EFFECT OF ANIMATED SELF-AVATAR IN VIRTUAL ENVIRONMENTS

by

Mukund Raj

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

School of Computing

The University of Utah

August 2013

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Mukund Raj**

has been approved by the following supervisory committee members:

William B. Thompson, Chair **5/30/2013** Date Approved

Miriah Meyer, Member **5/30/2013** Date Approved

Sarah H. Creem-Regehr, Member **5/30/2013** Date Approved

and by **Alan Davis**, Chair of

the School of **Computing**

and by Donna M. White, Interim Dean of The Graduate School.

ABSTRACT

Animated avatars are becoming increasingly prevalent in three-dimensional virtual environments due to modern motion tracking hardware and their falling cost. As this opens up new possibilities and ways of interaction within such virtual worlds, an important question that arises is how does the presence of an avatar alter the perception and performance of an action in a virtual environment when a user interacts with an object in the virtual environment through their avatar. This research attempts to answer this question by studying the effects of presence of an animated self-avatar in an object manipulation task in a virtual environment.

Two experiments were conducted as part of this research. In Experiment 1, the feasibility of an interaction system involving animated self-avatars to manipulate objects in a virtual environment was examined. It was observed that the presence of self-avatars had an affect on the performance of a subset of subjects. Male subjects with gaming experience performed similarly across both visual feedback conditions while female subjects who also had low gaming experience performed better in the condition with avatar feedback than in the condition without avatar feedback. In Experiment 2, we further analyzed the effect of presence of self-avatar visual feedback by looking at the effect of visual immersion in the virtual environment, task difficulty, and individual difference factors such as spatial ability and gaming experience. It was observed that difficult trials were completed significantly faster by subjects in the avatar feedback condition while in the case of the easy trials, there was no significant difference between performance of subjects in the avatar and sphere feedback conditions. No significant interaction was observed between visual feedback condition and either immersiveness or individual difference factors.

For my parents, Divya and ammamma

CONTENTS

CHAPTERS

APPENDICES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Dr. William Thompson and Dr. Sarah Creem-Regher for all the guidance and support provided by them to me over the past two years throughout the course of this research. I am also extremely thankful to Dr. Jeanine Stefanucci for her valuable input during the early stages of this research as well as her help to prepare me for my first conference talk. I would like to thank Dr. Miriah Meyer for her candid observations and feedback from the review of my work.

I would also like to thank Kaushik for all the help he provided at various stages of my research, Kristina for her assistance in analyzing the data, and fellow VPSC lab members Garrett, Lace, Kyle, Mike, Margaret, Grace, and David for their help and feedback while testing the prototypes. The support from my friends Satish, Joanita, Kaushik, Avantika, and Prasanna also helped me immensely during the course of this research.

Lastly, I would like to acknowledge the contribution of all those who volunteered to participate in this study as subjects.

This work was supported by the National Science Foundation under Grant No. 0914488 and 1116636.

CHAPTER 1

INTRODUCTION

Although the quality of visual realism of environments is improving at a tremendous pace, there is still a tremendous scope of improvement to make interaction with objects in such environments more natural.

1.1 Motivation

A particularly interesting development for the goal of advancing interaction methods with virtual objects is the availability of low-cost motion tracking and gesture recognition commodity hardware such as the Microsoft Kinect, Sony PlayStation Move, and the Nintendo Wii Motion Plus. Although the initial motivation for building these devices was their use as a gaming controller, other potential applications of such gesture tracking hardware were quickly realized to be far beyond gaming to also include areas such as education, training, medicine, research, and even building virtual three-dimensional (3D) representations of surroundings [Izadi et al. 2011]. This potential has been recognized by designers of next generations of motion sensing and gesture recognition devices which are being developed with specifically nongaming applications in mind. The Leapmotion Leap is one such device which is expected to be released in the near future. These devices make it possible to have virtual environments in which people could interact with objects in a more natural way than possible with conventional methods. Also, this interaction is possible without any cumbersome equipment needed to be worn by the user as in the case of current full body motion capture systems. The data from the gesture recognition hardware can be routed to animate realistic 3D virtual model representation of the users also known as avatars. It is possible that in the future, the orientation data from high-precision gyroscopes in the current generation of smart phones can be used to animate the wrist joint of a persons avatar with no need for additional specialized hardware.

In addition to the advances in tracking hardware and rendering capability, there have also been strides made in portability of realistic virtual environments. Web-based 3D graphic

APIs such as WebGL and Away3D are able to benefit from GPU-based hardware acceleration while rendering high-quality virtual worlds within a web browser itself. Essentially, we are seeing that building virtual worlds and animating self-avatars is now becoming possible at a fraction of the cost of the motion capture and rendering equipment that was required for such a setup only a few years ago. With self-avatar-based virtual worlds well poised to become more widespread, there is still a lack of understanding of how the presence of an animated self-avatar might influence interaction of users with the virtual environment. It is particularly interesting to know how the self-avatar effects the ability of the users while performing tasks in the virtual environment.

1.2 Research Problem

The main goal of this research is to see if the presence of a self-avatar has an effect on the performance of an object manipulation task in a virtual environment based on a desktop or flat panel display. A large-screen stereoscopic flat panel is able to provide a much higher level of visual immersion to the user in a virtual environment than a regular desktop display. In this study, we also investigate if the level of visual immersion of the user interacts with the presence of self avatars by simulating two different levels of visual immersion. In the context of this research, levels of visual immersion refers to different levels of visual information provided by the display itself as opposed to difference in the level of information in the rendering of the virtual world. The high visual immersion condition refers to a stereoscopic image displayed in a nonilluminated room while a low visual immersion condition refers to a nonstereo image displayed in a well-lit room.

The specific aims for this research were to answer the following questions and support them with data from user studies:

1. Is there an effect due to presence of an animated self-avatar on an object manipulation task in a flat-panel-based virtual environment?
2. Is the effect's strength affected by a higher level of visual immersion provided by a stereoscopic display?

In the context of this research, an avatar is a digitally rendered model representation of a person in a virtual environment. The avatar's motions are animated such that they mimic the actions of the user. Although the idea of interacting with objects in a virtual environment has been around for a while, rigorous studies using a scientific approach to understand the differences and similarities in operating in a virtual environment as compared to the working in the real world have only just started due to recent technological advances that make such an interface feasible. Of particular interest is the effect of interacting within a virtual environment via a self-avatar. A self-avatar is a first-person representation of the user themselves. Animating self-avatars in a virtual environment using the actual actions of the user can create a compelling level of visual immersion for the user in the virtual environment [Lok et al. 2003]. Applications of being able to do this properly are tremendous, especially in areas where the actual environment in the real world may be hazardous or cost-ineffective for training purposes. As we are likely to see an increasing number of virtual environments and users interacting with those environments via their virtual avatars, it important to understand how self-avatars contribute to the user experience and to use this information to enable more effective design and use of virtual worlds. Further, current studies mostly focus on visually immersive virtual environments such as HMDs. These still remain expensive and are not very common outside of research labs. The question of whether and how the presence of a self-avatar would have an effect on the interactions of a user within a virtual environment on a desktop display or flat panel display remains unanswered. It is important to answer this question since applications built over such displays are more likely to spread quickly and be used in applications as opposed to more expensive systems given similar effects.

The hypothesis of this research is that the presence of an animated self-avatar can help facilitate perception and action in a virtual environment. Specifically, with presence of a self-avatar, users would show an improvement in performance in an object manipulation task which involves matching the orientation of an unfamiliar object with the orientation of a similar target object at a different orientation. Another part of this research is to study if the stereoscopic cues may interact with the effect of presence of a self-avatar in an object manipulation task. Here the expectation is that higher visual immersiveness of the stereoscopic display would result in a stronger effect by making the visual feedback closer to that in the real world for a similar task.

An effect itself, if present, could be attributed to various reasons. The avatar figure may provide the users with a familiar reference scale with which to compare distances and sizes in the virtual environment. It is also possible that the avatar provides a familiar orientation indicator when manipulating an unfamiliar object in the virtual world that could aid performance. An additional frame of reference provided by the avatar may also be beneficial for the user to manipulate objects. Matching the visual feedback to the user while interacting with the virtual environment to the visual feedback expected in the real world while performing a similar task could facilitate a higher sense of embodiment while performing tasks in a virtual environment and in turn lead to better performance.

To recapitulate, understanding the effect induced by avatar presence in a virtual environment is currently an active research topic and the question whether there is an effect in action tasks due to presence of a self-avatar is still unanswered. It is further not known whether an effect, if any, will be strong enough on a desktop display as opposed to a more visually immersive virtual environment. The main goal of work described in this paper is to ascertain the effect of presence of a self-avatar on performance of an object manipulation task in a virtual environment. Further, we also explore if the effect in turn may also depend on other factors such as level of visual immersiveness of the display, difficulty of the task, or individual differences among users.

1.3 Overview

Chapter 2 of this thesis builds up the necessary background by exploring important prior research in perception, avatars in virtual environments, and interaction with objects in a virtual environments. Next, two studies, Experiment 1 and Experiment 2, are presented with a discussion of their results in Chapter 3 and Chapter 4, respectively. Experiment 1 is a pilot study that investigated the feasibility of an interface in which the user interacts with objects in the virtual world via an animated self-avatar. In this study, in addition to seeing that such an interface was indeed feasible, we also observe an effect of self-avatars on a qualified set of users based on individual differences. These findings from the pilot study motivated a design of Experiment 2 which focused on the effects of interaction between presence of a self-avatar and level of immersion, individual differences, and task difficulty. The interface in Experiment 2 was made easier to operate using observations made while conducting the pilot study. Finally, a general discussion analyzing the results from both experiments followed by some suggestions for future work are presented in Chapter 5.

CHAPTER 2

BACKGROUND

Prior work in the area of avatars has been largely focused on social interaction and presence in the virtual world. Studies investigating the effect of avatars on the cognition and action performance have only recently started and there is not a lot of research in this area yet. The work that has been done also mostly deals with visually immersive virtual environments and not on desktop-based virtual environments. The following paragraphs give a brief overview of existing literature on avatars in a virtual environment. Next, some relevant and interesting research related to interaction in a virtual environment is described. Also important to this research is selection of a task for evaluating performance, hence, research related to manipulation of objects in a virtual environment is also touched upon. People interacting with a virtual environment via an avatar animated using a motion tracking hardware can get a sense of embodiment within their avatar in several ways. These include efference which is a name given to motor signals from the central nervous system to the periphery of the body, afference which is the sensory information from sensors in the peripheral nervous system back to the central nervous system, and very importantly, visual feedback from viewing the avatar. Proprioception, which is an example of afference, is a strong cue for how users identify with the avatar [Balslev et al. 2007; Tsakiris et al. 2006].

Studies involving avatars in a virtual environment have looked at the effect of sense of presence experienced by the users. It was suggested by Slater et al. [1995] that the subjective amount of presence in the virtual environment experienced by the participants can be enhanced provided they associate strongly with the avatar. Also, Sanchez-Vives et al. [2010] concluded that illusions of ownership and proprioceptive displacement could be induced with only visuomotor stimulation, specifically, in absence of tactile stimulation. Six claims from the emerging viewpoint of embodied cognition which holds that cognitive processes are deeply rooted in the body's interaction with the world are distinguished and evaluated in Wilson [2002]. The claims that state we offload cognitive work to the environment and that environment is part of the cognitive system are consistent with the idea that avatars would aid in object manipulation in virtual environments by making the interface closer to the real world. The task in Lok et al. [2003] involved manipulating real objects while viewing a virtual simulation of the same objects and either generic or faithful self-avatars of the users hands. This work evaluated both task performance and subjective presence and concluded that the fidelity of the motion of the avatars was more important for a believable self-avatar than the visual fidelity of the avatar, though users indicated a preference for the more self-accurate avatar. We see from the work in Ban et al. [2012] that the shape of the object in a real hand need not match exactly with the shape of the virtual object being manipulated in order for us to perceive the shape of the real handled object to be similar to the virtual object being manipulated.

More recent work using better quality avatars and motion capture systems have shown that being able to see a rendered version of your own body increases the accuracy of distance judgments in virtual environments presented using a head-mounted display [Mohler et al. 2010; Phillips et al. 2010]. The work in Mohler et al. [2010] explored the effect of prior experience with a tracked self-avatar on the accuracy of subsequent distance judgments in a virtual environment. Although the above studies provide interesting insight about how the effects of user control of the avatar affected perception of the overall environment, they do not deal with how the presence or user control of the avatar affected performance in action tasks in the virtual environment. This question is especially interesting for tasks such as manipulation of virtual objects close to the avatar so as to see if performance in the virtual environments can be improved and brought closer to real-world performance of manipulating real objects with the hand.

There have also been studies evaluating the effect of a user avatar on interaction on a desktop display. We see the use of arrow-shaped symbols to represent a nonrealistic user avatar in Kadri et al. [2007]. It was found that the direction of the arrow influenced choice of the interaction strategy of users in a virtual object manipulation task. They found that if the avatar was represented by a left-pointing arrow, users preferred to grab the object from the right and vice versa. In other work, the effect of the realism of an avatar on a user's perception was also studied. In Hodgins et al. [1998] it was found that a viewer's perception of motion characteristics is affected by the geometric model used for rendering. It was observed that subjects were more sensitive to motion changes displayed through a polygonal model of the human body than through a stick figure model, leading to the conclusion that stick figures may not always have the required complexity to ensure that the subtleties of the motion are apparent to the viewer. McDonnell et al. [2012] investigated whether using a realistic rendering of a model of a person's upper body does in fact produce a more negative perceptual response than using a lower quality or a stylized rendering in accordance with the theory of the uncanny valley [Mori 1970]. It was observed that realistic virtual characters may be perceived well as long as the motion did not contain artifacts.

In Raj et al. [2012], we conduct a pilot study in which users manipulate 3D objects on the screen with and without the animated self-avatar present. It was seen that individual differences could have a significant effect on the performance of users in spatial manipulation tasks. In this study, we found an effect of the animated self-avatar feedback qualified by the individual differences among the subjects. The presence of the self-avatar as part of the interface seemed to benefit a subset of users. An effect of individual difference was also observed in Jurnet et al. [2005]. It was found that individual differences affects the sense of presence in virtual environments based on the personality of participants.

Other work surveyed involved evaluating avatar-based object manipulation interfaces in virtual environments. In Poupyrev et al. [1998], the authors perform one of the earliest comparisons of different methods of object manipulation in a virtual environment and provide a framework for conducting such a comparison. However, the avatars in the interface are not high fidelity and the study also does not consider the effect of individual differences. In Bowman and Hodges [1997], a high-level comparison of two methods of object manipulation was done. The first is the go-go technique in which the hand of the avatar is elongated to reach objects beyond natural reach and the other is ray-casting in which a virtual ray is cast from the hand toward the object that is to be manipulated. The focus here is on just subjective performance and again, individual differences are not considered.

On building interfaces for manipulating objects in the virtual world, Jacob et al. [1994] present and validate the hypothesis that matching the control structure of an input device with the perceptual structure of the task leads to better performance than when there is no such matching. It further adds that performance improves when the structure of the perceptual space of a graphical interaction task mirrors the control space of the input device.

One of the key earlier works involving rotation of 3D objects in the virtual world was done by Hinckley et al. [1994] which compared performance of users in rotating 3D objects in the virtual environment using four different interaction methods. It was found that 3D techniques of manipulation were more suited to object rotation than 2D techniques due to integrated degrees of freedom provided by the 3D interaction devices. It should be noted that the tactile feedback to the user in the case of the 3D interaction devices is closer to the actual feedback to a person when rotating an object in the real world. In Hinckley et al. [1997], a study is conducted of performance in a cooperative bimanual task in a physical environment. It introduces the concept that performance in hard and easy tasks are affected differently by interchanging the roles of the dominant and the nonpreferred hand. In hard tasks, performance is better when the nonpreferred hand orients an object and the dominant had manipulates the tool to interact with it while for an easy task, the performance did not vary significantly even after swapping the mapping between the task and the hands. In Hinckley et al. [1998], a similar study is conducted in a virtual environment in which users try to match the orientation of two different virtual objects by manipulating two physical controllers in two hands and the performance in this task is compared with the performance of this task conducted in unimanual mode. It is observed that the bimanual mode has benefits over the unimanual mode interface for appropriately matched tasks and interface. Some very useful concepts for design, evaluation, and analysis of input interfaces are expounded in Hinckley [2002] and Hinckley et al. [2004].

An interesting result that compared virtual-world task performance to real-world task performance in an action task was seen in Lok et al. [2003]. It was found that manipulating a real object in the hand brings performance of an object manipulation task in a virtual world closer to that of manipulating the object in the real world. This result is also is one of the key underpinnings of the interface for this study. We used an object orientation matching task in this research. Pani et al. [1995] found that it is harder to mentally rotate objects about axes that are oblique with respect to the environment than rotating them about nonoblique axes.

Presence of an orientation indicator has also been found to aid rotation tasks. This can be an object whose orientations are easily discernible and whose orientations are linked to those of the object being manipulated. These can be placed either overlapping the object or adjacent to the object. In Khan et al. [2008], an in-scene 3D widget that is used to help the user keep track of the object orientation and interact with the object was found to be very effective. Also in Ziemek et al. [2012], two different kinds of orientation indicators are compared for their ability to aid the users in a static orientation comparison task keeping in mind individual differences in the spatial abilities of the users. The study found varying effect of the presence of the orientation indicators on the users which depended upon their spatial abilities.

CHAPTER 3

EXPERIMENT 1

This experiment was a pilot study to analyze the effect of having a visual feedback of an avatar hand closely following the motion of the user's real hand while the user manipulated an object in a virtual environment. Since we did not find evidence of an earlier study using a commodity gesture-based controller to animate the avatar in real time for an object manipulation task, we wanted to test the feasibility of such a system being usable by general users who did not have any significant training or experience in using such an interface. The performance of subjects in the task with the visual feedback in the form of the avatar was compared with the performance of subjects in another condition in which the visual feedback of an accurate avatar was replaced by a hand-sized sphere. If the presence of the avatar affected the performance of the subjects in the task, the performance of the subjects in both conditions was expected to be significantly different from each other. However, if there was no benefit of the presence of the avatar in manipulating the objects, then no significant difference in performance of the subjects in both the conditions was expected.

3.1 Method

This experiment was designed keeping in mind the need to test the feasibility of a Kinect-driven object manipulation interface.

3.1.1 Participants

Twenty-three (13 male and 10 female) University of Utah students participated in the study for a compensation of $5 dollars. All participants used their right hand as their dominant hand. In addition, all participants gave written consent prior to participation but were naive to the specifics of the experiment until they had completed participation.

3.1.2 Apparatus

The system recovered the joint orientation information from the data streamed from the Microsoft Kinect in order to animate either the partial self-avatar of the user, or a sphere, depending upon the respective feedback condition to provide visual feedback to the participant. The virtual environment was rendered using WorldViz Vizard and FAAST [Suma et al. 2011] was used to read and process the Kinect data feed to get the user's joint orientations. In the self-avatar condition, participants viewed a virtual arm closely following the movements of their real arm. For those in the sphere condition, a white sphere was rendered instead of the avatar at the position where the hand of the avatar would have been rendered. Thirteen objects were presented to the participants: one for the practice trials, and 12 additional objects for the experimental test trials. The practice object was in the shape of a jack with 6 bars with different colored ends (see Appendix A). The objects used in the test trials were a subset of the anatomical objects used in Ziemek et al. [2012] created using digital embryos [Brady and Kersten 2003] (see Appendix A). The radius of the bounding sphere of the objects ranged from 0.16m to 0.29m and were rendered 1.4m in front of the avatar's position. All objects were displayed on an Asus ProArt Series 24.1 inch display. The distance of the participant from the display was 1.81m and the geometric field of view of the participant was matched with the rendered display field of view. Figure 3.1 show the setup of the experiment. The wireless IC3 was attached to a strap which was then fastened over the right hand of the participant such that the orientation sensor was situated over the back of the participant's right hand.

3.1.3 Design

A between-subject design was used with the visual feedback condition as a between-subjects factor. Two randomized orders of experimental trials were also manipulated between subjects. Participants performed 6 practice trials, and 12 experimental trials, for a total of 18 trials. Objects presented on each trial differed from the target orientation from 28 up to 177 degrees of rotation required to match the orientations by rotating the object about the most optimal axis.

3.1.4 Procedure

Participants were instructed that the goal of the experiment was to manipulate the object on the right side of the screen to match the orientation of the object on the left side of the screen as quickly as possible (see Figures 3.2 and 3.3). Participants were equipped with an InertiaCube mounted on a band that was strapped around their palm and a mouse held by the participant in their right hand, and led to the starting location. They were then informed of the two modes in which they would be able to manipulate the object: swipe or twist. The experimenter then demonstrated the distinction between the modes

Figure 3.1. Setup for Experiment 1. The subject position is 1.81m away from Asus ProArt Series 24.1 inch display

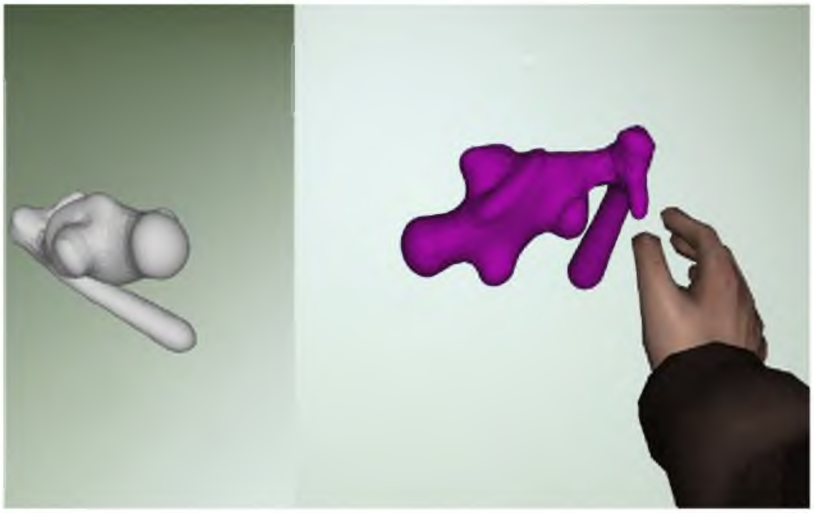


Figure 3.2. A screenshot of the self avatar feedback condition.

on a real object with their hands. The swipe mode used a drag motion performed with the arm similar functionally to methods such as the virtual sphere or arcball, where the object rotates in the direction of the drag. The twist mode referred exclusively to changes in the orientation due to rotation of the object about the wrist axes, closely resembling what would be expected in the real world. In both conditions, the rotation of the wrist about the axis joining the elbow to the wrist was also accurately mapped using orientation data from

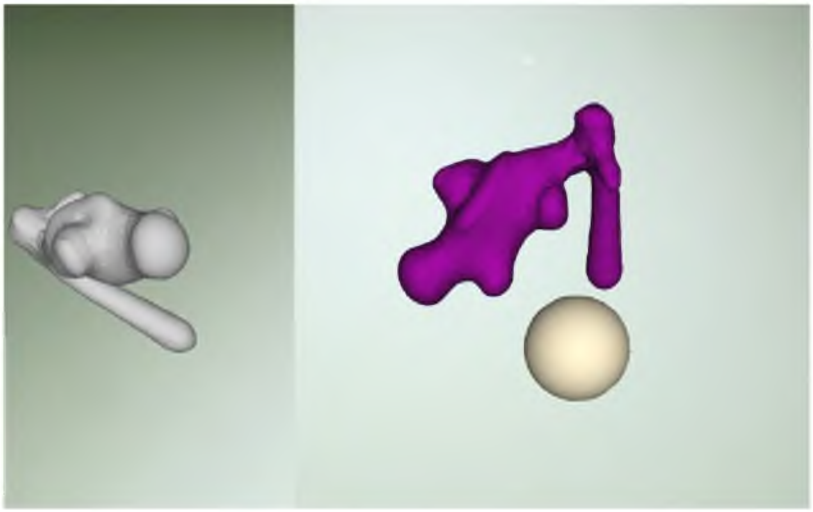


Figure 3.3. A screenshot of the sphere feedback condition.

the wireless IC3. Visual feedback for wrist rotation was evident in the hand condition but absent in the sphere condition. Finger joints were not animated. In both rotation modes, manipulation was only possible if the hand was close to the object, conveyed by a change in the color of the object. There was also the option of ratcheting to accomplish large rotations as a sum of smaller motions and scaling of the user action based on the speed of movement to maintain precision as well as range of manipulation.

Participants were informed that a left or right mouse click would perform the swipe or twist, respectively, and were asked to demonstrate their understanding of the distinction between the modes. Following the instructions, participants performed the practice and test trials. All trials began with the presentation of two objects at differing orientations. After three seconds, a prompt on the screen instructed the participant to begin. At this point, participants could use the two modes of object manipulation to match the objects. When the orientation of the object was within 15 degrees of the orientation of the target object, a "match detected" prompt appeared on the screen. During experimental trials, if 90 seconds elapsed between the onset of the trial before a match was detected, a prompt indicated that the allotted time had expired. This time interval was chosen to allow most of the trials to be completed but also to reduce the chance of frustration or fatigue on an individual trial level. After completion of the trials, participants were given a brief video game experience questionnaire, with rating scales ranging from 1-7 on first-person video game, gesture-based game, and third-person game (such as racing, sports) experience.

3.2 Results

We did not find a significant overall difference in the average rotation times between participants in the two visual feedback conditions; however, it was observed that female participants were significantly faster with avatar feedback while male participants performed similarly in both the conditions. On analyzing the number of timeouts per participant, we found a trend similar to what we found for the average times. There was no overall significant difference in the number of timeouts between participants in the two visual feedback conditions, but female participants had significantly lesser timeouts with avatar feedback while male participants in both visual feedback conditions had a similar number of timeouts. From the analysis of the usage of manipulation modes (swipe and twist), it was observed that males used both the modes equally while females employed a significantly larger amount of swipe mode as compared to twist mode. The visual display condition of the participant did not influence total number of left or right clicks for both male and female participants.

Time to rotate each object was averaged across the 12 trials for each participant to get the rotation time for each participant. Rotation time was slightly faster overall for the self-avatar (30.36 sec) versus the sphere conditions (34.81 sec). A 2 (visual feedback) x 2 (gender) ANOVA was performed on average response time. While there was not a significant difference overall between the self-avatar and sphere conditions ($p = .43$), there was a significant visual feedback x gender interaction, $F(1, 19) = 4.49$, $p < .05$, $\eta_{\mathrm{p}}^2 = .19$, showing that while males and females showed no difference in performance on the self-avatar task ($p = .84$), males outperformed females on the sphere task ($t(9) = 3.32$, $p < .01$). See Figure 3.4.

In the analysis of the number of timeouts per participant, we found that female participants had significantly more timeouts in the sphere condition as compared to the avatar feedback condition while male participants in the two visual feedback conditions did not show a significant difference in the number of timeouts. A 2 (visual feedback) x 2 (gender) ANOVA was performed on average number of timeouts (in which the participant did not complete the rotation task within the allotted time of 90 sec). Similar to the rotation time analysis, the only significant effect found was a visual display x gender interaction, $F(1, 19) = 6.76$, $p < .02$, $\eta_{\mathrm{p}}^2 = .26$, showing that there was no difference between male (1.43) and female (.80) performance for the self-avatar condition ($p = .52$), but significantly more timeouts for females (2.80) compared to males (.17) in the sphere condition ($t(9) = 3.30$, $p < .01$).

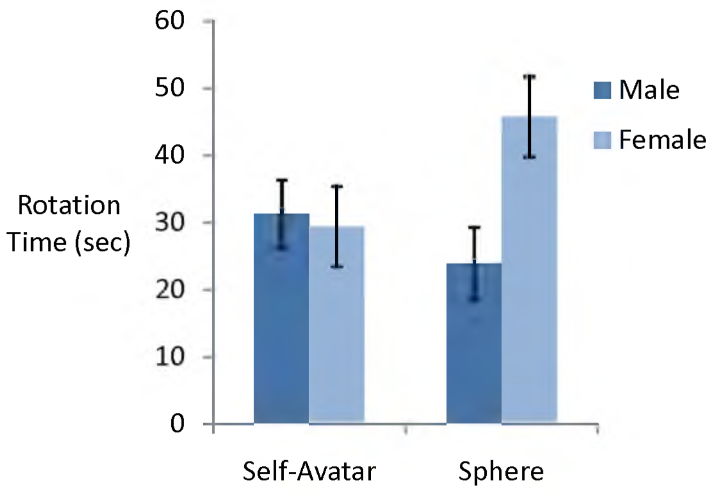


Figure 3.4. Mean rotation time (±1 SE) for the self-avatar and sphere conditions by gender.

Given the two different modes to rotate (swipe and twist performed with the left and right mouse clicks, respectively), we also analyzed total number of left and right mouse clicks. We used a 2 (visual feedback) x 2 (gender) x 2 (click: left vs. right) mixed ANOVA with left/right click as a within-subjects variable. The analysis showed a greater number of total left (87.32) versus right (66.28) clicks, $F(1,19) = 4.52$, $p < .05$, $\eta_{\mathrm{p}}^2 = .19$, as well as a click x gender interaction, $F(1,19) = 4.44$, $p < .05$, $\eta_{\mathrm{p}}^2 = .19$. The interaction revealed that males used swipe (left click) and twist (right click) modes equally ($p = .98$), but females used the swipe more than the twist mode, $t(9) = 2.27$, $p < .05$. See Figure 3.5.

Finally, we examined the influence of self-reported video game experience on rotation time by performing separate bivariate correlations between average rotation time and the three video game rating scales for both the self-avatar and sphere conditions. Gesture-gaming experience did not correlate with rotation time for either display conditions. Both first- and third-person game experience correlated with rotation time across the display conditions (see Figure 3.6), but effects were greater for the sphere versus the self-avatar display. (First-person: Self-avatar $R = -.55$, $p < .06$, Sphere $R = -.71$, $p < .02$; Third-person: Self-avatar $R = -.42$, $p < .16$, Sphere $R = -.77$, $p < .01$).

3.3 Discussion

This pilot experiment demonstrated the feasibility of using off-the-shelf technology, such as the Microsoft Kinect, to drive user interfaces that aid in manipulating 3D objects on a desktop display. The results also suggest that care should be taken to understand the

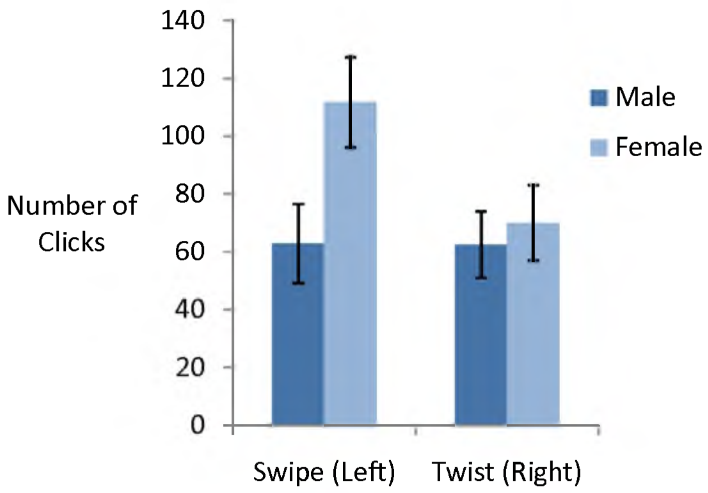


Figure 3.5. Average total number of left and right button clicks (±1 SE) by gender.

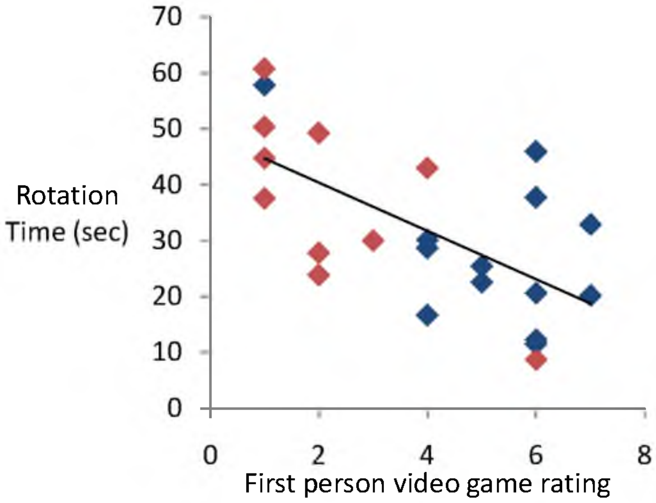


Figure 3.6. Correlation ($R = -.63$, $p < .001$) between first-person video game experience and rotation time. There is high overlap between gaming experience and gender (red symbols = females, blue symbols = males).

individual differences among users that could interact with visual feedback provided to the user. In this experiment, the gender and video game experience of the participant influenced performance, specifically in the sphere display condition which presented less information about the relationship between the orientation of the participant's real hands and the orientation of the displayed interface entity. Given the individual differences found only in the sphere condition, it is likely that the self-avatar provided additional body-based information that may be beneficial to a broader population of users.

There are a few experimental design issues to consider when interpreting the results of this experiment. First, there was a fixed maximum time limit for completion of each of the trials. This restriction was placed in order to reduce frustration and fatigue in the case the participant found a trial particularly challenging. Future work could give users unlimited time to more fully assess reaction time for successful manipulation, but it would need to find a solution for the participant fatigue problem. Also, while there were strong correlations between gaming experience and task performance, it is important to note that in the subject pool for this experiment, gender differences highly overlapped with videogame experience (see Figure 3.6). Thus, it is cannot be deduced from the results of this experiment whether the differences seen in the two feedback conditions as a function of individual differences will replicate with a greater range of female gamers or male nongamers. However, the individual differences affecting performance in the sphere condition are not too surprising given previous work demonstrating gender differences in spatial abilities, defined as a range of skills involving the mental representation and manipulation of information about geometric entities [Hegarty and Waller 2005]. The best documented gender-related performance difference in spatial abilities is a male advantage in mental rotation tasks, in which viewers are asked to determine the congruence between images of two static objects presented at differing orientations [Linn and Petersen 1985]. Also, some individuals find it very challenging to determine correspondence between multiple views of 3D objects, particularly when they are irregular shapes [Ziemek et al. 2012].

CHAPTER 4

EXPERIMENT 2

This experiment was conducted as a follow-up study to the pilot study described in Experiment 1. In the previous experiment, an effect of the presence of a self-avatar on the performance on the object manipulation task was observed for a subset of participants. In this experiment, the interface was made easier to use based on observations from the pilot experiment in order to minimize interference by uncontrolled factors such as a confusion in deciding between the left and right button click for selecting the desired mode of manipulation. The objective of Experiment 2 was to conduct a more rigorous study to analyze the effect of having a visual feedback in the form of an avatar on the performance of object manipulation tasks in a virtual environment. Additionally in this experiment, we also wanted to see if visual immersiveness of the virtual environment, individual differences of participants, or the difficulty of the task interacted with the visual feedback condition while performing the object manipulation task. This was a larger study with more subjects than in the pilot study. As in the earlier experiment, if participants in a particular condition were able to benefit from the presence of the avatar, then we expected a significant improvement in the performance of the group in that condition as compared to the other conditions. However, if there was no benefit due to the presence of the self-avatar in manipulating the objects, then we expected the performance of participants in groups with different visual feedback to be similar. In order to control for individual differences in spatial abilities among participants, all participants were tested for egocentric and exocentric spatial abilities before starting with the object orientation matching trials (see section 4.1.4.1).

4.1 Method

This experiment was designed keeping in mind the learnings from the interface of Experiment 1.

4.1.1 Participants

Seventy-three (37 male and 36 female) University of Utah students participated in the study for a compensation by either participation credit or $10 dollars. We only used data from subjects who used their right hand as their dominant hand. In addition, we excluded data of outliers who timed out on more than half of the trials. All participants gave written consent prior to participation but were naive to the specifics of the experiment until they had completed participation. The distribution of participants in each of the conditions was as follows - 18,17,18,15 for avatar present - immersive, avatar present - nonimmersive, avatar not present - immersive, and avatar not present - nonimmersive, respectively. The total number of subjects in all conditions whose data were included for analysis was 68. Gender and video game experience of participants were taken into account to have uniform distribution across the visual feedback as well as the immersiveness conditions.

4.1.2 Apparatus

As in the pilot study, the joint orientation information to animate the avatar and the sphere, respectively, was recovered from the data streamed from the Microsoft Kinect to provide feedback to the participant for each condition. The virtual environment was rendered using WorldViz Vizard and FAAST [Suma et al. 2011] was used to read and process Kinect data feed to get the user's joint orientations. In the self-avatar condition, participants saw a virtual arm closely following the movements of their real arm. For those in the sphere condition, a textured sphere instead of the white sphere used in the pilot study was rendered instead of the avatar at the position where the hand of the avatar would have been rendered. Twenty-three objects were presented to the participants: one for the practice trials, and 22 additional objects for the experimental test trials. The practice object was in the shape of a jack with 6 bars with different colored ends. The objects used in the test trials were a subset of the anatomical objects used in Ziemek et al. [2012] created using digital embryos [Brady and Kersten 2003] (see Appendix A). The hand of the avatar was stopped from colliding with the object by not letting it come closer than 0.2m from the center of the object. The display used was a Samsung 60" Class (59.9" Diag.) Plasma 8000 Series Smart TV (see Figure 4.1). The display was driven with a 1080p video feed at 60Hz refresh rate. In the stereo condition, we used top-bottom stereo mode with each feed for each eye at half of the complete resolution vertically. The distance of the participant from the display was 1.45m and the geometric field of view of the participant was matched with the rendered display field of view. The wireless IC3 was attached to a wrist brace which was then fastened over the right hand of the participant such that the orientation sensor

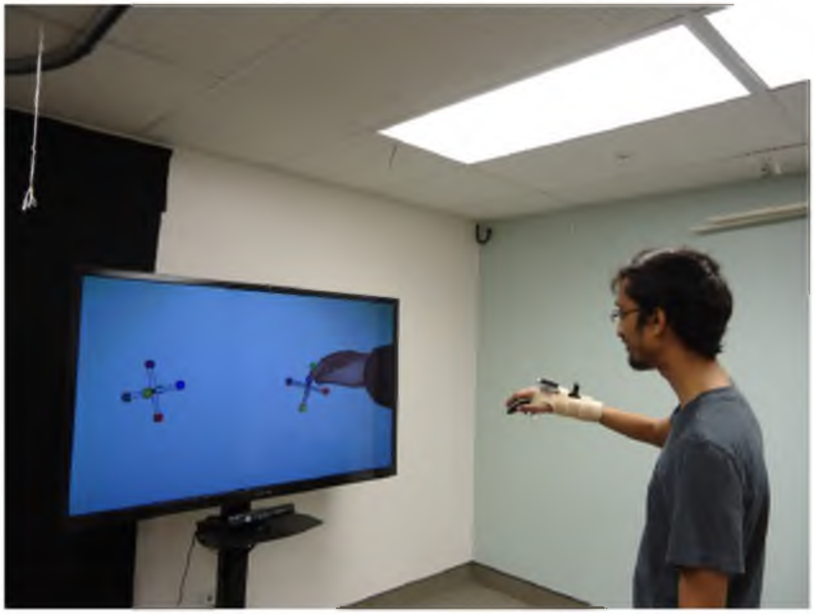


Figure 4.1. Setup for Experiment 2. The subject position is 1.45m away from Samsung 60" Class (59.9" Diag.) Plasma 8000 Series Smart TV

was situated over the back of the participant's right hand (see Figure 4.2). The wrist brace also restricted the motion of the wrist to rotation about a single axis along the direction of the wrist.

4.1.3 Design

This experiment used a mixed factorial design with two between-subjects factors and one within-subjects factor. The two between-subjects factors were visual feedback condition and visual immersiveness. The difficulty level of the 22 object matching trials that were common for all the subjects was the within-subjects factor (see section 4.2). In the high visual immersion condition, the trials are conducted with the stereoscopic condition enabled and room lighting switched off. The fusion distance of the stereo was set to the distance between the user viewpoint in the virtual world and the object in the virtual world to minimize the disparity when viewing the object. Also, the geometric and the display field of views were matched to make the simulation as realistic as possible. In the low visual immersion condition, the stereo mode of the display was disabled and the room lighting was not switched off in order to simulate lower level of visual immersiveness. For each of the visual immersiveness conditions, we used a between-subjects manipulation for testing for effects of self-avatars on the performance of users in the object manipulation task. The two conditions for visual feedback are the self-avatar present condition and the no

Figure 4.2. Close up of the wrist brace setup with the mounted Intersense InertiaCube3

self-avatar present or the sphere condition. The only difference between the two visual feedback conditions, as in the first experiment, was the visual feedback provided to the user. In order to test for the effect of visual immersiveness due to stereo on the strength of the effect of self-avatar, we again employed a between-subjects manipulation. To recapitulate, the four between-subject conditions, one of which was assigned to each subject, were avatar present - visually immersive, avatar present - nonvisually immersive, avatar not present - visuallyimmersive, and avatar not present - nonvisually immersive, respectively.

While selecting a set of 22 axes of rotation for each of the trials, it was ensured that there would be an even spread of axes whose orientations were close to the orientations of the coordinate axes and oblique axis that were oriented away from the coordinate axes. This was done with the assumption that axes at different orientations would make their respective trials have different difficulty levels. In this study, we define an axis as oblique-axis if the dot product of a unit vector along the axis and the unit vector along the closest coordinate axis is less than 0.65 and as near-coordinates axis if the dot product of a unit vector along the axis and the unit vector along the closest coordinate axis is greater than 0.85. In order to select the axes set (consisting of 22 axes of rotation, one for each trial), 22 axes were chosen by first selecting 22 points uniformly on a unit sphere. Then, the axis of rotation were the lines passing through the center and the selected points. Figure 4.3 shows sampling of points from a unit sphere. The amount of rotation needed for matching the orientations of the objects by rotating the manipulatable object along the optimal axes was selected from a group of uniformly spaced values between 21.6 degrees and 174.6 degrees with an

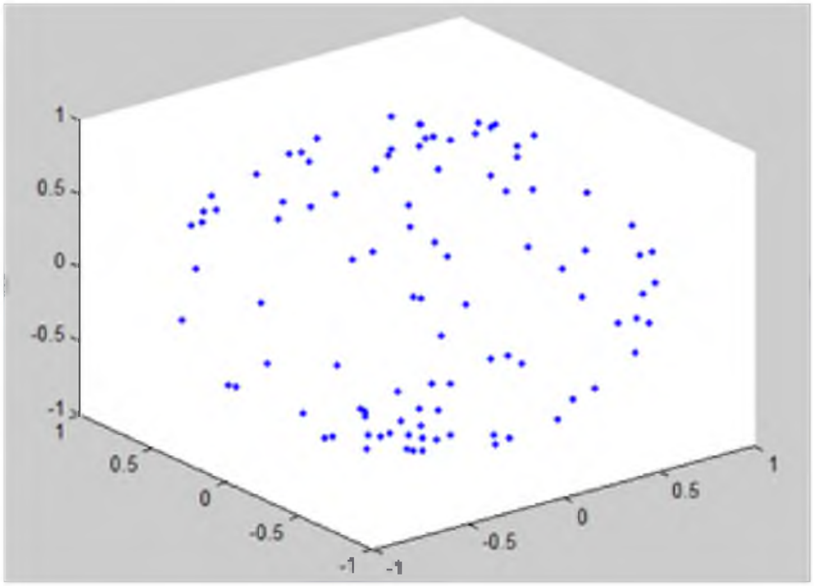


Figure 4.3. Sampling uniformally from a unit sphere for generating axis of rotation for each of the trials.

interval of 7.2 degrees between the values. Next, each of the 22 values for the amount of rotation was randomly assigned to a unique axis from the axes set. Now the two sets of axes and rotation amount were in the form of a single set of axis and rotation amount tuples. We then checked to see if this final set had at least three easy and hard trials by checking if it had at least three tuples each with oblique axis/large rotation and near-coordinates axis/small rotation pairs. If this was not the case, we tried again by starting with selecting a new axes set, assigning them each a different amount of rotation from the set we used earlier, and then testing if the new set of tuples satisfied our requirement of having hard and easy trials until this requirement was satisfied. Once the requirement was satisfied, the set of 22 tuples was stored and used for the whole study. Figure 4.4 shows the histogram of the dot product values of the final selected set of axes and the closest coordinate axis to each of those axes. The bins on the left side along the horizontal axis contain the oblique axes with low dot product values with the nearest coordinate axis while the bins on the right side along the horizontal axis contain the near-coordinates axis with high dot product values with the nearest coordinate axis. In the study, the orientations of the two objects was accepted as matched when their orientation differs by less than 15 degrees about the axis requiring the shortest rotation.

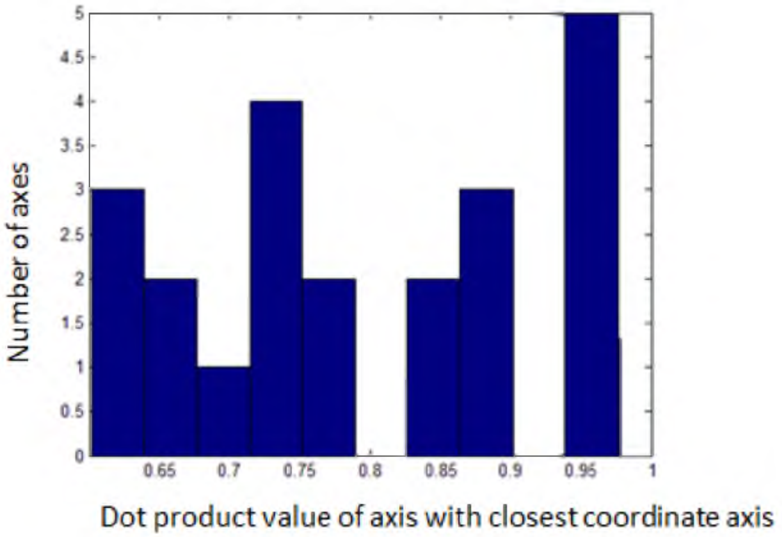


Figure 4.4. An evenly spread out histogram of the dot product of values of the final selected set of axes and the closest coordinate axis to each of those axes ensure that there are axes that are both near the coordinate axes as well as axes that are oblique. The bins on the left side along the horizontal axis contain the oblique axes with low dot product values with the nearest coordinate axis while the bins on the right side along the horizontal axis contain the near-coordinates axis with a high dot product values with the nearest coordinate axis.

4.1.4 Procedure

The participants in this experiment were tested for their spatial ability and then asked to perform a set of object manipulation tasks.

4.1.4.1 Questionnaire and Tests

After giving consent to participate, the participants were asked to fill in a questionnaire (see Appendix B) about their gaming experience and experience with gesture-based user interfaces. Their gender and age were also recorded in the questionnaire. Next, each participant was asked to complete two tests, first the paper folding test [Ekstrom and French 1976] to measure the exocentric spatial abilities and then a spatial orientation test [Hegarty and Waller 2004] to measure their egocentric spatial abilities. This was done in order to have data to analyze for interaction of spatial abilities of the participants as a factor for performance in the object manipulation task that came next. After completing the tests, each participant was led to a new room with the experiment setup and was given instructions for manipulating the object and completing the task.

4.1.4.2 User Interface

In order to rotate the object on the right side of the display to match its orientation with the orientation of the object on the left side, the user could chose between two different actions. The first action, also called the swipe motion, was similar to the swipe mode used in the first experiment. The second is called the twist and differed from the twist mode used in the first experiment with respect to the axis of rotation along which the rotation of the object occurred. Instead of the rotation happening about the axis along the orientation of the wrist, the rotation now occurred along the axis along the line joining the hand and the object. Also, the degrees of freedom of the wrist were now restricted using a wrist brace to just rotation about one axis which is a roll about the axis line joining the elbow joint and the wrist joint.

Further manipulation was only possible when the hand of the avatar is close to the object and when it is pointing towards the object. The orientation constraint was an additional constraint over the contact constraint in the pilot experiment. Also different from the interface in the first experiment is switching between the two modes of manipulation. While in the first experiment, the user needed to explicitly signal the desired mode of operation using the left or right mouse button, in the second interface, the mode switched automatically by detecting the kind of gesture made by the user. The decision of selecting a mode was done using the speed and type of hand motion gesture by the interface itself. However, the user still needed to signal gripping the object by pressing any one of the mouse buttons.

4.1.4.3 Task Description

The main experiment itself was an orientation matching task. Organic bone-shaped objects similar to those used in the first experiment were used for the orientation matching tasks. Additionally, the objects were smooth-shaded and pivot was fixed at the center of the bounding sphere of the objects. The area on the display was split into two regions by a vertical separator and an object at two different orientations was shown in the center of each of the regions. The subject needed to rotate the object on the right to match its orientation with the orientation of the object on the left. Each subject had to match the orientation of 22 distinct object pairs in 22 consecutive trials with one matching task in each trial. Before starting with the trials, each subject was given five practice trials with a practice object (see Appendix A). After the practice trials, 22 objects were presented in pairs at different orientations in successive trials for the subject to match the orientation. In each of the trials, the response time of the subject to complete the orientation matching task was recorded. When the orientation of the object was within 15 degrees of the orientation of the target object, a "match detected" prompt appeared on the screen. During experimental trials, if 60 seconds elapsed between the onset of the trial before a match was detected, a prompt indicated that the allotted time had expired. When the alloted time for a trial expired, a new trial was automatically loaded after recording the incomplete status of the current trial.

4.2 Results

Before analyzing the performance, the time taken by participants in each trial was normalized by dividing it by the average of the time taken by all the participants for a particular trial. This was done so that the effects of the shorter trials were not completely marginalized by the trials that, on an average, took longer to be completed. On analyzing the average normalized time for each participant to complete all trials, we observed that the visual feedback condition did not significantly influence the overall performance. However, on taking a closer look considering hard and easy trials separately, we found an effect of the presence of the avatar, qualified by the difficulty of the trial. It was observed that in the harder trials, participants in the avatar visual feedback condition performed significantly better than participants in the sphere visual feedback condition while in the easy trials, performance was similar for participants in both the visual feedback conditions (see Figure 4.5). We also found that the exocentric spatial ability tested by the paper folding test and the visual immersion condition of the participant also also had an influence on the performance in the orientation matching task. Participants were overall faster in the high visually immersive condition versus the low visually immersive condition. Performance was better across visual feedback conditions for participants with higher spatial ability test scores (see Figures 4.6 and 4.7).

We performed a 2x2x2 univariate ANOVA for 2(visual feedback) x 2(visual immersion) x 2(gender) conditions with the total normalized time taken to complete all trials as the dependent variable and first-person gaming score and both spatial ability test scores as covariates. We did not find a significant effect of either visual feedback $F(1,57) = 1.538$, $p < .220$, $\eta_{\text{p}}^2 = 0.026$, an interaction between visual feedback x visual immersion, $F(1,57) < 0.253$, $p < .617$, $\eta_{\text{p}}^2 < 0.004$, or an interaction between visual feedback x gender $F(1,57) < 0.495$, $p < .484$, $\eta_{\text{p}}^2 < 0.009$. We found that the paper folding test scores had a significant effect on the rotation time $F(1,57) < 5.372$, $p < .024$, $\eta_{\text{p}}^2 < 0.086$, as did visual immersion, $F(1,57) < 4.325$, $p < .042$, $\eta_{\text{p}}^2 < 0.071$

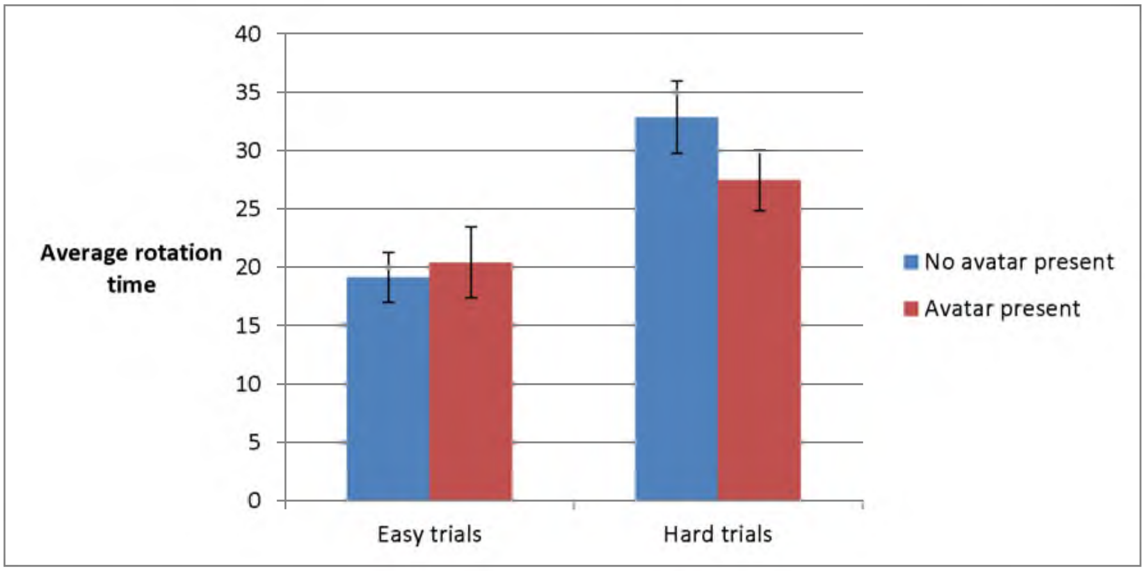


Figure 4.5. Average rotation time (unnormalized) with 95 percent confidence interval for easy and hard trials by visual feedback condition.

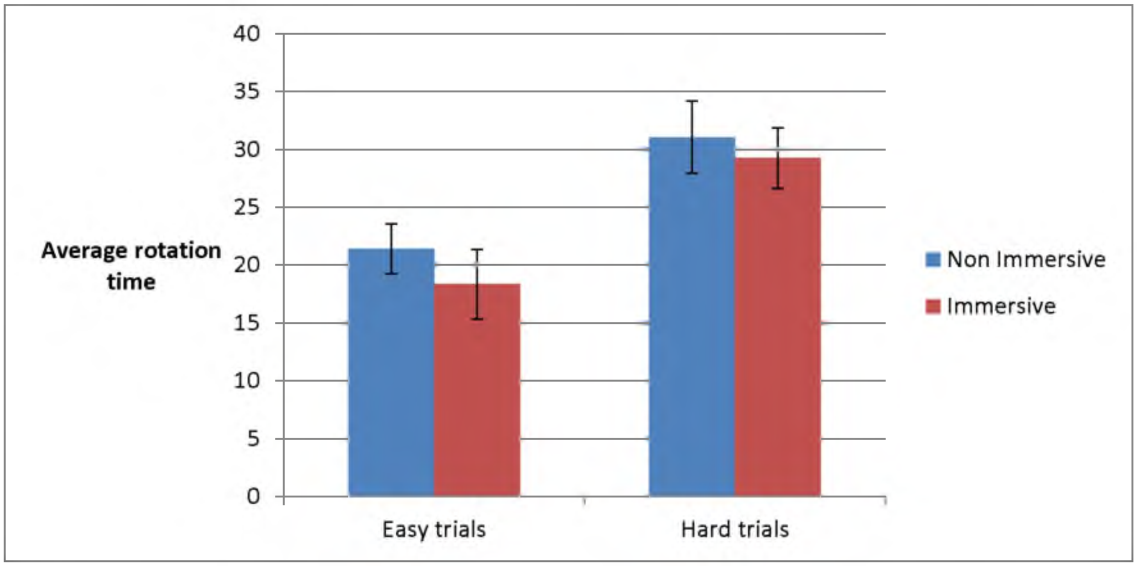


Figure 4.6. Average rotation time (unnormalized) with 95 percent confidence interval for easy and hard trials by immersiveness condition.

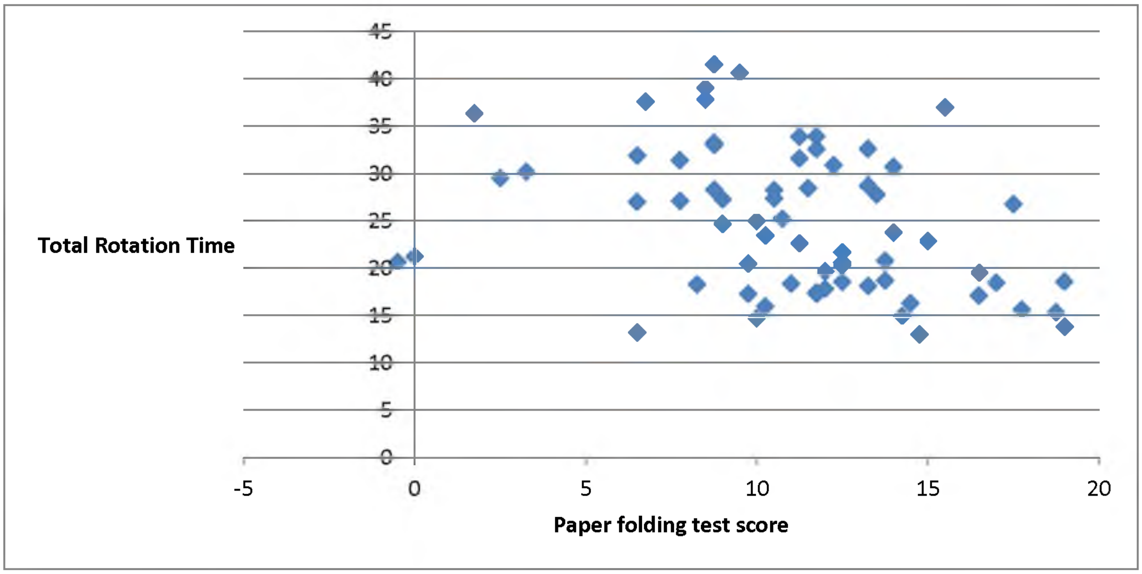


Figure 4.7. Average rotation time (unnormalized) versus the paper folding test score. The paper folding test measured the exocentric spatial abilites of the participants and was found to have a overall significant effect on average rotation time.

In order see if there was an interaction between visual feedback and task difficulty, we classified the tasks as hard and easy by looking at the average time taken for the tasks in a particular trial over all subjects before normalization as the indicator of difficulty. The trials with larger than median average completion time were classified as hard and the trials with smaller than median completion time were classified as easy. We then calculated an overall average rotation time for the 11 easy trials, and 11 difficult trials, based on a median split to do the analysis. We found that hard trials were completed significantly faster in the avatar feedback condition while in the case of the easy trials, there was no significant difference between performance in avatar and sphere feedback conditions. We also found an overall effect of visual immersion in addition to the effect of the paper folding test score. We ran a 2 (difficulty) x 2 (visual feedback) x 2 (visual immersion) x 2 (gender) repeated measures ANOVA with difficulty (2 levels) as within-subject and avatar, visual immersion, and gender as between subject variables, and gaming, spatial orientation test score, and paper folding test score as covariates. We found a strong avatar x difficulty interaction, $F(1,59) = 9.10$, $p < .004$, $\eta_p^2 < .138$. The overall effect of visual immersion, $F(1,57) = 4.32$, $p < .04$, $\eta_p^2 < .071$, and paper folding test scores, $F(1,57) = 5.37$, $p < .024$, $\eta_p^2 < .086$, were also significant .

Lastly, we analyzed the rotation times to see if there was an interaction between visual feedback condition and individual difference factors (gaming experience, spatial abilities). In this case, we did not find any interaction between visual feedback and any of the individual difference factors. Before doing this analysis, we first did a median split on all three individual difference factors to restrict them to just high and low values. This was needed to analyze the individual differences as a between-subjects variable in repeated measures ANOVA. We performed three separate repeated measures ANOVAs, 2 (feedback) x 2(gaming), 2 (feedback) x 2(paper folding score), and 2 (feedback) x 2(spatial orientation score), each with average time for easy and hard trials as a within-subjects variable. There was no significant effect observed here in any of the cases.

4.3 Discussion

We used the final average time taken for each trial as a measure of the difficulty for that trial instead of using trial difficulty computed according to a premeditated metric. This was done since there are a various possible factors such as object shape, trial number of the object, and axis orientation to consider that could interact with each other in unpredictable ways to make the rotation in a particular trial harder or easier than the others. For instance, a shorter rotation about an axis lying on the plane of the display could be harder than a larger rotation about an axis perpendicular to the surface of the display or an elongated object to be rotated about an oblique axis might be easier to match than a near symmetrical object about an axis perpendicular to the display.

The results of Experiment 2 did not indicate a significant overall effect of the presence of a visual feedback in the form of a self-avatar which is similar to what we observed in the pilot study. Also, there was no significant interaction of visual feedback with the level of visual immersiveness. However, we found a significant interaction between the avatar feedback condition and task difficulty. The performance of subjects in matching the object in the easy trials was similar across visual feedback conditions, but the time taken to match the objects was significantly less for harder trials in the condition with the avatar feedback compared to the condition with sphere feedback (see Figure 4.5). We also found an overall effect of visual immersion and paper folding test score with participants with high paper folding test score and those in the high visual immersiveness condition performing better. This is not surprising considering that the stereo display in the high visual immersiveness condition could help the participant choose a more appropriate axis of rotation to complete the task. Also, the object orientation matching task is inherently exocentric so the paper folding test score, which is designed to test exocentic spatial abilities, can be expected to have an effect on the performance in this task. However, there was no significant interaction between visual feedback and either spatial abilities or even gender as in the case of the pilot study.

CHAPTER 5

GENERAL DISCUSSION

With the advent of new low-cost gesture-based interfaces, such as the Microsoft Kinect, new modalities have become possible for general users to interact easily with 3D objects in desktop-based virtual environments. One such modality is interacting with objects in a virtual environment via an animated self-avatar. In this research, a set of two experiments was conducted to study the effect of the presence of a self-avatar in a virtual environment on the performance of an object rotation task in that virtual environment. Specifically, the questions that the experiments tried to answer were if an effect would be noticeable on a flat-panel display and whether the strength of the effect, if present, would be affected by factors such as visual immersion in the virtual environment, individual differences among the subjects, or the difficulty of the task.

5.1 Final Conclusion

A key finding of this research was the presence of an interaction effect between the visual feedback condition and task difficulty. It was observed in Experiment 2 that subjects in the avatar visual feedback condition performed better in the object manipulation task in trials with harder rotations while there was no significant difference in the performance of subjects in both visual feedback conditions in trials with easier rotations. This is an important result since it implies that interacting in virtual environment through an animated self-avatar could lead to significant performance improvements in more challenging tasks. In other words this would mean that the presence of an animated self-avatar could make a challenging task easier for the user to perform. Such an interface would find application in almost all domains where the user would need to interact with objects in a virtual environment.

In the first experiment, we saw the feasibility of two different visual feedbacks conditions during interaction, portraying feedback to the user's actions as either a self-avatar or a sphere. It was seen that the time taken to rotate the objects to match the target and the number of successful trials within the allotted time did not differ across display conditions when averaging across all users. However, gender differences that were also related to video gaming experience did influence performance in the sphere display condition which provided less of an egocentric frame of reference and was less anthropomorphic. Thus, it can be inferred that, when designing interfaces for object manipulation, individual differences in users' spatial abilities and experience should be taken into account in order to determine the interface that is most advantageous for the highest number of users. The pilot study led us to the idea that there might be an interaction effect due to the the visual feedback in the form of avatar or sphere interacting with another factor, possibly an individual difference factor among the subjects. The follow-up study or Experiment 2 was designed with the goal of exploring further the effect of such interactions.

In the second experiment, we recorded individual difference factors that were likely to have an effect on performance in the object rotation task. We tested subjects for their egocentric and exocentric spatial abilities in order to control for those factors. We also recorded first-person and third-person gaming experience along with their experience using gesture-based interfaces (see Appendix B). Additionally, we varied level of visual immersion as a between-subjects factor. Finally, we also classified the trials as easy and hard (see section 4.2) in order to analyze the effect of task difficulty as a within-subjects factor to see if task difficulty interacted with the visual feedback condition of the subjects.

As mentioned earlier, a significant interaction effect was observed between visual feedback and difficulty over all subjects in Experiment 2. Those in the avatar feedback condition did significantly better only on the harder task but performed similar to those in the sphere condition in the easy tasks (see Figure 4.5). This can expected to be the case if the presence of the avatar helped the subjects to keep track of the orientation of the object over more complex rotations. In the easy trials, this additional benefit may have been superfluous. We also found an overall effect of visual immersiveness and exocentric spatial ability test. This is easy to explain since it is likely that it would have been easy to select an axis of rotation given the stereo afforded in the high visual immersiveness condition. Also, the scores in the paper folding test indicate the exocentric spatial ability of the participant that was useful for the object rotation task.

As in the case of the pilot, we did not find an overall significant effect of visual feedback, although the mean rotation times for the avatar feedback condition were lower than that of the sphere condition. However, unlike in the pilot study, an effect of interaction of individual differences with visual feedback was absent in Experiment 2. One reason for this might be that the textured sphere used in the feedback for the sphere condition of the second experiment acted as a more efficient orientation indicator as compared to the uniform white sphere used in the pilot study. It should be noted that the interfaces in Experiment 1 and Experiment 2 had more significant differences that might have helped subjects with low gaming experience perform better (see 4.1.4.2). Also it is possible that the interaction effect became weaker as a result of the increased sample size in the second experiment representing a more general population. Lastly, we did not find a significant effect of interaction between visual immersion and visual feedback conditions. This would mean that it was visual feedback as an auxiliary orientation indicator rather than the general naturalness of the interface that was the dominant influence in the interface in Experiment 2. It is also possible that the interface needed to more closely match the real-world object manipulation for a significant interaction of visual immersion with visual feedback or that the level of visual immersion provided by a large screen stereo flat panel was not sufficient for an interaction with the visual feedback.

To conclude, although we did not observe a significant overall effect due to the presence of an avatar in performing an object orientation matching task, we did observe a significant interaction of the avatar feedback condition with other factors. The avatar feedback condition interacted with subjects' individual differences in the pilot study. In the second Experiment, we found a significant interaction between avatar feedback condition and task difficulty over all the subjects. The interaction of individual differences was marginalized in Experiment 2, lending to the possibility that this is a weak effect if present. As mentioned before, it is important to note that there were some differences in the interface between the two experiments and also a difference in the subject pool size.

5.2 Future Work

To the best of our knowledge, this research is the first study investigating the effect of a Kinect-driven self-avatar in an object manipulation task with user testing and it is important for other independent studies to reproduce the results for them to be validated further. Although the avatar-based interface was made as close as possible to interaction in the real world with available resources, much improvement is still possible. One useful enhancement would be a higher fidelity motion capture system with less noise and lag. Also tracking and animating the avatar's fingers may further aid the user to manipulate objects in a virtual environment. It would be interesting to see if more realistic motions of the hand provided by a higher fidelity system together with animation of the fingers of the avatar would help the user become more efficient than with the avatar feedback used in this research. Also an important question that remains unanswered in this research is the impact of the realism of the avatar on the effects of the presence on the animated self-avatar. It would be interesting to find out if the effects observed in the experiments of this study would still be observed if the avatar's hand were not rendered as a polygonal model but more crudely instead as a sphere and five cylinders signifying the palm and five fingers or even as a wire-frame representation. Lastly, it is important to note that upgrades in the hardware needed for improving fidelity of the interface come at an additional cost. However, it is likely that such hardware will become available at a much lower cost in future, improving odds of such an interface becoming popular in applications in domains such as gaming, architecture, simulation, and education.

APPENDIX A

PRACTICE AND TRIAL OBJECT FIGURES

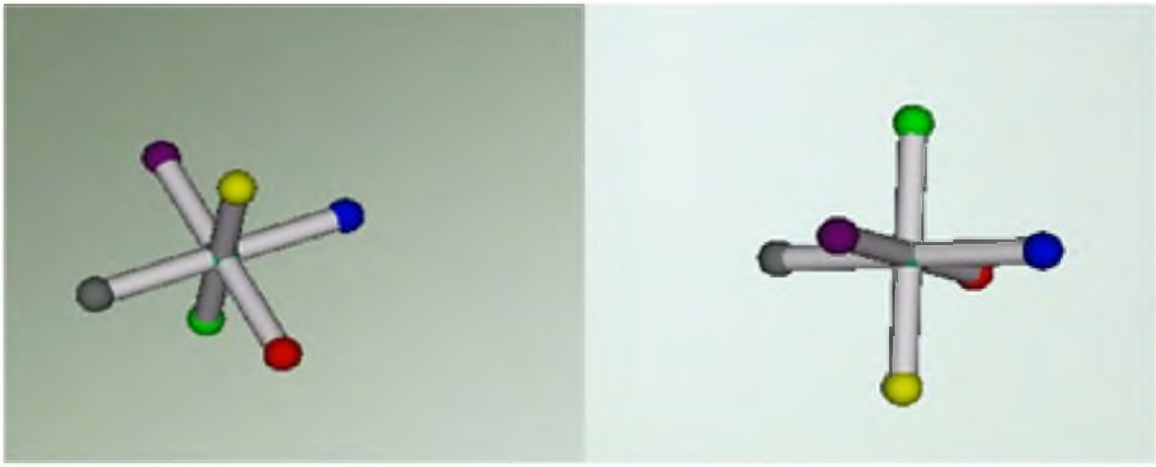


Figure A.1. Practice object used in Experiment 1 and 2

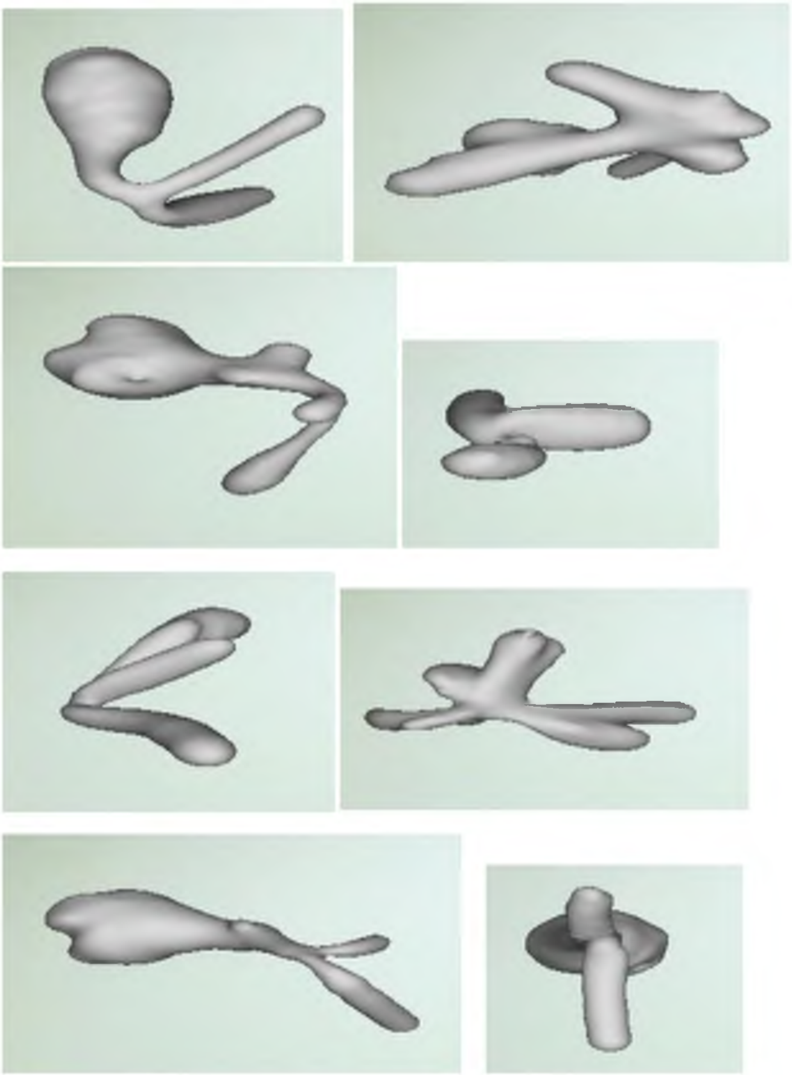


Figure A.2. Objects used in trial 1

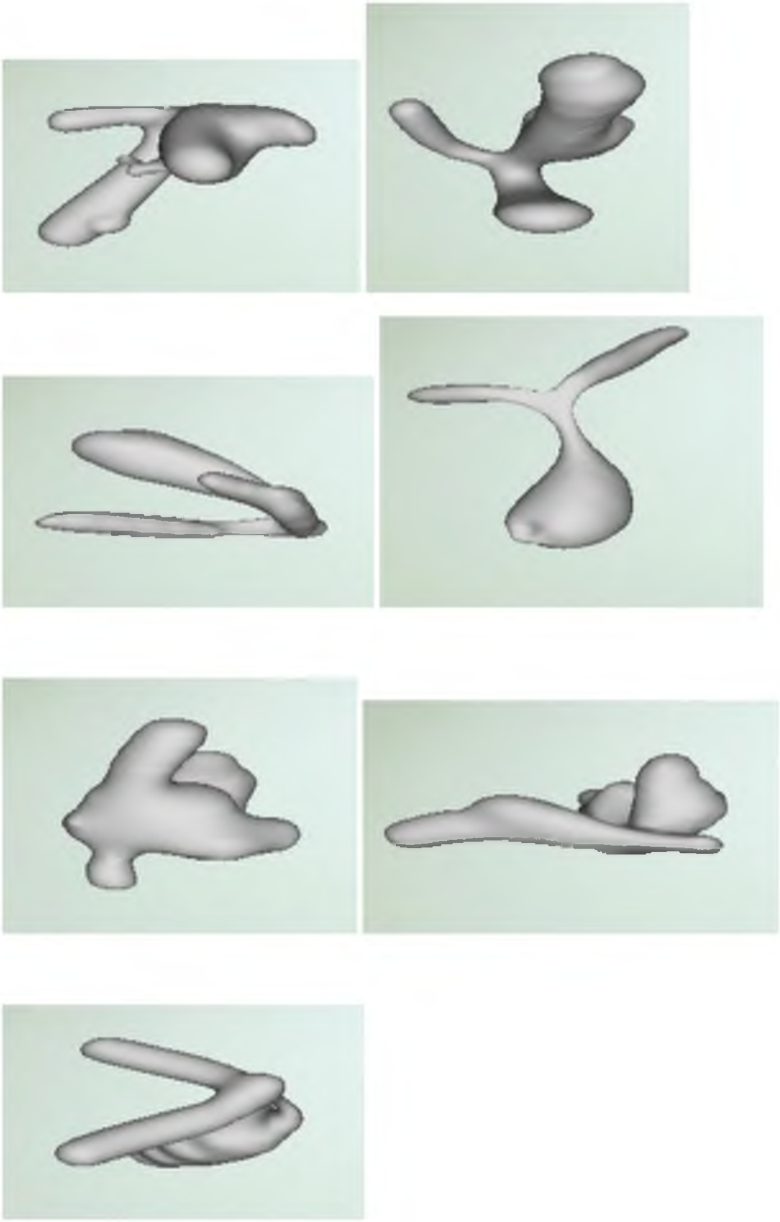
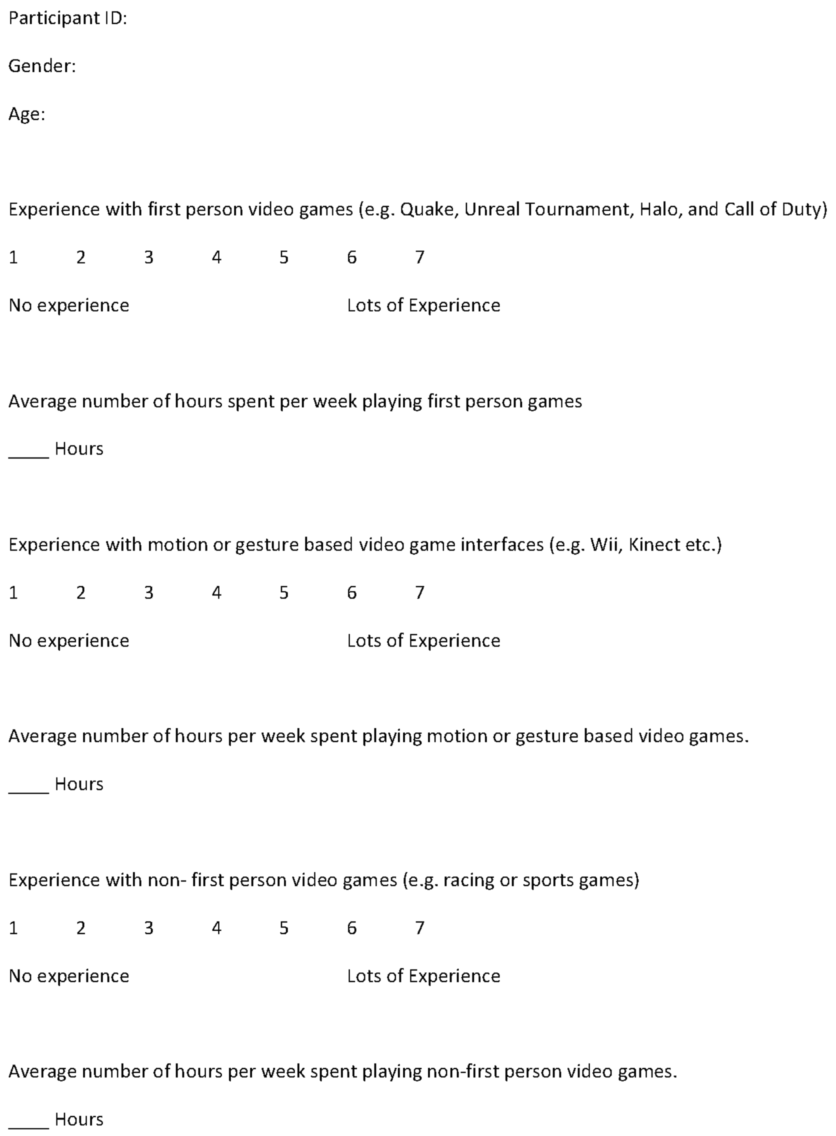


Figure A.3. Objects used in trial 2

APPENDIX B

QUESTIONAIRE GIVEN TO PARTICIPANTS



Participant ID:

Gender:

Age:

Experience with first person video games (e.g. Quake, Unreal Tournament, Halo, and Call of Duty)

1 2 3 4 5 6 7

No experience Lots of Experience

Average number of hours spent per week playing first person games

____ Hours

Experience with motion or gesture based video game interfaces (e.g. Wii, Kinect etc.)

1 2 3 4 5 6 7

No experience Lots of Experience

Average number of hours per week spent playing motion or gesture based video games.

____ Hours

Experience with non- first person video games (e.g. racing or sports games)

1 2 3 4 5 6 7

No experience Lots of Experience

Average number of hours per week spent playing non-first person video games.

____ Hours

Figure B.1. Questionaire given to participants for collecting age, gender, and gaming experience data

REFERENCES

[Balslev et al., 2007] Balslev, D., Cole, J., and Miall, R. C. (2007). Proprioception contributes to the sense of agency during visual observation of hand movements: evidence from temporal judgments of action. *Journal of Cognitive Neuroscience*, 19(9):1535–1541.

[Ban et al., 2012] Ban, Y., Narumi, T., Tanikawa, T., and Hirose, M. (2012). Magic pot: Interactive metamorphosis of the perceived shape. In *ACM SIGGRAPH 2012 Emerging Technologies*, SIGGRAPH '12, pages 14:1–14:1, New York, NY, USA. ACM.

[Bowman and Hodges, 1997] Bowman, D. A. and Hodges, L. F. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D graphics*, I3D '97, pages 35–ff., New York, NY, USA. ACM.

[Brady and Kersten, 2003] Brady, M. J. and Kersten, D. (2003). Bootstrapped learning of novel objects. *Journal of Vision*, 3(6):413–422.

[Ekstrom and French, 1976] Ekstrom, R. B. and French, J. W. (1976). *Manual for Kit of Factor-referenced Cognitive Tests.*

[Hegarty and Waller, 2004] Hegarty, M. and Waller, D. (2004). A dissociation between mental rotation and perspective-taking apatial abilities. *Intelligence*, 32(2):175–191.

[Hegarty and Waller, 2005] Hegarty, M. and Waller, D. (2005). Individual differences in spatial abilities. In Shah, P. and Miyake, A., editors, *Handbook of Higher-level Visuospatial Thinking*, pages 121–169. Cambridge University Press, New York.

[Hinckley, 2002] Hinckley, K. (2002). Input technologies and techniques. In Jacko, J. A. and Sears, A., editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pages 151–168. Lawrence Erlbaum Associates.

[Hinckley et al., 2004] Hinckley, K., Jacob, R. J. K., and Ware, C. (2004). Input/output devices and interaction techniques. In Tucker, A. B., editor, *Computer Science Handbook*, chapter 20. CRC Press, 2nd edition.

[Hinckley et al., 1994] Hinckley, K., Pausch, R., Goble, J. C., and Kassell, N. F. (1994). Passive real-world interface props for neurosurgical visualization. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems: Celebrating Interdependence*, pages 452–458. ACM.

[Hinckley et al., 1998] Hinckley, K., Pausch, R., Proffitt, D., and Kassell, N. F. (1998). Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction*, 5(3):260–302.

[Hinckley et al., 1997] Hinckley, K., Pausch, R., Proffitt, D., Patten, J., and kassell, N. (1997). Cooperative bimanual action. In *ACM CHI*, pages 27–35.

[Hodgins et al., 1998] Hodgins, J. K., O'Brien, J. F., and Tumblin, J. (1998). Perception of human motion with different geometric models. *Visualization and Computer Graphics, IEEE Transactions on*, 4(4):307–316.

[Izadi et al., 2011] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. (2011). Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 559–568, New York, NY, USA. ACM.

[Jacob et al., 1994] Jacob, R. J. K., Sibert, L. E., McFarlane, D. C., and Mullen, J. M. P. (1994). Integrality and separability of input devices. *ACM Trans. on Comput-Hum Interaction*, 1(1):3–26.

[Jurnet et al., 2005] Jurnet, I. A., C., C. B., and Maldonado, J. (2005). Individual differences in the sense of presence. In *Proceedings of Presence 2005: The 8th International Workshop on Presence*, pages 133–142.

[Kadri et al., 2007] Kadri, A., Lcuyer, A., and Burkhardt, J.-M. (2007). The visual appearance of user's avatar can influence the manipulation of both real devices and virtual objects. In *3DUI*, page 11. IEEE.

[Khan et al., 2008] Khan, A., Mordatch, I., Fitzmaurice, G., Matejka, J., and Kurtenbach, G. (2008). Viewcube: A 3D orientation indicator and controller. In *ACM Symp. on Interactive 3D Graphics*, pages 17–25.

[Linn and Petersen, 1985] Linn, M. C. and Petersen, A. C. (1985). Emergence of characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56(6):1479–1498.

[Lok et al., 2003] Lok, B., Naik, S., Whitton, M., and Brooks, J. F. P. (2003). Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(6):615–628.

[McDonnell et al., 2012] McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)*, 31(4):91.

[Mohler et al., 2010] Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., and Buelthoff, H. H. (2010). The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. In *Presence: Teleoperators and Virtual Environments*, pages 230–242.

[Mori, 1970] Mori, M. (1970). The uncanny valley. *Energy*, 7(4):33–35.

[Pani et al., 1995] Pani, J. R., William, C. T., and Shippey, G. T. (1995). Determinants of the perception of rotational motion: Orientation of the motion to the object and to the environment. *JOURNAL OF EXPERIMENTAL PSYCHOLOGY HUMAN PERCEPTION AND PERFORMANCE*, 21:1441–1456.

[Phillips et al., 2010] Phillips, L., Ries, B., Kaeding, M., and Interrante, V. (2010). Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In *Proceedings of the 2010 IEEE Virtual Reality Conference*, VR '10, pages 115–1148, Washington, DC, USA. IEEE Computer Society.

[Poupyrev et al., 1998] Poupyrev, I., Ichikawa, T., Weghorst, S., and Billinghurst, M. (1998). Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In *Computer Graphics Forum*, volume 17, pages 41–52. Wiley Online Library.

[Raj et al., 2012] Raj, M., Creem-Regehr, S. H., Rand, K. M., Stefanucci, J. K., and Thompson, W. B. (2012). Kinect based 3d object manipulation on a desktop display. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '12, pages 99–102, New York, NY, USA. ACM.

[Sanchez-Vives et al., 2010] Sanchez-Vives, M. V., Spanlang, B., Frisoli, A., Bergamasco, M., and Slater, M. (2010). Virtual hand illusion induced by visuomotor correlations. *PLoS ONE*, 5:e10381.

[Slater et al., 1995] Slater, M., Usoh, M., and Steed, A. (1995). Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 2(3):201–219.

[Suma et al., 2011] Suma, E. A., Lange, B., Rizzo, A., Krum, D., and Bolas, M. (2011). FAAST: the flexible action and articulated skeleton toolkit. In *IEEE Virtual Reality*, pages 247–248.

[Tsakiris et al., 2006] Tsakiris, M., Prabhu, G., and Haggard, P. (2006). Having a body versus moving your body: How agency structures body-ownership. *Consciousness and Cognition*, 15(2):423–432.

[Wilson, 2002] Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9:625–636.

[Ziemek et al., 2012] Ziemek, T., Creem-Regehr, S., Thompson, W., and Whitaker, R. (2012). Evaluating the effectiveness of orientation indicators with an awareness of individual differences. *ACM Trans. Appl. Percept.*, 9(2):7:1–7:23.