Clicking versus grabbing: An investigation into the effect of interaction metaphor on fine-motor tasks in virtual reality

Geoff Musick* Clemson University
Carrie Russell† Clemson University
Austin Searight‡ Clemson University
Boston Ellmyer§ Clemson University

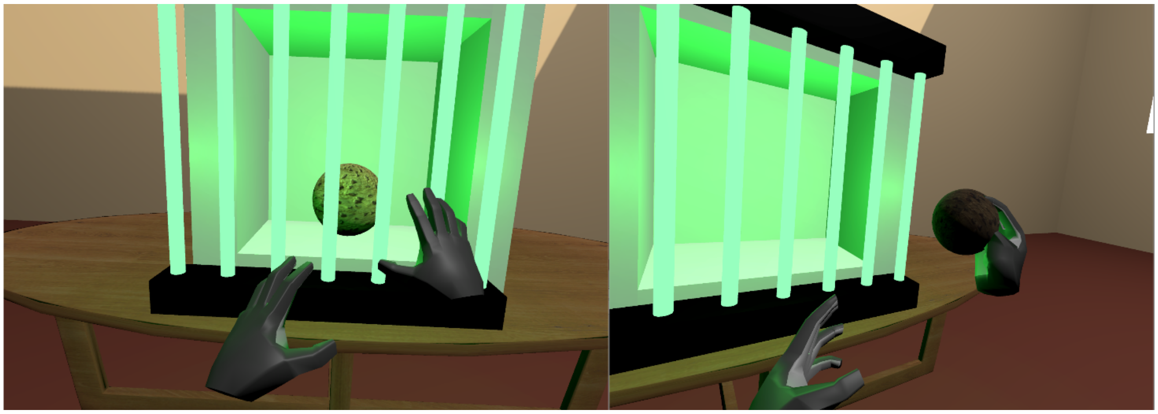

Figure 1: Leap Motion hand tracking condition - grabbing a sphere in the box

Abstract

Performing fine motor tasks in Virtual Reality has become increasingly common in recent years with the improvement in technology, processing power, displays, image quality, and improved tracking. One prominent metaphor for fine motor tasks are for the participant to hold a controller in each hand. A benefit of this metaphor is the precise and reliable outcome corresponding to a button press. A disadvantage of this metaphor is the limited gestures and perhaps decreased immersion this affords. Alternatively, hand tracking or gesture interaction metaphors which have a high level of immersion, have their own limitations. In this study, we investigated the effect of interaction metaphor on fine motor tasks. Comparing the interaction metaphor of the HTC Vive Controller to that of the Leap Motion hand gesture metaphor, we observed the success rate of participants attempting to grab objects in a near field box with a door that closed at various frequencies. Further, we compared the total time by condition, success rate by size of object, sense of presence, and workload for comparing the two conditions. Though the extent of our findings were limited due to the small sample size, the time it took for participants to complete the task was far greater for the gesture condition. Importantly, we address important recommendations for this study to be performed fully in the future.

Index Terms: Human-centered computing—Interaction Metaphor—Hand Tracking—;

1 Introduction

Near field fine motor tasks are important in Virtual Reality. From training scenarios to gaming applications, Virtual Reality users need to be able to successfully and consistently interact with objects that are within close proximity with little to no frustration. As hand tracking in Virtual Reality is gaining in popularity due to improved technology and the increased immersion that comes with it, it is important to investigate whether this assumed increased immersion is true. Further, we are interested as to whether gesturing as an interaction metaphor significantly decreases performance.For the purpose of our study we investigated how successfully participants were able to insert their hand and retrieve items from a box that has a sliding door that opens and closes at different frequencies. To measure this interaction metaphor difference, we compared the frequency of sliding doors against the number of items successfully retrieved per n attempts for each condition.

*e-mail: gmusick@g.clemson.edu
†e-mail: crusse4@g.clemson.edu
‡e-mail: asearig@g.clemson.edu
§e-mail: bellmye@g.clemson.edu

2 Experiment Simulation

We built our experiment simulation using the Unity 3D game engine. Participants are immersed in the virtual environment using the HTC-Vive Pro headset. Furthermore, participants interacted with the environment either using an HTC-Vive controller or with natural gestures using the Leap Motion hand tracking system (see Figure 2 and Figure 1 respectively). The simulation involves a box that is closed on 5 sides with a 6th side that is periodically exposed by a door that opens and closes at a given frequency. This box was centered in an environment that resembled a sports gymnasium. The point of the simulation is for participants to attempt to grab spheres that spawn in the box without tripping the alarm mechanism associated with the moving door. Throughout the simulation, 48 balls will spawn to represent 48 trials. These trials varied in ball size and door frequency. The details of these trials are described in more detail within the User Study section below. In addition to the experimental task of selecting objects, the scene has been gamified in order to increase user engagement. After successfully grabbing an object, participants had the opportunity to shoot a basket that was approximately 5 meters to the right of where participants were sitting. Furthermore, a scoreboard above the basket displayed the user's score. Points were deducted for triggering the alarm, while points were added for successful grabs and successful basket attempts (see Figure 3). Multi-modal feedback came in the form of both visual and audio for participants. For visual feedback, the shading of the spheres darkened when either the participants hands or controller were within the interactable range of the object. This shading reverted when leaving the interactable zone. Additionally, the color of the lasers associated with the door changed from green to red when an alarm was triggered. The color of the scoreboard cells also responded based on the level of success for the trial. Audio feedback involved different clicking sounds for both selection and deselection of the spheres. Additional audio cues came in response to spheres hitting the ground, goals being scored, and a sound for the alarm being triggered. Before the main scene that contained the 48 trials, participants were able to calibrate and learn their method of interaction with a training phase. In this special scene, participants were seated in front of a large number of spheres that they could pick up and drop to practice the interaction metaphor. Furthermore, participants were surrounded by baskets associated with the numbers 0-9. Using the spheres, participants would try to score baskets at the instruction of the experiment administrator to attempt to record their participant identification number. If participants were unable to record their number, the administrator could override their selection.

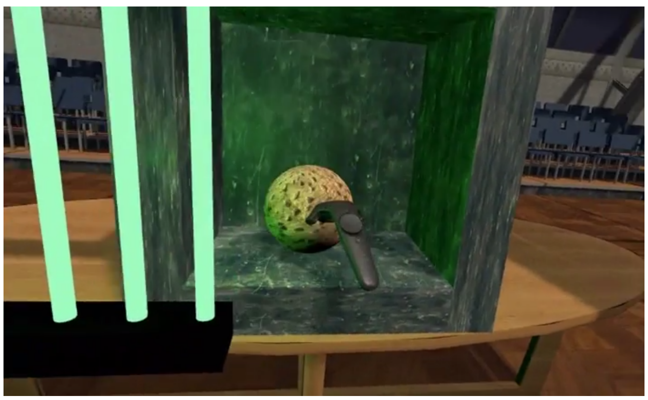

Figure 2: HTC Vive Controller condition - grabbing a sphere in the box

3 User Study

Upon arrival at the lab facility, participants were randomly assigned to one of two experimental conditions: 1) controller or 2) natural gesture. The participant was informed of the purpose of the study and provided a printed informational letter to read. After consenting to participate in the study, the participant sat at a computer workstation and was directed to the Clemson Qualtrics website, where they completed and submitted online surveys on demographics, experience with gaming technology, and experience with immersive virtual reality. Following completion of the online survey, the participant was given a five minute demonstration on how to use the virtual reality equipment and interact in the immersive virtual reality environment using the HTC Vive headset and either

- HTC Vive controllers or
- Leap Motion tracker and natural hand gestures.

Upon completion of the demonstration, the participant began the practice portion of the virtual reality experience.

Within the immersive virtual reality environment, the participant first encountered a training scene that was used to provide information about the experimental task. The participant went through a guided tutorial, where they practiced ‘grasping’ and ‘releasing’ an object that was on a virtual desk.

During the experimental portion of the study, the immersive virtual reality environment involved a gymnasium environment with a box at eye level of the participant in a seated position. The box had a sliding door that opened and closed at different rates, depending on the experimental trial. The participant proceeded to reach in and grab a virtual object from within the box. The size of the virtual object was either ’large’ or ’small’, depending on the experimental trial. The participant completed a total of 8 trials for each set, one trial for each of the 8 frequencies. After completing a set, the participant would move on to the next set. 6 sets were completed in total to allow for 3 trials for each object size at each frequency. This made a total of 48 trials. When the participant had completed all 48 trials, the experimenter would signal to the participant that the trials have ended. The participant would then remove the HTC VIVE headset and the experimental portion of the study was complete.

The participant then sat at a computer workstation and would go to the Clemson Qualtrics website, where they completed and submitted the igroup Presence Questionnaire (IPQ) and the NASA TLX questionnaire. Upon completion of the survey, the participant would be debriefed on the purpose of the study, potential impacts of research, and principal investigator contact information for questions or follow-up. The participant was then dismissed from the study session.

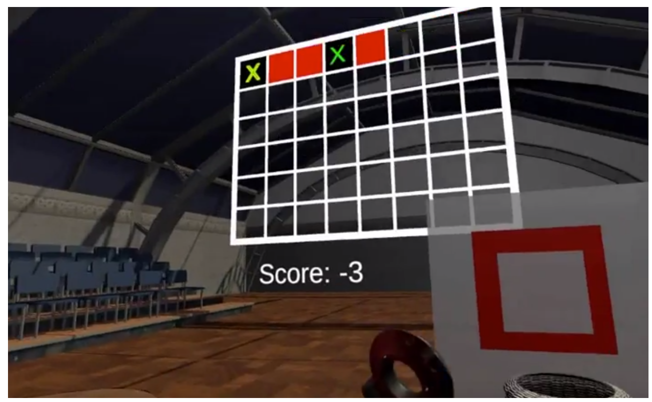

Figure 3: Scoreboard and gymnasium environment

4 Results

4.1 Interaction Metaphor

Several analyses were performed to investigate Research Question 1: Does the type of interaction metaphor affect performance in a VR fine motor, near-field task?

A two-tailed t-test was performed to analyze the effect of type of interaction metaphor (natural gesture, controller) on task performance. The result was not significant at α 0.05, as t(8) = 1.64, p = .1179. This indicates that there is no significant difference in success rates between the gesture (M = 0.78, SD = 0.21) and controller (M = 0.88, SD = 0.07) conditions.

A two-tailed t-test was also performed to analyze the effect of interaction metaphor (natural gesture, controller) on total time to complete. The result was statistically significant at α 0.05, as t(8) = -2.61, p = .03125. This indicates that the time needed to complete the task is significantly higher in the natural gesture condition (M = 773.82, SD = 3.04) than the controller condition (M = 393.49, SD = 47.42), as shown in Figure 4.

4.2 Target Object Size

Further analyses were performed to investigate Research Question 2: Does target object size affect performance in a VR fine motor, near-field task?

A two-tailed t-test was performed to analyze the effect of object size (small, large) on task performance. The result was not significant at α 0.05, as t(8) = -0.399, p = .6947. This indicates that there

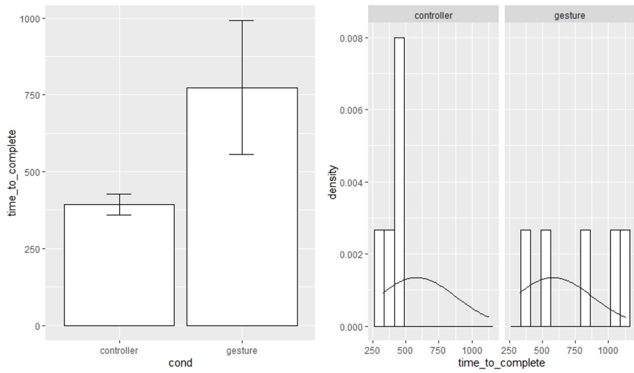

Figure 4: Total time to complete the experiment shown by condition

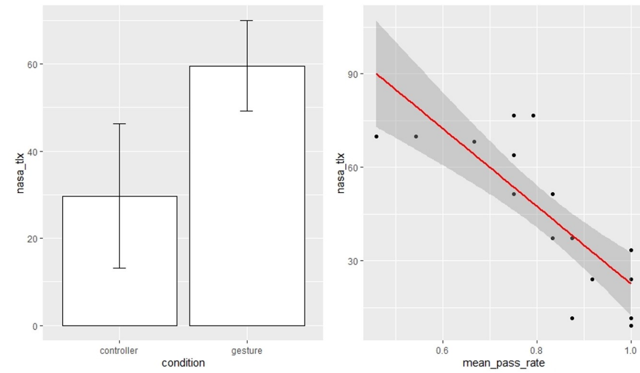

Figure 5: NASA TLX scores shown by condition.

is no significant difference in performance for either condition when interacting with small (M = 0.84, SD = 0.16) and large (M = 0.81, SD = 0.17) target object.

A multifactor analysis of variance (ANOVA) was performed to analyze the interaction effect of condition (natural gesture, controller) and object size (small, large) on task performance. The result was not significant at $\alpha = 0.05$, as $F(1, 16) = 0.164$, $p = .69$, as shown in Figure 8.

4.3 Frequency

Additional analyses were performed to investigate Research Question 3: How does door speed (frequency) affect performance. As shown in Figure 9, mean success rate declines as door speed increases for all conditions; the natural gesture condition has lower mean success rates than the controller condition for both target object sizes; and performance was different at increasing speeds for all conditions between the small and large target objects.

4.4 Workload

A two-tailed t-test was performed to analyze the effect of interaction metaphor (natural gesture, controller) on perceptions of workload. The result was statistically significant at α 0.05, as t(8) = -2.304, p = .05. This indicates that the perceptions of workload are significantly higher in the natural gesture condition (M = 59.53, SD = 15.42) than the controller condition (M = 29.67, SD = 24.54).

Considering the non-normality and the large number of same scores for the NASA TLX (workload) variable, a Kendall's tau was also performed to determine the correlation between workload and performance. The result was statistically significant at α 0.05, as $\hat{=}$ -0.69, p = 5.105e-05. Given these results, we can conclude that participant's overall perceptions of workload are negatively correlated with task performance. This means that as performance decreases, perceptions of workload increase as shown in Figure 5.

4.5 Immersive Tendencies

A two-tailed t-test was performed to analyze the difference in self-reported immersive tendencies between participants in the natural gesture and controller conditions. The result was statistically significant at α 0.05, as t(8) = -2.79, p = .01209. This indicates that participants in the natural gesture condition (M = 91.8, SD = 10.36) had significantly higher immersive tendencies than participants in the controller condition (M = 80, SD = 8.46). These differences can be seen in Figure 6.

4.6 Sense of Presence

A two-tailed t-test was performed to analyze the difference in self-reported sense of presence between participants in the natural gesture and controller conditions. The result was statistically significant at

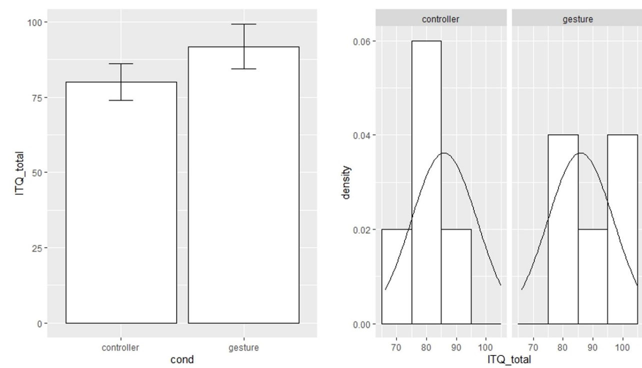

Figure 6: Immersive tendencies shown by condition.

α 0.05, as t(8) = -2.29, p = .0339. This indicates that participants in the natural gesture condition (M = 56, SD = 6.73) had significantly lower sense of presence in the virtual environment than participants in the controller condition (M = 61.4, SD = 3.73). These differences can be seen in Figure 7.

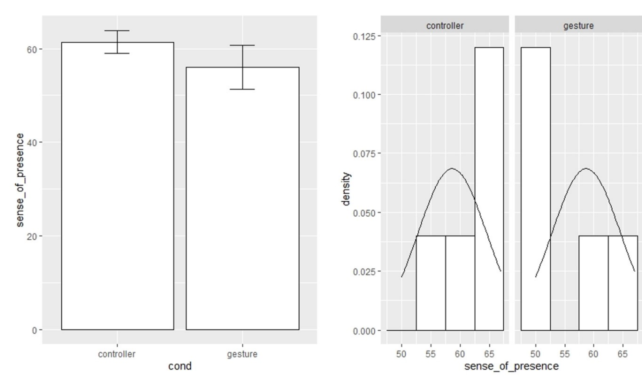

Figure 7: Sense of presence in the virtual environment, shown by condition.

5 Conclusion

Based on our pilot user study, we found that there is no difference in overall task success rate by interaction metaphor or object size. However, there did seem to be a trend favoring both smaller objects and the controller interaction metaphor. We expect that there might be a significant difference, especially for interaction metaphor, if the sample size were increased to an appropriate number. Our most distinct finding was that it takes longer to complete a task using natural gestures. This difference was large and significant. From observational evidence, we know that this mostly comes down to users in the gesture condition taking many more attempts to select the objects. Additionally, we found that workload is higher when using natural gestures, and as this workload increases, the performance also decreases. Finally, despite higher tendencies for immersion (pre), natural gestures led to a lower perceived sense of presence in the virtual environment (post). As this was a pilot study, the primary goals of the study were to take findings and apply them to a formal study. Our recommendations for future studies are as follows:

- Change the frequencies of the door to 25, 40, 55, 70, 85, 100, 115, 130 cycles/min. Increasing frequencies will better show the effect.
- Record every attempt to grasp/select each object. We expect to see more attempts in the gesture condition.
- Record the time it takes to grasp for each trial.
- Record how each attempt is failed (hand or ball collision) if applicable.
- Run at least twice as many participants to have a higher sample size.
- Implement setting for left-handed participants so that they are not disadvantaged by the direction the door is closing.

With these additions and everything we learned from our pilot study, we expect to see interesting differences between gesturing and using controllers as an interaction metaphor for near field grasping tasks.

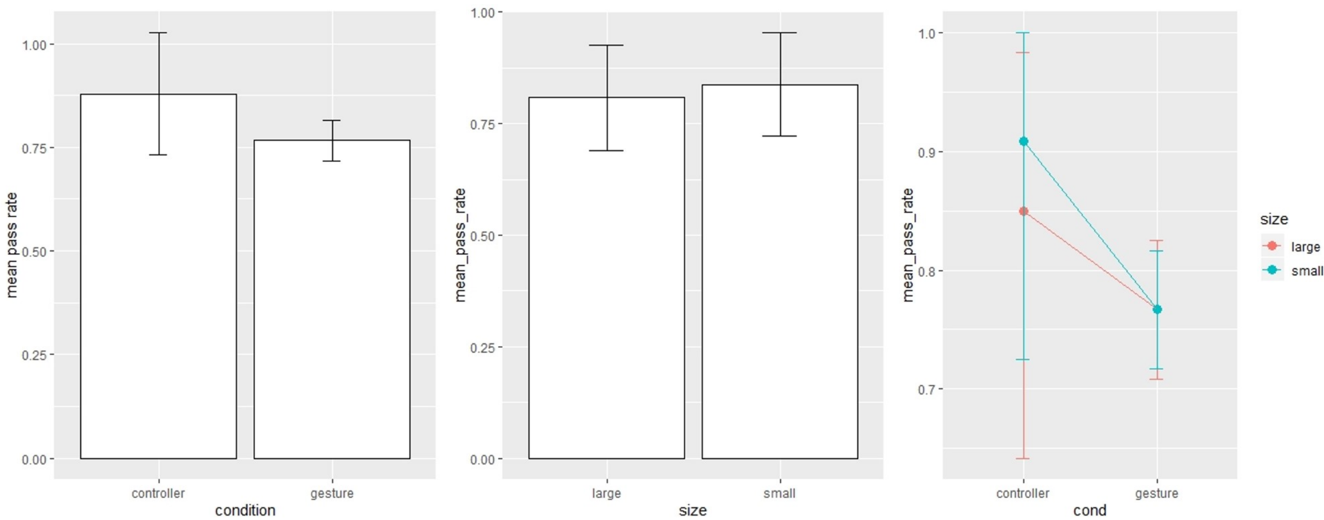

Figure 8: Target object size shown by condition

Figure 9: Mean success rates for all conditions across increasing door speeds.