.

Before our experiment, we assumed that two factors could affect people's performance in affordance judgment tasks in virtual environment. The first factor we assumed is the fidelity of the avatar, which was eliminated by our result. As in our study, the thresholds of subjects experienced low fidelity line-avatar have no significant differences to the thresholds of high fidelity full-avatar. The second assumption was that the difference of thresholds between no-avatar condition and full-avatar condition is caused because the self-avatar provides scale information about subject's own body. This assumption is supported by the results of our line-avatar, due to the fact that the thresholds in these two conditions (line-avatar and full-avatar) are similar; while the thresholds in the conditions with avatar and without avatar are significantly different. And the most noticeable similarity between these avatars is they were all carefully calibrated to represent subjects' own body dimension. This assumption needs further investigation. One way of studying this would be to partially scale the self-avatar and analyze subject's threshold, for example, we could stretch avatars leg length while compress the length of torso so the subject could experience the self-avatar at the same eye height.

We felt it was necessary, as part of our experiment, to provide a virtual mirror in order to help users observing their own models. According to our results, the presence of this virtual mirror may provide additional information to subjects as the thresholds were offset by a small amount, compared with our prior study. This is supported by subjects' feedback during the experiment, as some of them mentioned they had a better understanding about their self-avatar when looking into the virtual mirror. The effect of virtual mirror is also worth further investigation.

During our pilot study, we ran our experiment in a within subject manner, which was, every subject experienced all three condition. Although the order of conditions for each subject was randomized, the thresholds of three conditions of each subject were relatively close. It is most likely because the subjects studied the virtual environment in their first condition, and used those experiences in the following conditions.

In conclusion, for virtual environment that seek to simulate the scene of the real world, actions and the possibilities of action are important components. Having these affordances and actions match their counterparts in real world would increase the similarity of the virtual environment to the real world. And, our study supports that a size matched self-avatar is critical in increasing this similarity while the form of self-avatar is relatively not important. The reason why the self-avatar has this effect remains unproven.

As our results demonstrated, the fidelity of self-avatars is not critical in affordance judgment tasks, we would infer that it is the scale information that the self-avatar provides that really affects people's performance. Thus, we believe further investigation should be done in how people perceive this scale information. For example, the difference of people's affordances when they experience the same self-animated avatar in either a first person perspective view or third person perspective view. In another example, the difference of people's affordances when they either have normal view of their self-avatar, or only have sight on part of their self-avatar (the other part is rendered invisible). Additionally, as found in our study, a high fidelity self-avatar is not needed in distance estimation tasks, then it is worth further study in what situations a high fidelity self-avatars is required for better performance.

References

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol., 59*, 617-645.

Beall, A. C., Loomis, J. M., Philbeck, J. W., & Fikes, T. G. (1995). *Absolute motion parallax weakly determines visual scale in real and virtual environments.* Paper presented at the IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology.

Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. *Journal of Computer-Mediated Communication, 3*(2), 0-0.

Biocca, F. A., & Rolland, J. P. (1998). Virtual eyes can rearrange your body: Adaptation to visual displacement in see-through, head-mounted displays. *Presence: Teleoperators and Virtual Environments, 7*(3), 262-277.

Bodenheimer, B., Meng, J., Wu, H., Narasimham, G., Rump, B., McNamara, T. P., . . . Rieser, J. J. (2007). *Distance estimation in virtual and real environments using bisection*. Paper presented at the Proceedings of the 4th symposium on Applied perception in graphics and visualization, Tubingen, Germany.

Botvinick, M., & Cohen, J. (1998). Rubber hands' feel'touch that eyes see. *Nature, 391*(6669), 756-756.

Creem-Regehr, S. H., Willemsen, P., Gooch, A. A., & Thompson, W. B. (2005). The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual environments. *Perception, 34*(2), 191-204.

Durlach, N., & Slater, M. (2000). Presence in shared virtual environments and virtual togetherness. *Presence: Teleoperators and Virtual Environments, 9*(2), 214-217.

Eby, D. W., & Loomis, J. M. (1987). A study of visually directed throwing in the presence of multiple distance cues. *Perception & Psychophysics, 41*(4), 308-312.

Foley, J. (2007). Visually directed action: Learning to compensate for perceptual errors. *Journal of Vision, 7*(9), 416-416.

Franchak, J., & Adolph, K. (2014). Affordances as Probabilistic Functions: implications for development, perception, and decisions for action. *Ecological Psychology, 26*(1-2), 109-124.

Fukusima, S. S., Loomis, J. M., & Da Silva, J. A. (1997). Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 86.

Geuss, M., Stefanucci, J., Creem-Regehr, S., & Thompson, W. B. (2010). *Can I pass?: using affordances to measure perceived size in virtual environments.* Paper presented at the Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception: Classic Edition*: Psychology Press.

Grechkin, T. Y., Nguyen, T. D., Plumert, J. M., Cremer, J. F., & Kearney, J. K. (2010). How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Trans. Appl. Percept., 7*(4), 1-18. doi: 10.1145/1823738.1823744

Green, D., & Swets, J. (1966). Signal detection theory and psychophysics. 1966. *New York*.
Hay, J. C., Pick Jr, H. L., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic science, 2*(1-12), 215-216.
Henry, D., & Furness, T. (1993). *Spatial perception in virtual environments: Evaluating an architectural application.* Paper presented at the Virtual Reality Annual International Symposium, 1993., 1993 IEEE.
Hillis, K. (1999). *Digital sensations: Space, identity, and embodiment in virtual reality*: U of Minnesota Press.
Interrante, V., Ries, B., & Anderson, L. (2006). *Distance perception in immersive virtual environments, revisited.* Paper presented at the Virtual Reality Conference, 2006.
Interrante, V., Ries, B., Lindquist, J., Kaeding, M., & Anderson, L. (2008). Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence: Teleoperators and Virtual Environments, 17*(2), 176-198.
Jones, J. A., J. Edward Swan, I., Singh, G., Kolstad, E., & Ellis, S. R. (2008). *The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception*. Paper presented at the Proceedings of the 5th symposium on Applied perception in graphics and visualization, Los Angeles, California.
Knapp, J. M., & Loomis, J. M. (2004). Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators and Virtual Environments, 13*(5), 572-577.
Kunz, B. R., Creem-Regehr, S. H., & Thompson, W. B. (2009). Evidence for motor simulation in imagined locomotion. *Journal of Experimental Psychology: Human Perception and Performance, 35*(5), 1458.
Lin, Q., Rieser, J., & Bodenheimer, B. (2012). *Stepping over and ducking under: the influence of an avatar on locomotion in an HMD-based immersive virtual environment*. Paper presented at the Proceedings of the ACM Symposium on Applied Perception, Los Angeles, California.
Lin, Q., Rieser, J., & Bodenheimer, B. (2015). Affordance Judgments in HMD-Based Virtual Environments: Stepping over a Pole and Stepping off a Ledge. *ACM Trans. Appl. Percept., 12*(2), 1-21. doi: 10.1145/2720020
Lin, Q., Rieser, J. J., & Bodenheimer, B. (2013). *Stepping off a ledge in an HMD-based immersive virtual environment*. Paper presented at the Proceedings of the ACM Symposium on Applied Perception, Dublin, Ireland.
Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 906.
Loomis, J. M., & Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual environments. *Virtual and adaptive environments, 11*, 21-46.
Loomis, J. M., Silva, J. A. D., Philbeck, J. W., & Fukusima, S. S. (1996). Visual perception of location and distance. *Current Directions in Psychological Science*, 72-77.
Mark, L. S., Balliett, J. A., Craver, K. D., Douglas, S. D., & Fox, T. (1990). What an actor must do in order to perceive the affordance for sitting. *Ecological Psychology, 2*(4), 325-366.
McManus, E. A., Bodenheimer, B., Streuber, S., Rosa, S. d. l., B, H. H., #252, . . . Mohler, B. J. (2011). *The influence of avatar (self and character) animations on distance estimation,*

object interaction and locomotion in immersive virtual environments. Paper presented at the Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, Toulouse, France.

Meehan, M., Insko, B., Whitton, M., & Frederick P. Brooks, J. (2002). Physiological measures of presence in stressful virtual environments. *ACM Trans. Graph., 21*(3), 645-652. doi: 10.1145/566654.566630

Messing, R., & Durgin, F. H. (2005). Distance perception and the visual horizon in head-mounted displays. *ACM Transactions on Applied Perception (TAP), 2*(3), 234-250.

Michaels, C. F. (2003). Affordances: Four points of debate. *Ecological Psychology, 15*(2), 135-148.

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., & Bülthoff, H. H. (2010). The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. *Presence: Teleoperators and Virtual Environments, 19*(3), 230-242.

Nevin, J. A. (1969). SIGNAL DETECTION THEORY AND OPERANT BEHAVIOR: A Review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. 1. *Journal of the Experimental Analysis of Behavior, 12*(3), 475-480.

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*: Basic books.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research, 41*(24), 3145-3161.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on psychological science, 1*(2), 110-122.

Richardson, A. R., & Waller, D. (2005). The effect of feedback training on distance estimation in virtual environments. *Applied Cognitive Psychology, 19*(8), 1089-1108.

Richardson, A. R., & Waller, D. (2007). Interaction with an immersive virtual environment corrects users' distance estimates. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*(3), 507-517.

Ries, B., Interrante, V., Kaeding, M., & Anderson, L. (2008). *The effect of self-embodiment on distance perception in immersive virtual environments*. Paper presented at the Proceedings of the 2008 ACM symposium on Virtual reality software and technology, Bordeaux, France.

Ries, B., Interrante, V., Kaeding, M., & Phillips, L. (2009). *Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments*. Paper presented at the Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology, Kyoto, Japan.

Rieser, J. J., Ashmead, D. H., Talor, C. R., & Youngquist, G. A. (1990). Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception, 19*(5), 675-689.

Sahm, C. S., Creem-Regehr, S. H., Thompson, W. B., & Willemsen, P. (2005). Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Transactions on Applied Perception (TAP), 2*(1), 35-45.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., . . . Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS one, 1*(1), e39.

Slater, M., Frisoli, A., Tecchia, F., Guger, C., Lotto, B., Steed, A., . . . Sanchez-Vives, M. V. (2007). Understanding and realizing presence in the Presenccia project. *Computer Graphics and Applications, IEEE, 27*(4), 90-93.

Slater, M., Khanna, P., Mortensen, J., & Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *Computer Graphics and Applications, IEEE, 29*(3), 76-84.

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2008). Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience, 2*.

Slater, M., & Usoh, M. (1994). Body centred interaction in immersive virtual environments. *Artificial life and virtual reality, 1*, 125-148.

Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact., 2*(3), 201-219. doi: 10.1145/210079.210084

Thelen, E. (1995). Motor development: A new synthesis. *American psychologist, 50*(2), 79.

Thompson, W. K., Willemsen, P., Gooch, A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence, 13*(5), 560-571.

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., & Frederick P. Brooks, J. (1999). *Walking > walking-in-place > flying, in virtual environments*. Paper presented at the Proceedings of the 26th annual conference on Computer graphics and interactive techniques.

Waller, D., & Richardson, A. R. (2008). Correcting distance estimates by interacting with immersive virtual environments: Effects of task and available sensory information. *Journal of Experimental Psychology: Applied, 14*(1), 61.

Warren Jr, W. H., & Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 371.

Warren, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 683.

Willemsen, P., Colton, M. B., Creem-Regehr, S. H., & Thompson, W. B. (2004). *The effects of head-mounted display mechanics on distance judgments in virtual environments.* Paper presented at the Proceedings of the 1st Symposium on Applied perception in graphics and visualization.

Willemsen, P., Gooch, A. A., Thompson, W. B., & Creem-Regehr, S. H. (2008). Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments, 17*(1), 91-101.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review, 9*(4), 625-636.

Zhang, H., Yu, C., & Smith, L. B. (2006). *An interactive virtual reality platform for studying embodied social interaction.* Paper presented at the Proceedings of the CogSci06 Symposium Toward Social Mechanisms of Android Science.

Aymerich-Franch, L., Kizilcec, R. F., & Bailenson, J. N. (2014). The relationship between virtual self similarity and social anxiety. *Frontiers in human neuroscience, 8*.

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences, 110*(31), 12846-12851.

Bodenheimer, B., & Fu, Q. (2015). *The Effect of Avatar Model in Stepping Off a Ledge in an Immersive Virtual Environment*. Paper presented at the ACM Symposium on Applied Perception.

Cohen, M. M., Walker, R. L., & Massaro, D. W. (1996). Perception of synthetic visual speech *Speechreading by humans and machines* (pp. 153-168): Springer.

Fox, J., Bailenson, J. N., & Tricase, L. (2013). The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior, 29*(3), 930-938.

Gibson, J. J. (2014). *The Ecological Approach to Visual Perception: Classic Edition*: Psychology Press.

Grassi, M., & Soranzo, A. (2009). MLP: A MATLAB toolbox for rapid and reliable auditory threshold estimation. *Behavior research methods, 41*(1), 20-28.

Green, D. M. (1990). Stimulus selection in adaptive psychophysical procedures. *The Journal of the Acoustical Society of America, 87*(6), 2662-2674.

Green, D. M. (1993). A maximum - likelihood method for estimating thresholds in a yes – no task. *The Journal of the Acoustical Society of America, 93*(4), 2096-2105.

Hill, H. C., Jinno, Y., & Johnston, A. (2003). Comparing solid body with point-light animations.

Hodgins, J. K., Brien, J. F., & Tumblin, J. (1998). Perception of human motion with different geometric models. *Visualization and Computer Graphics, IEEE Transactions on, 4*(4), 307-316.

Lin, Q., Rieser, J., & Bodenheimer, B. (2012). *Stepping over and ducking under: the influence of an avatar on locomotion in an HMD-based immersive virtual environment*. Paper presented at the Proceedings of the ACM Symposium on Applied Perception, Los Angeles, California.

Lin, Q., Rieser, J., & Bodenheimer, B. (2015). Affordance Judgments in HMD-Based Virtual Environments: Stepping over a Pole and Stepping off a Ledge. *ACM Trans. Appl. Percept., 12*(2), 1-21. doi: 10.1145/2720020

Linkenauger, S. A., Leyrer, M., Bülthoff, H. H., & Mohler, B. J. (2013). Welcome to wonderland: The influence of the size and shape of a virtual hand on the perceived size and shape of virtual objects. *PloS one, 8*(7), 1-16.

Michaels, C. F. (2003). Affordances: Four points of debate. *Ecological Psychology, 15*(2), 135-148.

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*: Basic books.

Thelen, E. (1995). Motor development: A new synthesis. *American psychologist, 50*(2), 79.

Thompson, W. K., Willemsen, P., Gooch, A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence, 13*(5), 560-571.

van der Hoort, B., Guterstam, A., & Ehrsson, H. H. (2011). Being Barbie: the size of one's own body determines the perceived size of the world. *PloS one, 6*(5), e20195.

Vignais, N., Bideau, B., Craig, C., Brault, S., Multon, F., Delamarche, P., & Kulpa, R. (2009). Does the level of graphical detail of a virtual handball thrower influence a goalkeeper's motor response? *Journal of sports science & medicine, 8*(4), 501.

Vinayagamoorthy, V., Brogni, A., Gillies, M., Slater, M., & Steed, A. (2004). *An investigation of presence response across variations in visual realism.* Paper presented at the The 7th Annual International Presence Workshop.

Yee, N., & Bailenson, J. N. (2009). The difference between being and seeing: The relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychology, 12*(2), 195-209.