DEGREE PROJECT IN MEDIA TECHNOLOGY,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2021

Grasping Techniques With Hand Tracking in Virtual Reality and Its Effect on the Sense of Presence

SOFIA BLOMGREN

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Abstract

Interactions in virtual reality are mostly done using handheld controllers, but hand motion tracking is becoming popular as an alternative. This study explores different types of grasping techniques with hand motion tracking, for objects that should be held in a specific way in order to be used. Can the same techniques and interaction behaviors be used with hand motion tracking as with physical controllers, or will there need to be adjustments made? Do different grasping techniques change the sense of presence while interacting with objects through hand tracking in virtual reality environments, and which grasping technique is preferred for hand tracking in VR?

A common grasping technique in games, using handheld controllers, is that objects automatically snap into the right grip when the user gets close enough. In this study, user tests were conducted to compare how this kind of assisted grasp compares to the user instead being visually guided to how to grasp the object, in the sense of feeling of presence, using hand motion tracking.

The results suggest that an assisted snap grasp can be used with hand tracking in VR without too much impact on the sense of presence, as it was the preferred grasping technique in this study. The guided grasp was not as successful as the snap grasp, since it was perceived as too restricting. The results don't necessarily mean that no kind of guided grasp would work, or even be preferred over an assisted snap grasp. A guided grasp might work well if it's not too strict and the design accommodates the many different ways people usually interact with objects.

Sammanfattning

Interaktion i VR utförs ofta men hjälp av fysiska handkontroller, men handspårning börjar bli ett populärt alternativ. Den här studien utforskar olika typer av metoder för att greppa objekt med handspårning, för objekt som behöver hållas på ett specifikt sätt för att de ska gå att använda. Kan samma greppmetoder och interaktionsbeteenden användas med handspårning som med fysiska handkontroller, eller måste de justeras? Ändrar olika greppmetoder närvarokänslan vid interaktion med objekt med handspårning i VR och vilken greppmetod föredras för handspårning i VR?

En vanlig greppmetod i spel, där fysiska handkontroller används, är att objekt automatiskt klickar in i rätt grepp i handen när användaren kommer till räckligt nära. I den här studien genonfördes användartester för att undersöka hur närvarokänslan påverkas vid den typen av automatiskt grepp, jämfört med att användaren istället får visuell hjälp med hur de ska greppa ett objekt, vid användning av handspårning.

Resultatet föreslår att automatisk hjälp att greppa objekt kan användas med handspårning i VR utan större påverkan på närvarokänslan, eftersom det var den greppmetod som föredrogs i denna studie. Greppmetoden med visuell hjälp var inte lika omtyckt, eftersom den upplevdes som för strikt. Resultatet betyder nödvändigtvis inte att ingen liknande greppmetod med visuell hjälp skulle fungera, eller till och med vara den metod som föredras över ett grepp med automatisk hjälp. En greppmetod med visuell hjälp skulle eventuellt kunna fungera bra om den inte är för strikt och om den även tar hänsyn till de olika sätt som användare interagerar med objekt.

Grasping Techniques With Hand Tracking in Virtual Reality and Its Effect on the Sense of Presence

Sofia Blomgren
EECS
KTH Royal Institute of Technology
Stockholm, Sweden
sofiablo@kth.se

ABSTRACT

Interactions in virtual reality are mostly done using handheld controllers, but hand motion tracking is becoming popular as an alternative. This study explores different types of grasping techniques with hand motion tracking, for objects that should be held in a specific way in order to be used. Can the same techniques and interaction behaviors be used with hand motion tracking as with physical controllers, or will there need to be adjustments made? Do different grasping techniques change the sense of presence while interacting with objects through hand tracking in virtual reality environments, and which grasping technique is preferred for hand tracking in VR?

A common grasping technique in games, using handheld controllers, is that objects automatically snap into the right grip when the user gets close enough. In this study, user tests were conducted to compare how this kind of assisted grasp compares to the user instead being visually guided to how to grasp the object, in the sense of feeling of presence, using hand motion tracking.

The results suggest that an assisted snap grasp can be used with hand tracking in VR without too much impact on the sense of presence, as it was the preferred grasping technique in this study. The guided grasp was not as successful as the snap grasp, since it was perceived as too restricting. The results don't necessarily mean that no kind of guided grasp would work, or even be preferred over an assisted snap grasp. A guided grasp might work well if it's not too strict and the design accommodates the many different ways people usually interact with objects.

INTRODUCTION

Throughout history, humans have used different kinds of media to express their emotions, ideas and experiences. From media like music and dance, to radio, photography, video and computer graphics, new inventions and technology have expanded the possibilities of expression. A more recent medium in this progress is virtual reality (VR), which allows users to step into and experience a world, created with computer graphics, as if they were actually in that world.

The main *physical* component of VR is the headset which the user wears over their eyes and through which they're allowed to see the virtual environment (VE). The headset collects several kinds of input, based on the user's behavior, to make changes in the VE. One very important kind of input in VR is the tracking of the user's physical movement, either just their head or several parts of their body. Physical movement can be tracked in several different ways, for example visually by cameras. Another kind of input is the one based on an actual decision the user makes in the VE by for example using some sort of handheld controller, like clicking a button or turning a wheel [15].

Many popular VR headsets[1] come with some sort of handheld controllers, one for each hand. These controllers are used to track the user's hand movement so it can be replicated in the VE. Buttons on the controllers are commonly used to make decisions, for example picking something up. In this report, "grasp" will be used to refer to the action of reaching for and picking something up, while "grip" will be used to refer to the way an object is held.

Oculus Quest is a VR headset that's very suitable for using at home since it requires little setup time and little space. While some other headsets require setup of external sensors for tracking, also known as outside-in tracking, the Oculus Quest headset has its sensors inside the headset, called inside-out tracking, allowing the headset to track the physical movement of the user by itself [12]. In late 2019, Oculus released hand tracking for their Quest VR headset. This feature enables users to navigate in and interact with VEs using only their hands, instead of using handheld controllers. This kind of technology has been around for a while but with Oculus Quest being very accessible to users who want to use VR in their homes, this opens up for a wider spread than before. This means that game developers will start creating VR games based on hand tracking, and also add hand tracking to some of their already existing games. There are pros and cons of both handheld controllers and hand tracking and a big difference between

[1] For example Oculus Quest, HTC Vive and HP Reverb

the two is the actual tracking possibilities. Controllers can usually be tracked from any position and angle around the headset. If the user wants to pick up something that they know is located behind them, they can technically do that without turning around. With hand tracking, however, the hands must be seen by the tracking cameras to be used. Another difference is interacting with free hands opposed to holding a physical controller. What does that change mean in terms of interactions? Can the same techniques and interaction behaviors be used with hand tracking as with physical controllers, or will there need to be adjustments made?

RESEARCH QUESTION

Do different grasping techniques change the sense of presence while interacting with objects through hand tracking in virtual reality environments, and which grasping technique is preferred for hand tracking in VR?

BACKGROUND

Five Key Elements of VR

In *Understanding Virtual Reality* [16] Sherman and Craig describe five key elements of VR. The first key element is the participant, the person who experiences the VE. Different participants will experience the same VR application in different ways, since their own imagination is a big part of the experience and will differ from person to person based on their own capabilities, interpretations and background.

The creators of the application are the second key element of VR. They create the application, but the VR experience can be seen as a team effort between the creator and the participant.

A virtual world can exist in many forms. It can exist as an idea in an author's head or written down as a movie script, but when shown through a medium like a movie, it becomes an experience. The same applies to VR. When a virtual world is shown through the medium of VR it becomes an experience. The virtual world is the third key element of VR and will in this report be referred to as the *virtual environment* (VE).

The fourth key element, immersion, is hard to put a finger on and is often described in slightly different ways depending on where you look. Generally, it can be described as how engaged and involved the user gets in the VR experience [16][10]. The word immersion is sometimes used interchangeably with the word presence, which is incorrect. When compared to each other, immersion stands for an objective point of view and is more based on the technology used, while presence stands for a subjective point of view and is based on the human reaction to the immersion [18]. Sense of presence can also be described as to which degree the user feels present in the VE [19]. This report will mostly focus on the sense of presence. Witmer and Singer's study from 1998 [22] is widely used for measuring sense of presence and was therefore chosen to be used in this study. Their study contains extensive presence- and immersive tendency questionnaires, from which questions can be picked depending on their relevance to the study. Their presence questionnaire measures the participant's sense of presence in the VE and the immersive tendency questionnaire measures how easily the participant tends to be immersed in things in everyday life.

The fifth and last key element according to Sherman and Craig [16] is interactivity. For a VE to seem authentic and make the participant experience it as if they were there, it should respond to the participant's actions. In many games, VR or not, interactions can include picking things up and putting them down again. An interaction more specific to VR is changing the point of view simply by the participant moving their head, just like in real life.

Tracking of Headset

It's very important that the head movements of the user are tracked correctly so the view of the VE changes realistically as the user moves around.

Degrees of Freedom

The orientation of the head can be represented by different values along the X, Y and Z axes, having their origin at the center of the user's head. This kind of tracking gives three degrees of freedom (DoF) and is known as 3 DoF tracking. Some trackers also measure position and that's called 6 DoF tracking [7].

Visual Tracking

There are several different kinds of sensors that can track these movements, including inertial, GPS and magnetic. A common type of tracking is visual tracking, which also has different kinds of variations within itself.

One variation of visual tracking is called *inside-out* tracking and is distinguished by the trackers being placed on the moving object itself, in this case the VR headset. The trackers map the environment and measure the changes as they move, to translate that into movement in the VE. One type of marker assisted inside-out tracking is called *lighthouse tracking* and it uses external base stations which emit infrared light for the trackers on the headset to read. This can make the tracking more reliable, but won't work without prior setup of the base stations.

Another variation of visual tracking is called *outside-in* tracking and is basically inside-out tracking reversed, meaning that the VR headset is tracked by external trackers. This kind of tracking is very reliable since the headset can be designed for optimal tracking, but just as with lighthouse tracking prior setup of external trackers is required [7][2].

Oculus Quest, the headset used in this study, uses 6 DoF inside-out tracking. This means that no external trackers and no prior external setup is required to use it, which allows for use in a smaller space and opens up for easier use at home.

Interactivity and Input

Aside from the headset, VR games and applications often also need some other way of collecting user input. A way for the user to make decisions and interact with the VE other than just moving around.

Handheld controllers

Handheld controllers are commonly used for both tracking the position of the user's hands and allowing the user to make decisions in the VE. Traditional gaming input solutions like buttons and joysticks are sometimes placed on these controllers, but more and more VR specific solutions are being added as well. Some examples are the Oculus Half Moon, which allows some basic finger gesture recognition, and the Reactive Grip, which gives the user haptic feedback [2].

Hand Tracking

Several attempts have been made to remove the handheld controllers, replacing them with different kinds of hand and/or gesture tracking and allowing the user to interact with the VE only by moving their hands and fingers. Some examples of technologies used for hand tracking are biofeedback, which measures electrical signals in the user's arms, optical tracking and data gloves [2].

Hand tracking can be done in several ways and while some technologies can give great tracking results, they might not be the best option in other aspects. One technology is marker based tracking, in which the user wears gloves with markers on them that can be read by a dome of cameras surrounding the user. This technique can give great results, but isn't very convenient for at home users. Another technology is using a single depth sensing camera, which can give good results and is a lot more convenient than using an entire dome of cameras. Oculus Quest, which is used in this study, uses its four already built in monochrome cameras placed at the front of the headset. These cameras are better than the single depth sensing camera in several ways, some of them being a wider field of view, smaller size, lighter weight and less expense. Those things make them great for at home use, but this technology also comes with a harder computer vision problem. These cameras lack direct 3D information about the hands and the visual of the hands will vary a lot from user to user depending on their environment [11]. The hand tracking on Oculus Quest starts with the four monochrome cameras first detecting the hands, which allows for the hand detection network DetNet to roughly compute a bounding box for each hand. The images are then cropped at the edges of the bounding boxes and those cropped images are sent to the hand keypoint network, KeyNet. KeyNet adds keypoints and predicts a 2D location heatmap and a 1D relative distance heatmap. A hand model is then shaped based on those keypoints. The information from the two previous frames are also used to predict the position in the next frame, to improve the estimation of each movement [8]. For this kind of hand tracking, a good visual of the actual hand is critical. The hands need to be placed in view of the cameras, placed at the front of the headset, for them to be tracked. As soon as a hand goes outside the field of view, it's not possible to use it in the VE. Another big problem is occlusion, meaning that it's also important for the user's hands to be in an angle that makes it possible for the cameras to read the hand movement. If some of the fingers are covered by the rest of the hand, the movements won't be tracked correctly [13]. This creates further problems with hand tracking and is something designers creating VR experiences for hand tracking need to have in mind and possibly design new interaction techniques for.

Handheld Controllers vs Hand Tracking

While it might seem like eliminating the controllers and letting the user interact with the VE using only their hands should create a more realistic experience, that might not be the case. To create a realistic experience in VR, interaction feedback is crucial [3][6][20]. With hand tracking the user can still be provided with visual and auditory feedback, but by eliminating the controllers the possibility of giving haptic feedback through controller vibrations is eliminated as well. Examples of common use of haptic feedback in VR are when the user touches or collides with something in the environment, when the user hovers over an item that can be picked up and when the user holds an item and it collides with something else.

Imagine picking up an item using hand tracking. Visually in the VE the hand will close around the object and stop moving when it reaches the edges, just as in the real world. The user's actual hand, though, will be able to keep moving and bending and feel as if it goes through the object without any haptic feedback. Will the user get used to that, or will it completely break the sense of presence? Does it matter to begin with?

Several studies have shown that controller naturalness improves the sense of presence. Controller naturalness means that the controller used to perform an interaction in the VE somehow mimics the way that same interaction

Figure 1. From left to right: Free grasp, Snap grasp, transparent Guide grasp hand.

would be performed in the real world, for example using a steering wheel to control a driving game. A game controller with *no* relation to real life interactions can start feeling more natural over time, as the user learns the interface, but a controller which allows the user to use real life knowledge will open up for a higher sense of presence from the beginning [17][21]. This implies that VR users might have gotten used to in-game grasping with standard controllers with buttons, but also that grasp interactions should benefit from using hand tracking since it's closer to reality. However, there are more aspects to look at. One study examined VR implementations of games where the user would naturally hold something in their hands, like tennis, and looked at what had the most impact on the sense of presence depending on which kind of controller was used. It was shown that when the user had some sort of controller or realistic object in their hand, the sense of presence was mostly impacted by *perceived reality*, but when playing with empty hands, *interactivity* had the biggest impact on the sense of presence [14]. This indicates that natural interaction is an important part of games with hand tracking and that the interaction techniques used in games with handheld controllers might not be directly transferable. When interaction suddenly becomes more important for the sense of presence, some changes might need to be made and new techniques considered.

Grasping with Hand Tracking

Just like in everyday life, it's common in games and other VEs for some objects to require a specific grip in order for the user to be able to use it. For example, a gun wouldn't be useful in a shooting game if it was held upside down or backwards. A common way to solve this is that those objects automatically snap into the right hand position when the user gets close enough. This is supposed to make it easier to interact with things and ensures that the object is held in the correct way. With the previous reasoning that interactions need to be more realistic when not using controllers [14], will this interaction feel unnatural with hand tracking? When removing the controllers, and with the user possibly being more set on the VE to actually respond as the real world, will an assisted grasp decrease sense of presence rather than being a helpful tool?

Visual information is critical in a reach-and-grasp movement. Before grasping an object, the brain analyzes the distance, angle, shape etc. to plan for the correct arm movement and to correctly shape the hand even before contact [5][9]. By just observing an action being performed, the human mind replicates that same action and prepares the observer to perform it themself. This is called common coding [4]. Ambrosini et al. showed in their study [1] that when presented with two different objects, which both require different kinds of grip to be picked up, people are more likely to pick up an object which they've previously viewed someone grasping rather than an object they've viewed someone just reaching out for and touching. Could visual guidance based on the required grip for each item therefore be an efficient and more natural grasping technique for a VE controlled by hand tracking?

METHOD

In this study, user tests were conducted to compare how an assisted grasp, where objects snap into the right position, compares to the user instead being visually guided to how to grasp the object, in the sense of feeling of presence, while using hand tracking in VR. The hypothesis was that a guided grip would be preferred over an assisted snap grip.

Test Cases

To test the hypothesis, three test cases were created in Unity 3D for Oculus Quest. All three test cases were set in the same VE and were based on the user picking up and putting down objects, but differed in grasping behavior. The three different test cases used the same scene setup and the same interactions, but with different grasping techniques, shown in Figure 1. The grasping techniques were the following:

<u>Snap grasp</u>

This test case had an assisted grasp that snapped the objects into the right position. The hand also locked into a predetermined pose. The hand shape was completely designed beforehand and the user only had to get close enough to the object. This grasping technique was based on how objects with specific grip requirements are commonly handled in VR games where physical controllers are used.

Guided grasp

In this test case, objects needed to be picked up from a specific direction with a specific hand angle. A static transparent representation of the hand showed the specific grip, guiding the user to the required hand pose. This grasping technique was based on studies showing that by viewing an action being performed, the human mind replicates that same action and prepares to perform the same action [4].

Free grasp

In this test case, objects could be grabbed in any way and from any direction or angle. A simple grasping technique, with no special constraints or guidance. Since the study is based on objects requiring a specific grip, this kind of free grasp wouldn't be a valid solution in an actual VE for the kind of problem researched in this report. This test case was the only one performed by all participants and was used to allow for a better comparison between the test groups, since the test environment would differ between participants due to the tests being held remotely.

Test Environment

The VE mainly consisted of a table with two white squares on top of it, between which the interaction would happen. Different objects would appear on top of the left square, one at a time, and the participant would pick them up and move them to the right square. When an object was placed on the right square it would disappear, also triggering the next object to appear on top of the left square. In case of an object being dropped out of reach, making it impossible to use it as a trigger to receive the next object, six cubes were available on the table. If needed, the participant could place one of the cubes on the right square and the next object would be triggered to appear that way. The basic Free grasp applied to the black cubes in all three test cases, since the use of them wasn't supposed to affect the outcome of the study. To minimize distraction from the interactions, no other specific details were added to the scene and Unity's default skybox was used for the surroundings. See Figure 2 for the start view of the VE.

Test Objects

Each test included 15 object movements, three iterations of five unique objects. 15 movements in total to give the participant some time in the VE and a chance for them to get comfortable in the environment. Five unique objects, instead of 15, so that the participant would get to try the same kind of grasp several times, giving them an opportunity to improve their technique during the test [17].

Figure 2. Test scene start view.

After the three iterations of the five objects, no more objects appeared and the test was completed.

The objects were chosen based on each having one specific grip with which it's commonly held, for example objects with handles and objects that would be difficult to use if they were held in a different way than that common way. All objects were also everyday objects that wouldn't be too unfamiliar to the participants. They were all also objects that would make sense to be seen on a table.

The objects (shown in Figure 3) were the following:

- Milk carton
- Coffee pot
- Frying pan
- Pencil
- Key

Evaluation and Expected Result

The hypothesis was that the best feeling of presence is achieved when the user is guided towards a specific hand placement. This was based on previous studies showing that interactions should be more realistic when using hand tracking instead of handheld controllers to increase sense of presence [14], and that visual guidance of someone grasping an object helps the human mind to prepare for grasping that same object [4][1].

To evaluate the hypothesis, several kinds of data were collected from the participants:

- Pre-study form, collecting information about previous knowledge of VR and hand tracking.
- Recording of the video call, capturing the participant's movements in the real world.
- Recording of the view inside the VR headset, capturing the participant's movements in the VE.
- Presence and immersion tendency questionnaires.
- Short verbal interview, structured around a few questions about the interactions.

Figure 3. From left to right: Milk carton, coffee pot, frying pan, pencil, key.

The evaluation was mostly based on the verbal interview and the presence questionnaire, while the other data was collected to help find other helpful or interesting information about the different grasping techniques.

Presence Questionnaire (PQ)

Several studies have been made to figure out how to measure presence. Witmer and Singer's study from 1998 [22] is widely used and was therefore chosen to be the one used in this study. Their study contains extensive presence- and immersive tendency questionnaires, from which questions can be picked depending on their relevance to the study. Questions in both questionnaires are answered on a seven point scale from least to most, with number four being a neutral middle.

To measure and compare the feeling of presence, the participants filled out a presence questionnaire after completing each test case. All questions were chosen from the study made by Witmer and Singer and were based on either general feeling of presence, or interactions in the VE.

The questions were the following:

1. How involved were you in the virtual environment experience?
2. How natural did your interactions with the environment seem?
3. How aware were you of events occurring in the real world around you?
4. How much did your experiences in the virtual environment seem consistent with your real-world experiences?
5. How quickly did you adjust to the virtual environment experience?
6. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
7. Did you learn new techniques that enabled you to improve your performance?
8. Were you involved in the experimental task to the extent that you lost track of time?

All questions were rated on a scale from 1 to 7, 1 being the least, 7 being the most and 4 being a neutral middle. The most positive impact on presence was indicated by 7 for all questions except question number 3, where the 1 meant the most feeling of presence. The scale for question number 3 was therefore inverted before added to the charts in the results, for an easier overview.

User Tests

Ten people, ages 24-45, participated in the study. Due to COVID-19, all tests had to be held remotely over video call. This also meant that all participants needed access to their own Oculus Quest headset, resulting in most participants having medium or advanced previous experience of VR. The participants were divided into two groups of five (Group A and Group B) for the tests. Based on experience levels, participants were spread out as evenly as possible between the two groups.

Tests were held remotely over video call, one participant at a time, with the participant having no previous knowledge of the goal of the study. They had received the scenes ahead of the session to transfer them onto their headset, but had been asked to not open the scenes beforehand.

Both Group A and Group B started out with the Free grasp scene, followed by either the Snap grasp scene or the Guided grasp scene, respectively.

Each session began with a short description of the structure of the test. The participant was then asked to open the Free grasp scene and received a short description of what they were supposed to do within the scene. The participant was then asked to complete the scene on their own, without talking unless necessary. The interviewer was muted while the participant was in the VE, to further decrease distraction.

After completing the Free grasp scene, the participant was sent a link to a presence questionnaire to be completed right away.

The participant was then asked to complete the group specific scene, followed by the same presence questionnaire as before. At this point they also filled out an immersion tendency questionnaire.

At the end of the test the participant got to answer some questions verbally and also discuss the differences between the two test cases, which grasping technique they preferred and why, and what they thought of the interactions in general.

RESULTS

The results are divided and compared in a few different ways. First they're compared within the two different test groups. Then all three test cases are compared at once. Lastly, the pros and cons are shown for each test case separately. The most positive impact on presence was indicated by 7 for all questions except question number 3, where the 1 meant the most feeling of presence. The scale for question number 3 was therefore inverted before added to the charts in the results, for an easier overview.

Test Group A (Free + Snap)

When asked during the final discussion which grasping technique they preferred, 4 out of 5 participants in Group A said they preferred the Snap grasp. Several people mentioned how the technology of hand tracking isn't perfect yet and that they felt that they had to be more precise in the Free grasp scene, which can be a struggle with hand tracking when compared to using other controllers. They felt that the Snap grasp helped them achieve what they wanted faster. Participant A3 claimed to feel a slight worry every time they were to pick something up with the Free grasp, since they were unsure that they would succeed. Participant A4 said they like when things just work, which made them enjoy the Snap grasp more. However, participant A2, who preferred the Free grasp scene, seemed to have had no trouble with the Free grasp but experienced it harder to let go of objects while using the Snap grasp.

The responses in the presence questionnaire are consistent with the final discussion (See Figure 4). For questions 1 to 7, the average scores show a higher sense of presence for the Snap grasp compared to the Free grasp. Question 8 shows a slightly lower score for the Snap grasp than the Free grasp. When compiling the total average score for all questions, the sense of presence seems to have been higher in the Snap grasp scene compared to the Free grasp scene (See Figure 5).

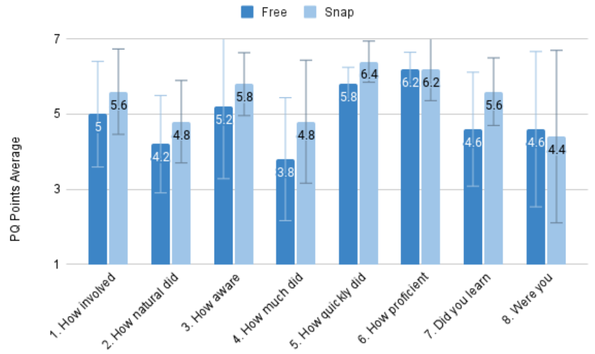

Figure 4. PQ average points per question - Group A

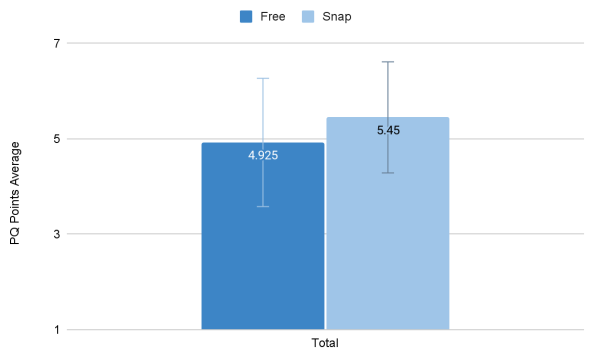

Figure 5. PQ average points total - Group A

Test Group B (Free + Guided)

When asked during the final discussion which grasping technique they preferred, all 5 participants said they preferred the Free grasp over the Guided grasp. They all mentioned how it was hard or felt unnatural to lift certain objects in the Guided grasp scene, since the guiding ghost hand in some cases didn't reflect the way they would normally interact with that object. Participant B2 experienced trouble lifting the pencil specifically, since they had previously broken a finger and now needed to accommodate certain hand poses which the scene didn't allow. Participant B5 talked about how they're usually very careless while using VR, since they know they can't really break anything. As an example they mentioned how they remembered trying to lift the coffee pot from the top in the Guided grasp scene, even though they thought the ghost hand made it very clear they were supposed to use the handle. Participant B4 found the Guided grasp too restricting and said that it might've worked better if it was a little bit easier to choose where to grasp the object. They mentioned the frying pan as an example, where they wanted to be able to grasp anywhere on the handle instead of at one specific spot. Participant B1 suggested that a more diffuse guide, rather than an actual hand with a specific shape, might have made them prefer the Guided grasp scene.

The responses in the presence questionnaire are consistent with the final discussion for Group B as well (See Figure 6). For questions 1 to 7, the average scores show a lower sense of presence for the Guided grasp compared to the Free grasp. Just as for Group A, question 8 stands out from the rest by receiving a lower score for the otherwise preferred grip. When compiling the total average score for all questions, the result clearly shows that the participants experienced a higher sense of presence in the Free grasp scene compared to the Guided grasp scene (See Figure 7).

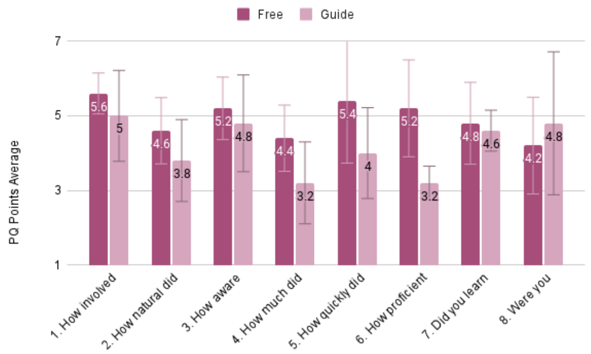

Figure 6. PQ average points per question - Group B

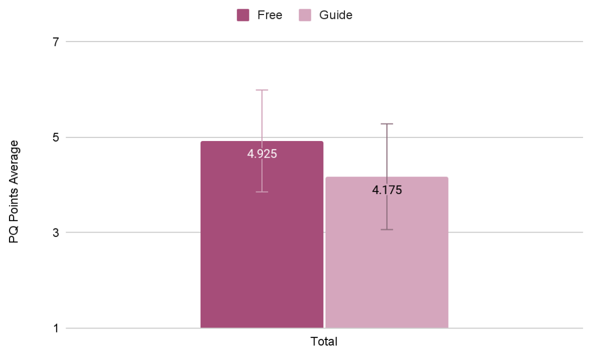

Figure 7. PQ average points total - Group B

Comparison of All Test Cases

It's clear from the discussions that the participants, at the end of the test, had different views on the Free grasp scene depending on which test group they were in, since they compared that scene to the group specific test case. Looking at the results from the presence questionnaire though, which was filled out before they interacted with the group specific test case, the two groups seem to have experienced a quite similar sense of presence while in the Free grasp scene. When compiling the total average score for all questions, both groups' score ended up the same (See Figure 8). This shows that comparing the results from the two group specific cases should be possible, even though some test factors differed between the participants.

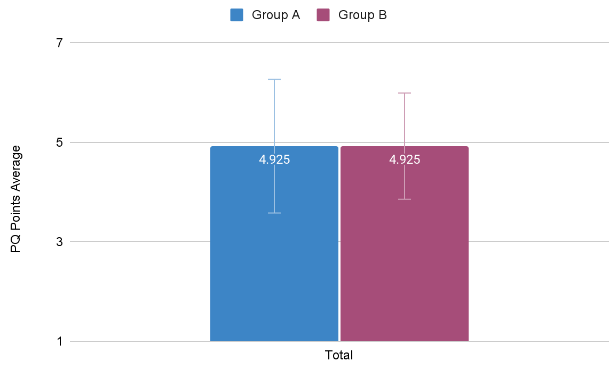

Figure 8. PQ average points total - Free Grasp Scene

Finally, when comparing the results from the presence questionnaires from the Snap grasp and Guide grasp scenes, it's clear yet again that the Snap grasp was preferred (See Figures 9 and 10).

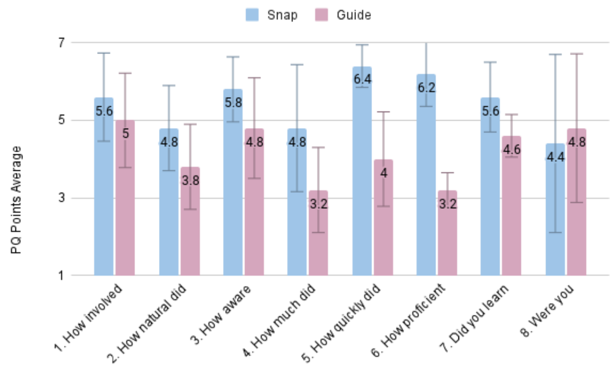

Figure 9. PQ average points per question - Snap vs Guide

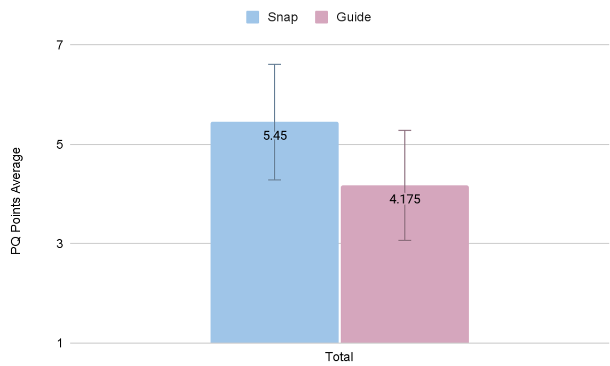

Figure 10. PQ average points total - Snap vs Guide

Free Grasp Scene

Both groups started out with the Free grasp scene, which was meant as a basic starting point. Since the participants compared this scene to one of the two other scenes, the comments varied based on group. The pros mainly came from Group B and the cons mainly came from Group A. Keep in mind that this grasp was only a basic starting point and wouldn't work as a valid final option when creating interactions for these kinds of objects. It was only used as a comparison to the other tests.

Pros

The Free grasp allowed participants to pick up the objects in any way they wanted, which opened up for each participant to decide how to interact with the objects.

Cons

Participants seemed to notice the technical limitations of hand tracking more clearly with this grasp than in the other test cases.

Snap Grasp Scene

This grasp was only tested by Group A and was discussed in comparison to the Free grasp.

Pros

Participants seemed to experience the least technical difficulties with this grasp. It also made them feel very efficient and they trusted the assisted grasp to help them achieve what they wanted.

Cons

Some participants did mention that it felt a little bit unnatural that the object and hand moved in a way they didn't control themselves, at the point when the object snapped into the hand.

Guided Grasp Scene

This grasp was only tested by Group B and was discussed in comparison to the Free grasp.

Pros

The participants didn't mention any pros for this grasp, although they did seem to understand the meaning of the guiding hand even though it didn't work in several other ways.

Cons

The participants felt forced to grasp the objects in ways they weren't used to, which made the whole experience feel unnatural and like too much work.

DISCUSSION

The purpose of this study was to compare how an assisted grasp, where objects snap into the right position, compares to the user being visually guided to how to grasp the object, in the sense of feeling of presence, while using hand tracking in VR. The hypothesis was that a guided grasp would be preferred over an assisted snap grasp, but the results show that most participants preferred the assisted grasp.

Participants in Group B perceived the Guided grasp to be more strict than the Free grasp. This study was based on objects that *need* a specific, or more strict, grip and for which the Free grasp wouldn't be an actual final option. However, Group A found the Snap grasp much more efficient than the Free grasp, which leaves the Snap grasp as the most successful test case in this study.

That doesn't mean that an assisted snap grasp ultimately is better than *any* guided grasp. There are several improvements that can be made to a guided grasp, which might make it the preferred grasping technique in the end. The thing that stood out the most from the Guided grasp discussions is that people interact with objects in many different ways, even though one might believe that there's only one common way to do so. Based on feedback from the results of Group B, the Guided grasp would possibly improve if it wasn't as strict. This could be done by making the guide more diffuse and allowing for more individual decisions, that would still make it possible to use the object. More research should also be made beforehand on how people usually grasp objects and how different every person's grip of the same type of object can be.

One thing that's interesting about the results from the Guided grasp test is that many participants complained about having to shape their hand in a way they're not used to, which actually isn't completely correct. The guiding ghost hand was shaped in a specific way, but the actual *shape* of the participant's hand wasn't important in order for the VE to allow them to pick up an item. The only constraints were the direction and the rotation of the hand, but the participant could've shaped their hand in any way they wanted. It would be interesting to further test this kind of guided grasp but with a more diffuse guide than an actual hand shape to allow for more individual interpretation by the user.

Method Criticism

Since the tests had to be done remotely, due to COVID-19, each test was done in a different environment. Space, lighting and external distractions are things that differed between the participants. To ensure a great user experience for everyone, VEs need to work in all kinds of environments. In a study like this, though, it's not ideal for the environment to be a variable and this can have affected the outcome of the study.

For the hand tracking to work, the headset needs to be able to correctly track the hands at all times. Because of restrictions, like bad lighting, some of the participants experienced problems with tracking. Participant B2 had to restart the Free grasp scene because of tracking issues and especially mentioned trouble letting go of objects.

The tests being held remotely also meant that each participant needed access to their own Oculus Quest headset. This meant that most participants work or study within this field and are very used to using VR. Because of this, many of them might have had a lot of previous experience using an assisted snap grasp, which could be a reason for most participants preferring it. More accurate results might have been reached if the participants had a wider range of previous knowledge.

Most of the negative feedback for the Guided grasp was that it didn't allow the participants to pick up the items the way they normally would. This wasn't the plan for this test case, the restrictions were there so the participants wouldn't for example pick up the coffee pot by the top, rather than by the handle. The restrictions could've been better optimized technically. Research could also have been made beforehand to see how people commonly pick up the items used in the study.

Future Work

The participants didn't have to actually use any of the objects in any of the scenes, they were only asked to move them from one place to another. If the participants would've had to use the objects in the scenes, for example pouring coffee from the coffee pot, they might've had a different view of the different grasps. For example, some of the participants tried lifting the objects incorrectly just because they could, but if they were told to actually use the objects that wouldn't have been as easy. This could be something to look at in future research.

The specific way the participants were guided in the Guided grasp scene in this study wasn't preferred over the other grasps, but that doesn't mean that an improvement of the guided grasp wouldn't be preferred over an assisted snap grasp. Participant B1 suggested that a more diffuse guide, rather than an actual hand with a specific shape, might've made them prefer the Guided grasp. It would be interesting to do a similar study, but with a more diffuse guide and with less strict constraints, to see if a different version of a guided grasp might be preferred.

In this study the assisted snap grasp was preferred. A similar study could be made in which this assisted snap grasp with hand tracking would be tested and compared to the same kind of grasping technique using physical handheld controllers.

CONCLUSION

According to the participant discussions during the user tests, users don't want to have to slow down and really focus on how to interact with objects when performing a task in VR. For their interactions to feel natural, the most important aspect isn't that the VE and the interactions exactly reflect the real world. It's more important that everything works smoothly and doesn't conflict too much with the user's intentions.

This study suggests that an assisted snap grasp can be used with hand tracking in VR without too much impact on the sense of presence. A guided grasp might work well, but it shouldn't be too strict and the design needs to make sure to accommodate the many different ways people interact with things in the real world.

ACKNOWLEDGMENTS

I want to thank both my supervisors who have helped me during this entire process. Kristoffer Benjaminsson at Fast Travel Games for technical guidance and for helping me build the test scenes and Björn Thuresson for academic guidance in everything from choosing a research question to writing the report. I also want to thank everyone who participated in the user tests.

REFERENCES

[1] Ambrosini, Ettore, Costantini, Marcello, & Sinigaglia, Corrado. (2011). Grasping with the eyes. Journal of Neurophysiology, 106(3), 1437-1442.

[2] Anthes, Christoph & García Hernandez, Rubén & Wiedemann, Markus & Kranzlmüller, Dieter. (2016). State of the Art of Virtual Reality Technologies. 10.1109/AERO.2016.7500674.

[3] Bozzacchi, C., Volcic, R., & Domini, F. (2014). Effect of visual and haptic feedback on grasping movements. Journal of Neurophysiology, 112(12), 3189-3196.

[4] Chandrasekharan, Sanjay & Mazalek, Alexandra & Nitsche, Michael & Chen, Yanfeng & Ranjan, Apara. (2010). Ideomotor Design: using common coding theory to derive novel video game interactions. Pragmatics & Cognition. 18. 2-313. 10.1075/pc.18.2.04cha.

[5] Desmurget, Michel & Pélisson, Denis & Rossetti, Yves & Prablanc, Claude. (1998). From Eye to Hand: Planning Goal-directed Movements. Neuroscience & Biobehavioral Reviews. 22. 761-788. 10.1016/S0149-7634(98)00004-9.

[6] Dinh, H., Walker, Hodges, Chang Song, & Kobayashi. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. Proceedings IEEE Virtual Reality (Cat. No. 99CB36316), 222-228.

[7] Gourlay, Michael J, & Held, Robert T. (2017). Head‐Mounted‐Display Tracking for Augmented and Virtual Reality. Information Display (1975), 33(1), 6-10.

[8] Han, S., Liu, B., Cabezas, R., Twigg, C. D., Zhang, P., Petkau, J., ... & Wang, R. (2020). MEgATrack: monochrome egocentric articulated hand-tracking for virtual reality. ACM Transactions on Graphics (TOG), 39(4), 87-1.

[9] Janssen, Peter & Scherberger, Hans. (2015). Visual Guidance in Control of Grasping. Annual review of neuroscience. 38. 10.1146/annurev-neuro-071714-034028.

[10] Kim, G., Biocca, F.: Immersion in virtual reality can increase exercise motivation and physical performance. In: International Conference on Virtual, Augmented and Mixed Reality, pp. 94–102. Springer, Cham (2018)

[11] Oculus. (2019). *"Hand Tracking Deep Dive: Technology, Design, and Experiences"*. Retrieved February 5, 2021. https://youtu.be/gpQePH-Ffbw

[12] Oculus. *"Oculus Device Specifications"*. Retrieved January 3, 2021. https://developer.oculus.com/learn/oculus-device-specs/

[13] Parger, M., Tang, C., Xu, Y., Twigg, C., Tao, L., Li, Y., ... & Steinberger, M. (2020). UNOC: Understanding Occlusion for Embodied Presence in Virtual Reality. arXiv preprint arXiv:2012.03680.

[14] Shafer, Daniel & Carbonara, Corey & Popova, Lucy. (2014). Controller Required? The Impact of Natural Mapping on Interactivity, Realism, Presence, and Enjoyment in Motion-Based Video Games. Presence. 23. 267-286. 10.1162/PRES_a_00193.

[15] Sherman, William R. & Craig, Alan B. (2018). The Morgan Kaufmann Series in Computer Graphics, Understanding Virtual Reality (Second Edition), Chapter 4, Pages 190-256.

[16] Sherman, William R. & Craig, Alan B. (2019). The Morgan Kaufmann Series in Computer Graphics, Understanding Virtual Reality (Second Edition), Chapter 1, Pages 4-58.

[17] Skalski, Paul & Tamborini, Ron & Shelton, Ashleigh & Buncher, Michael & Lindmark, Peter. (2011). Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. New Media & Society. 13. 224-242. 10.1177/1461444810370949.

[18] Slater, M.: A note on presence terminology. Presence Conn. 3(3), 1–5 (2003)

[19] Slater, M.: Immersion and the illusion of presence in virtual reality. Br. J. Psychol. 109(3), 431 (2018)

[20] Swapp, D., Pawar, V., & Loscos, C. (2006). Interaction with co-located haptic feedback in virtual reality. Virtual Reality, 10(1), 24-30.

[21] Tamborini, Ron & Bowman, Nicholas. (2010). Presence in video games. Immersed in Media: Telepresence in Everyday Life. 87-109.

[22] Witmer, Bob G., & Singer, Michael J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence, 7(3), 225-240.