

*Deictic pointing and mobile device  
interaction techniques in 360-degree  
videos*

---

MASTER'S THESIS  
GROUP 161031





**AALBORG UNIVERSITY**  
STUDENT REPORT

**Department of Architecture,  
Design and Media Technology**  
Rensburggade 14  
9000, Aalborg Øst  
Telephone +45 9940 8793  
<http://create.aau.dk>

**Title:**

Deictic pointing and mobile device interaction techniques in 360-degree videos

**Theme:**

Master's Thesis

**Project period:**

Spring 2016, Semester 10  
February 1st - May 25th

**Project group:**

MTA161031

**Members:**

Teemu Luhtapuro

**Supervisors:**

Martin Kraus

**Pages:** 49

**Abstract:**

The consumer market has different mobile devices and interaction techniques to experience 360-degree videos. Deictic pointing is referential communication where a person wants to draw another person's attention to an object. Deictic pointing often occurs in e.g. guided museum tours where the guide points to draw attention to objects or areas in and outside the field of vision. Thus, it was necessary to investigate if people were capable of perceiving deictic references outside the field of vision. This work investigated differences between mobile interaction techniques when experiencing 360-degree videos. The work tested participants' ability to detect and determine the target of deictic pointing directed outside the field of vision. Three interaction techniques were used in the experiment: click and drag, IMU-based rotation, and mobile virtual reality. These were compared to a real world experience to look for differences between the real world results and the interaction techniques. The test showed, that the real world results were significantly better than the interaction techniques. The results also indicated that the interaction techniques were equal. Thus, it does not matter which interaction technique is used to experience the 360-degree videos including deictic pointing gestures.



# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>State of The Art</b>	<b>5</b>
2.1	Motivation . . . . .	5
2.2	Gestures . . . . .	7
2.3	Deictic Pointing in Videos and Virtual Environments . . . . .	9
<b>3</b>	<b>Implementation</b>	<b>11</b>
3.1	Ricoh Theta S . . . . .	11
3.2	Unity . . . . .	14
3.3	Interaction Techniques . . . . .	16
<b>4</b>	<b>Limitations of Design</b>	<b>21</b>
<b>5</b>	<b>Experiment</b>	<b>23</b>
5.1	Hypothesis . . . . .	24
5.2	Participants . . . . .	24
5.3	Equipment . . . . .	25
5.4	Experiment Apparatus . . . . .	26
5.5	Procedure . . . . .	27
5.6	Questionnaire . . . . .	28
<b>6</b>	<b>Results</b>	<b>29</b>
6.1	Results from The Questionnaires . . . . .	31
6.2	Observations . . . . .	34
<b>7</b>	<b>Discussion</b>	<b>37</b>
7.1	Possible Future Implementations . . . . .	39
<b>8</b>	<b>Conclusion</b>	<b>43</b>
<b>A</b>	<b>Experiment Questionnaire</b>	<b>47</b>



# 1 | Introduction

Applying 360-degree spherical equirectangular videos as virtual tour videos, is a new concept in the consumer market. The consumer market consists of multiple devices and methods for exploring these videos, and therefore, it is necessary to know the main differences between these methods. As users behave differently, while experiencing these videos, it is unclear how nonverbal pointing in and outside the field of vision affects the user's actions in these devices and methods.

For instance, when the host in e.g. a museum or in real-estate tour, wants to draw attention to specific objects, they very often point at these objects. However, these settings often include multiple objects, and therefore, when using 360-degree videos as a way to experience locations such as these, it is important to know how well a person can receive these deictic references and determine, which object the host points at.

According to Kita [10], the gesture pointing is one of the fundamental blocks of human communication and has been a very important tool for human communication through the ages. These gestures have been studied in multiple areas and we humans develop the skill to detect gesture pointing already in the first years of our lives. A study by Jancovic et al. [9] provides information on gesture detection by infants, which is the early stage of the human development. As development continues, the use of gestures increases with age, and when humans are adults, the development from conventional gestures transform towards more complex gestures. Gesture pointing, also known as deictic pointing, occurs frequently in verbal communication and is very often an unconscious activity to lessen the cognitive load [12] during a conversation. As pointing is a very powerful gesture, in order to draw a person's attention towards an object – the referent, it is very often used in museum- or real-estate tours. Studies [23] [5] show that people are capable of interpreting nonverbal gesture pointing and to read people's body language, even through video. Therefore, it is interesting to investigate whether people are capable of detecting these deictic references in 360-degree videos on mobile devices.

The experiment in this project focuses on how people experiencing different interaction techniques on mobile devices, can interpret deictic pointing in 360-degree videos. Studies[27] have been conducted on accuracy of detecting gesture pointing in collaborative virtual environments, but not much has been studied on 360-degree videos.

The work investigates the differences of the deictic pointing in the mediated 360-degree videos on different mobile devices, which consumers could use for virtual tours and events. During

the project, a test was conducted to compare three platforms to test nonverbal pointing in a 360-degree panorama viewing/virtual reality. The purpose of this test was to identify possible differences between these three interaction techniques. The test included a 360-degree panoramic video, which included nonverbal pointing to target areas. The video was tested in three different platforms; a Youtube styled click and drag, an IMU-based rotation application, and a mobile virtual reality application.

Many studies claim, that the real world situation is the best reference situation. Therefore, it is essential to test and compare the different interaction techniques with the real world results. However, as it is commonly believed that the real world test is the best of these interaction techniques, it is predicted that the IMU-based rotation and mobile virtual reality will be close to each other. Lastly, the click and drag technique is expected to be the least successful of these platforms.

This thesis is structured as follows: To set a theoretical basis for the project, Chapter 2 starts by explaining the motivation for the research and then continues with reviewing the research on gestures and gesture pointing. In Chapter 3, implementation is discussed in five parts: video, Unity, click and drag navigation, IMU-based rotation, and mobile virtual reality. In Chapter 4 the limitations of the project's design are discussed and in Chapter 5 the experiment is explained. Chapter 5 starts with a description of the experiment and then continues with the experimental apparatus and the procedure itself. Section 6 has the results of the experiment and the questionnaires that show all outcomes of the experiment. Lastly, is the discussion and conclusion where the results will be discussed and evaluated for the final conclusion.

It is important to mention, that this study is concerned with mobile virtual reality. However, in order to make a simple definition, this is referred to as virtual reality throughout the project.

## 2 | State of The Art

### 2.1 Motivation

As the devices and the virtual reality tools are becoming more accessible for the consumer market, it is high time to investigate this field. However, the raising popularity of 360-degree videos in consumer products are causing the consumer market to offer multiple devices, ready for experiencing these 360-degree videos. Thus, depending on the hardware, the users have many different options to choose between when they want to experience these 360-degree virtual experiences. As the chances are high, that users will apply mobile devices to experience these videos, it is necessary and interesting to investigate the differences between these mobile devices and navigational options they carry.

Multiple high tech companies have expanded their product line to virtual reality and 360-degree cameras. For example, the company Matterport has been in the news a lot as they provide 360-degree cameras allowing users to visit real-estate apartments and houses. [26] Alongside this, Nokia and Qualcomm are both investing in 360-degree technology. Nokia has released a high-end 360-degree camera, Ozo, which is advertised as the world's first professional 360-degree video camera. [17] Qualcomm, on the other hand, has invested millions of dollars in a 360-degree video company called 360fly which provides products and software for 360-degree videos. [24] Alongside these few mentioned ones, more cameras exist. Examples of these are the Ricoh Theta S, the Sphericam, the Samsung 360, and the LG 360 cam. This shows, that there is a great interest in these devices and lately, they have increasingly been used in virtual tours and for event purposes.

As mentioned earlier, there is an increasing interest in using virtual reality in different consumer market fields such as in tourism or real-estate sales. Lately, virtual tours in distant places like museums and historical sites have become popular alongside with high-end virtual panorama real-estate tours. However, they have not completely conquered the market yet, and therefore, there is an opportunity to investigate some fundamental parts of the field. For example, how mobile devices differ from each other as the market is filled with countless different devices to experience these virtual tours. In these tours, the user very often has a tour guide or -host to guide the user through the virtual environment which can be either a 360-degree video or a digital environment. This raises an interesting question as the tour guides in these different areas often depend on one of the basic communication gestures, being gesture pointing.

Imagine a host that narrates and guides the users through the video by describing the environment and pointing out interesting facts and objects in the video. The moment when the host points at something, it becomes interesting, as the users with multiple different devices are experiencing the same video with different interaction techniques and therefore, they might perceive these pointing gestures differently. Even in the real world, it is highly possible that a tour guide is pointing somewhere outside of the viewer's field of vision to attract the user's attention to a referent. Therefore, it is important to know if the person immersed in a 360-degree video is capable of detecting the direction and the target of gesture pointing.

When experiencing 360-degree videos, depending on the device, a user is experiencing a certain level of telepresence. Telepresence is the sense of being in a mediated environment, e.g, a remote or virtual environment. Telepresence was conceptualized by Marvin Minsky [15] in the 1980s and has grown a huge research field.

As telepresence is the sense of being in mediated environments, Minsky [15] also argued that the more the senses are used, the more immersed you are in the mediated environment. Therefore, virtual reality would use more natural senses than the click and drag technique on a tablet screen which would mean that virtual reality would be the logical choice for a natural interaction method for 360-degree videos and gesture pointing in those videos. As virtual tours try to be substitutes for real world tours, the logical choice would be virtual reality glasses for experiencing this virtual tour. However, when moving to mobile devices as interaction platforms, which includes smartphones and tablets as devices for experiencing the 360-degree videos, the level of immersion is not as high quality as the perfect virtual reality or high-end virtual reality powered by powerful computers.

According to Bowman[2], interaction techniques are methods for accomplishing tasks that allow users to navigate, select and manipulate the content with inputs devices. Examples of such input devices can be keyboards, computer mice, trackballs, joysticks and touch screens. [2] However, this project focused on the modern mobile device input methods that are being used in the consumer market.

The touch screen input is the most popular input method on the modern mobile devices. The touch screen input allows user to control and navigate content on touch screen devices. For example, to navigate on a web page, user uses the finger swipes to move around the web page.

Other method on modern mobile devices that allows interaction differently is the inertial measurement unit based input. [18] These devices have a built in technology that detects the device's orientation in 3D space. This technique is often used in applications to navigate or rotate objects by orienting the device. For example, on mobile racing games this technique is often used to steer a car by orienting the device to left or right.

The mobile device interaction techniques, this thesis is focusing on, are the following: Click and drag, IMU-based rotation and virtual reality. In short, the click and drag technique is the most popular of the interaction methods, as the interaction is done by finger sweeping and dragging on the device's touch screen. Depending on the application, the application will react to the finger input accordingly, and in this case, the finger is used to rotate the camera

in a 360-degree video. The Inertial measurement unit(IMU)-based rotation is a interaction technique that allows the users to move and rotate handheld mobile device in order to explore the 360-degree video. The virtual reality interaction technique uses the same technology as the IMU-based rotation. However, instead of using hands to rotate the mobile device around, the mobile device is located in a low budget headset. This interaction type allows the person to use their head and body naturally to rotate around the 360-degree video.

## 2.2 Gestures

Throughout history, gesture pointing has been a fundamental part in conversation and human interaction. Kita [10] describes gesture pointing to be a “foundational building block of human communication” and a unique human behavior that distinguish us from other primates. However, gesture pointing is used for multiple purposes and therefore, detecting these gestures is a complex and very often a psychological phenomenon. To gain a better understanding of gesture pointing, it is necessary to investigate what gestures in general are, and from there, advance to the gesture pointing itself.

According to McNeill[13], gestures are a byproduct of the speech production and therefore, gestures are closely connected to verbal communication. Furthermore, gestures are the first communicative tools of a human baby. Very often, babies rely on these gestures to communicate their needs, and as babies grow, they rely even more on hand gestures as their communicative tool. According to Jancovic et al.[9], when the babies start growing older the use of communicative gestures increases and eventually these basic gestures transform to more complex and more precise gestures for the verbal and non-verbal communication.

These gestures are the communicative tools that very often occur unconsciously and therefore, we do not pay attention to them. However, they are a fundamental part of human communication and even the smallest gesture movements can make an impact. Furthermore, gesture pointing is just one of the types of multiple gestures. According to McNeill [14], there are at least four different distinct ways of using gestures. These four ways are iconic, metaphoric, beat and deictic gestures. The iconic type is where a person is using hand gestures to depict an action sequence that is also measured to give a better understanding of the situation. For example, if a person is speaking of throwing a ball in a conversation, that person uses his body language to mimic the action while speaking. Therefore, iconic gestures depict an action, or object, in a conversation to illustrate what is being said.

Metaphoric gestures are the most complex gestures out of the four types. According to McNeill [13], metaphoric gestures are another type of referential gestures, but instead of having a reference to concrete objects and events, metaphorical gestures use this concrete reference to represent an abstract thought. For example, making a gesture to one’s heart can indicate love or care.

Beats are gesture movements that do not have a direct meaning, however, they are very often visible with speech, and are small and quick movements of hands, fingers, or arms.[14] According to Lucero et al. [12], beat gestures are often visible when a speaker tries to look for a missing

word. When they occur, they do not need to be in a specific position or space, they are just performed where the hand is situated the moment the beat gesture is happening. According to Lucero et al.[12], beat gestures are simple repetitive movements that are not as cognitively taxing as the iconic gestures. Therefore, these gestures happen unconsciously and do not have symbolic meaning but instead they emphasize words.

Deictic gestures are pointing movements, which are performed with the pointing finger or any extensive object or bodypart that allows pointing like e.g stick or leg. [14] Deictic gestures are a type of referential communication where the whole purpose of the gesture in a communication is to direct the collaborator's attention towards a referent. A referent is an object or a thing that is referred to. Deictic gestures are gestures that can provide more ways to pointing than just direct finger pointing. According to Sauppé and Lite [22], there are several kinds of deictic gestures to draw attention to the referent, for example Sauppé and Lite introduced in their article the following gestures: presenting, touching, exhibiting, sweeping, grouping and pointing. Furthermore, they applied these deictic gestures on a human-like robot so they could evaluate the effectiveness of these gesture in referential communication. Their study showed, that different deictic gestures have different levels of effectiveness in communicative accuracy. For instance, their results demonstrated that physical gestures had the highest effectiveness. Deictic pointing gesture in their study seemed to have the best results from the non-physical gestures. [22]

Lempers [11] conducted a study on infants and studied finger pointing with 3 to 12 months old infants. He studied whether the infants were capable of understanding the referential meaning of the pointing finger and if the infants could find the referent either in the field of vision or outside of it. The study by Lempers[11] showed that already infants can react to deictic references to outside of the field of vision. Therefore, based on Lempers [11] arguments, an adult is capable of detecting these deictic references as well. A similar study by Callaghan and Corbit [3] states, that infants of 10 months are already capable of perceiving and detecting the referent from a gesture pointing in and outside the field of vision. Furthermore, a study of Jancovic et al. [9] indicates, as babies get older towards an adult age, the use of gestures also increases. However, the basic gestures transform from conventional gestures to more complex ones. Kita[10] concludes in her study that understanding and producing finger pointing is a unique skill and is a skill that shares attention towards objects with deictic gesture which only humans can fully understand. Furthermore, the study also has a big emphasis on the fact, that when babies use gesture pointing they can increasingly detect these, the older they are. Kita also concluded, that the pointing is one of the routes that eventually lead towards learning the language itself.

According to McNeill[14], gestures are very closely correlated with linguistic speech with adults and younger people. Interpreting deictic gestures without linguistic cues is a process that uses a lot of information about the gesture and the environment itself. McNeill says, that when we are determining the direction and the referent of a deictic gesture, we use all the information of the situation, orientation, expectation, the gesture itself and linguistic cues which according to McNeill [13], are very important for the interpretation of gestures as the gestures are closely correlated to speech channels. According to Bangertter and Oppenheimer[1], the detection of

pointing is highly accurate. This is comparable to a study made by Gale and Monk, who studied the accuracy of detecting the eye-gaze in a face-to-face conversation and video mediated links. Their study concluded, that people are highly capable of detecting one's eye-gaze directions and targets. Therefore, Bangerted and Oppenheimer[1], concluded that detection accuracy of pointing can be compared to Gale and Monks [5] study and the interpretation accuracy of pointing is as good as detecting eye-gaze in a conversation. Furthermore, a study by Schmidt [23] investigated attention-directing gestures aka deictic gestures and their ability to contribute to attention focusing towards surrounding or referents in an event. Schmidt found, that with no linguistic cues as helpers, adults were capable of perceiving and interpreting the deictic gestures.

## 2.3 Deictic Pointing in Videos and Virtual Environments

Many extensions and cues are used to direct attention towards the referent. Especially using tools to apply these pointing gestures more effectively. However, other extensive tools are used that extend the length of the pointing to become more accurate. A good example is laser pointers during a class. Laser pointers have been popular in the past, and demonstrate a perfect accuracy where the pointing gesture is directed at and what the referent is. There have been more ways of drawing attention towards a referent. Especially, an article by Rusch et al. [21], explored the idea of using augmented reality to improve the detection of possible objects and cues that require attention in a truck driver's environment. While driving, the augmented reality indicated of possible cues in the field of vision and increased the driver's attention for possible threats on the road.

However, using typical hand gesture pointing is still used in the consumer market quite a lot. Especially the newly arrived and popular 360-degree videos where people do not use any extensive tools to point towards the referents.

Moreover, there have been studies that investigated the gesture pointing in games and especially collaborative virtual environments. In a study about virtual avatars, Moore et al. [16] stated, that gestures work differently in collaborative virtual environments, especially in games. They say, that gesture pointing becomes difficult in games and thus, text or verbal communication is often needed in order to fully understand the gesture. However, Wong and Gutwin [27] studied accuracy of pointing in and out of collaborative virtual environments(CVE) to measure differences between CVE and real world results. The study concluded, that the CVE pointing is not yet accurate and therefore, CVE does not give a good support for pointing yet.

A study by Gale and Monk [5] investigated the accuracy of gaze awareness in a video-mediated environment. They concluded, that the participant's estimation of gaze was no more difficult over a video link than on face-to-face real world conversation. As the study has tested just the eye gaze awareness and estimations, it cannot prove if the gesture pointing can be estimated as accurately in a mediated video environment as in the real world. As the project revolves around 360-degree videos, this study gives an opportunity to test if the gesture pointing has the same effect as Gale and Monk's study.

Therefore, many of these earlier mentioned studies have been focusing on differences between the real world and one specific digital platform like virtual reality. However, Wong and Gutwin's [27] study on pointing accuracy and interpretation provided information that there is a difference between virtual environments and the real world. These studies raise an interesting question; if there are any significant differences between gesture pointing in different 360-degree video-mediated environments and navigational methods, and how do they compare with the real world situation? For example, if there are any differences between virtual reality or tablet navigation (rotational click and drag). However, the study by Gale and Monk compared to Wong and Gutwin's collaborative virtual environment, was in a video-mediated environment, and therefore, brings a question if the pointing can be perceived as well as the eye gaze in a video-mediated environment or as a gesture pointing in CVE which was not as good as real world gesture pointing accuracy.

As mentioned in the motivation section 2.1, the whole motivation for the project's idea comes from virtual tours in real-estate, tourism and museum tours. The section also stated, that the tour guides very often rely on deictic pointing to refer to objects in the surroundings and very often the pointing is directed outside of the user's field of vision. Therefore, it is necessary to investigate if the users can interpret the deictic references to outside the field of vision in 360-degree videos in a mediated environment. If users can interpret deictic references from the mediated environment, is there a difference between multiple consumer market devices?

## 3 | Implementation

The section explains how the implementation process is carried out, starting with videos and how they were created. Then it proceeds to the Unity game engine in order to explain how everything was done in this engine to make the applications ready for the experiment.

As mentioned in Chapter 5, the experiment consists of four test. To have a different number sequence in each test, four videos were needed. For each video, an application was created which meant that each test had four different applications. The reason for this method was because the applications did not have a good inbuilt system to change the videos for the application and therefore, separate applications were needed.

### 3.1 Ricoh Theta S

The videos, that were used in the experiment, were recorded with a Ricoh Theta S, omnidirectional camera. Ricoh Theta S in a video recording mode runs at full hd resolution 1080p and does 30 frames per second. Alongside the video recording capabilities the camera has a 12 megapixel photo quality. However, the most important part of the camera is that it is meant for the 360-degree video recording with its two cameras that point to opposite directions. Each camera covers an 180 degree fisheye view and therefore, the whole system covers 360-degrees around the camera. [20]



Figure 3.1: (a) Example of the raw dual fish eye format from the Ricoh Theta S. (b) Example of the equirectangular video converted by Ricoh Theta's official application.

When exporting videos from the camera, the videos were in a dual fish eye format. See Figure 3.1 (a). In order to create the equirectangular video for the spherical viewing, conversion software was needed. Ricoh Theta’s official app from its manufacturer included a software that converts the video into an equirectangular video clip (see Figure 3.1). As seen in the Figure 3.1 (b), the conversion is very well made, and the software does a great job on making a seamless conversion to an equirectangular video. The video exported from the software was in .MP4 format which is easily recognised by many media players. The software usage was very straightforward as it asks the user to drag and drop a spherical image or video from the camera to the application. Once the user drops a file from the camera to the software, it automatically starts converting the file to an equirectangular form and after it allows to watch the material in a 360-degree mode or in the equirectangular form.

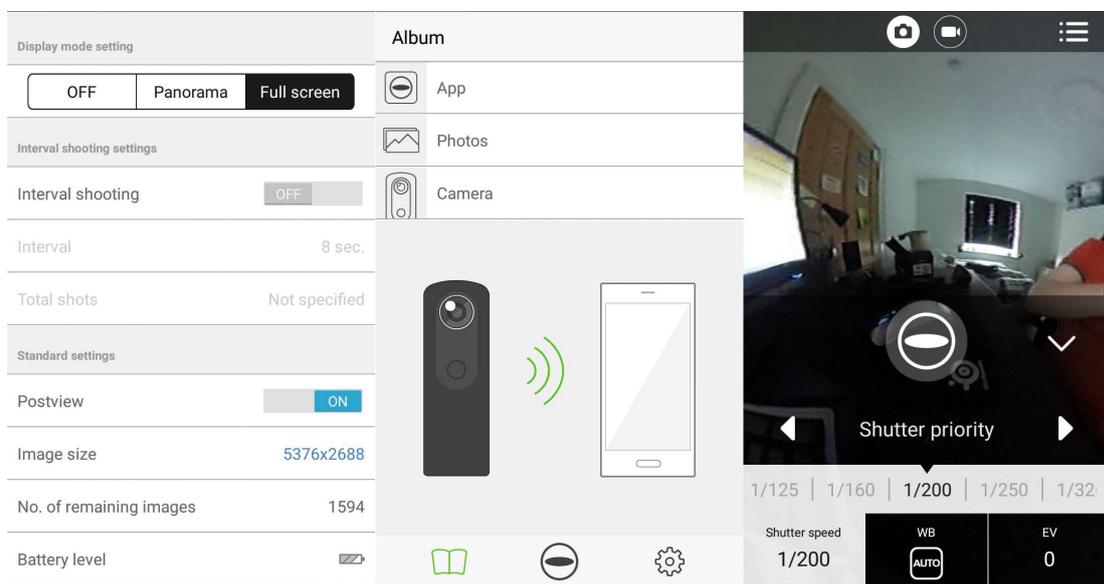


Figure 3.2: Ricoh Theta S camera application, where the camera can be controlled and set ready for filming.

The Ricoh Theta S camera is not a typical camera that has all the settings buttons on the camera itself. Instead of traditional buttons and switches, the Ricoh Theta S camera is controlled by a smartphone application that has all the features that a normal camera has in its menu. It includes controls from video resolution to exposure settings, and even a live view is possible from the camera to get a brief glimpse on how the camera is set up. A Wi-Fi connection is required in order to connect the camera with the smartphone. While taking pictures and videos, the inbuilt Wi-Fi of the camera can easily transport the files to the user’s smartphone. (See Figure 3.2)

### 3.1.1 Creating The Videos for The Experiment

Filming was done in classroom 3.529 at the Aalborg University Create building. The experiment required a lot of space so the classroom seemed a logical choice for the purpose, as the room

was not required to have any special features for the experiment. The numbers used for the experiment were attached to a wall of the room. The numbers were in order from 1 to 12, were the size of an A5 paper, and had 15cm space between them on the wall as seen in the Figure 3.3. The number papers were 158 cm from the floor and the participant's position was centered between number 6 and 7.

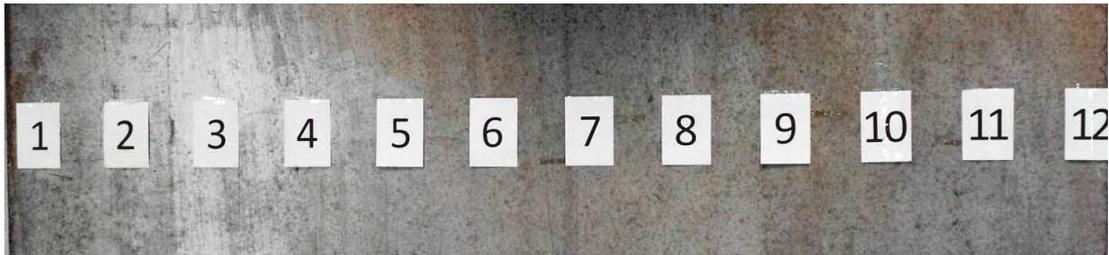


Figure 3.3: The numbers on the wall are a5 size and have 15cm distance between each other. Same number setup was used in the experiment.

The setup was tested multiple times to make sure that the camera, the wall, and the host were in a comfortable distance from each other. The safe distance between the host and the wall was three meters and the camera/participants sitting in the middle on the point of 1.5 meters from the wall and the host. This setup of the filming was necessary to remember, because in the experiment, the real world test was part of the experiment and required the same identical environment from the video to make results that can be compared to the other three tests.

As mentioned in Chapter 5, the experiment had four different videos with different number sequences in order to not use the same sequence to lessen the memorization of one number sequence. All the videos needed to be in the setup, but only one thing required changing and that was the pointing sequence on the videos. Before the real experiment, it was necessary to investigate the duration for each pointing gesture that was pointing to the numbers at the wall. Multiple tests were held to find out a duration for each gesture pointing in the experiment. The pointing duration required to have enough time for the participants to detect the gesture and determine the target. To be sure, that everybody had enough time to react on the pointing in each test and technique, multiple tests were carried out, and as a result, 10-12 seconds were found to be most efficient duration for each pointing gesture.

An intro clip was created for the test to instruct the participants in what was going to happen. The intro clip explained the general idea of the test and what the participants should be doing. Making the final clip required a video-editing software to compile the final videos. In this case, Adobe Premiere Pro was used to combine and cut the videos in the right order. First, adding the intro video, and then the final clip with a transition effect between the two videos.

## 3.2 Unity

Using Unity for the project's purpose was the most ideal choice when knowing what hardware was going to be used. Both devices, Samsung S2 and Toshiba Excite 7.7 tablet, had Android version 4.1.1 which meant that they were not supported by the latest 360-degree video and the Google cardboard applications. The devices given for the experiment were outdated and could not be upgraded to the newer operating system versions. Therefore, it was necessary to build and create the application mainly for these two devices. Choosing the Unity game engine as the development platform for the project was due to the easy use and previous knowledge in that engine.

The first thing that needed to be created for the application was the sphere object. This sphere object was going to have the equirectangular video as a texture. In order to make the object work for this purpose, creation of the UV map (Figure 3.4 (b)) on the object was required. Doing this, allowed the Unity to add an equirectangular video as a UV map texture on the created sphere object. This texture feature is called "movie texture" in Unity and it works as a normal bitmap texture but instead of a picture, it is a video. The UV texture includes the information, where in the sphere, each pixel is supposed to be located at. In Unity, the camera needed to be in the centre of the sphere so the camera was surrounded by the sphere. To see the equirectangular video on the surface of the sphere, the normals of the sphere object needed to be inverted. Due to the inverted normals, the camera was capable of seeing the texture inside of the sphere and have a 360-degree video environment, see Figure 3.4 (a).

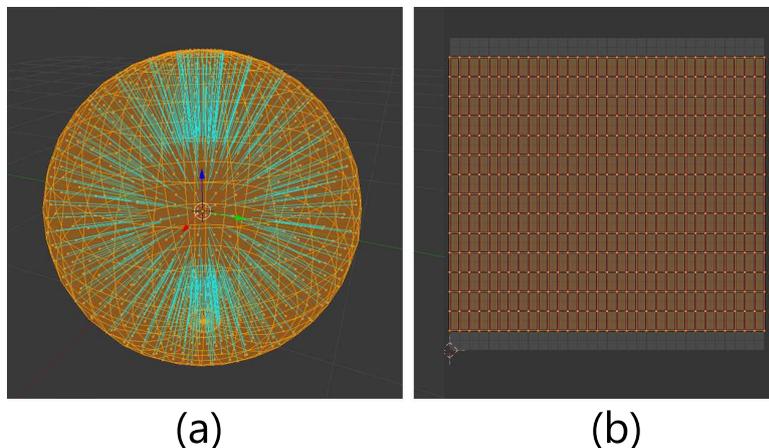


Figure 3.4: (a) This is the sphere created in Blender, the blue lines represent the direction of the polygon normals and in this case the normals of the sphere has been inverted to point inside so the video will be playing towards the center point of the sphere. (b) UV map is required in order to attach the movie texture correctly on the sphere.

In order to start the video on the movie texture, it required a script that makes video play when starting the application. Writing the following script from the Unity manual "`((MovieTexture) GetComponent<Renderer>().material.mainTexture).Play();`" was required to start the video that is attached to an object as texture in the Unity Engine. [25] By creating a C# script and

adding the script mentioned to the start function, made the movie texture work. The application started working in the Windows' testing mode, and demonstrated how the application would work in general. However, the next step was to test this new application on the Android devices. Due to the Unity's own coding, Unity does not support movie textures for android yet so the application did not work.

### 3.2.1 Plugin

An alternative approach was necessary for the applications because the movie textures did not work on the Android devices. Unity's third party plugin called "Easy Movie Texture" [8] provided an alternative solution to the problem. The plugin uses the same logic as before by wrapping a video texture around a sphere with surface normals pointing inwards, but instead of using Unity's own video codec players, the plugin makes the application use the Androids own media player to play the video texture on the sphere surface and therefore, creating 360-degree video playable on the Android devices. The plugin featured one negative aspect, the play mode in Unity did not show or play the video texture. This was because of the plugin script that played the video texture with the Android's own media player. Therefore, it was not capable of playing the videos with Unity's media codecs.

There was trouble with the video from that Ricoh Theta S software converts. The video that is produced from the software is in a format of .MP4 and when this video was tested in the Unity with the new plugin, it did not run. For a unknown reason, the video from the Ricoh Theta software needed to be converted again to .mp4 format.

The package includes a demo scene for all of the modes it is capable of running which includes streaming, spherical 360-degree video, and normal texture plane. Alongside of these scenes, it brings a menu GUI to control the videos on the mobile application. However this GUI was not necessary for the purpose of the projects and was required to be deleted as it was covering a lot of screen and disturbing the experience. The sample scene "Demo\_sphere" is connected to a script called "MedaiPlayerSampleSphereGUI.cs" which creates a GUI media control panel for the app. The GUI script is attached to the scene camera in order to remove the script from interfering with the application screen. The script could be simply be unattached from the scene camera by using the "Remove Component" option.

Media Player Ctrl is the script attached to the sphere object, that controls which video the application would be playing. As can be seen in the Figure 3.5, the "Str File Name" is the section that determines which video will be played. However, the script is looking for that file name from a specific location called "StreamingAssets" folder. During the development of the applications it occurred that depending how many clips there was in the folder, Unity included all the clips to the application even though, the intention was to include only one. This resulted in a huge application file sizes which the applied mobile devices could not handle. Furthermore, this was the reason for the media player GUI buttons in the demo scene. However, by deleting other clips from the asset folder, Unity allowed to build applications with a much smaller file size.

Testing the applications with the working videos resulted in a new problem as the tablet, used

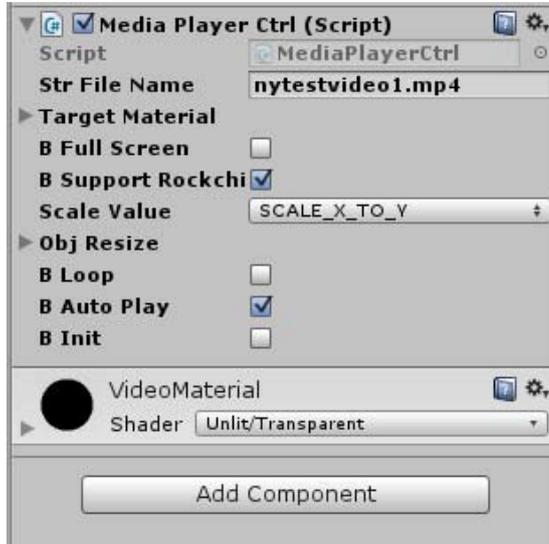


Figure 3.5: The picture shows what options the Media Player Ctrl script provides for the developers. "Str file name" is the section which determines what video the application will be having.

for experiment, did not provide any sound while the Samsung S2 phone had sound in the videos. This was resolved by converting sound from the videos and then be added in Unity as an audio source. The audio gathered from the videos were in a Wave sound format. Somehow, the audio did not work in one of the devices but making an external audio source in the Unity engine seemed to be the easiest fix for the problem.

### 3.3 Interaction Techniques

Interaction techniques are techniques that provide a way to interact with a mobile device. The following interaction techniques are used for the experiment: click and drag, IMU-based rotation, and virtual reality. These interaction techniques were chosen as they are the most popular techniques of the mobile devices in the consumer market.

#### 3.3.1 Click and Drag

The click and drag interaction was chosen for the experiment as it is the most common interaction technique in mobile devices. This is exactly the same interaction technique that can be found in YouTube's 360-degree cameras if a person does not have a device that supports gyroscopic rotation. It requires the user to use an index finger or any other finger to navigate through the operating system. To interact, the user is just required to tap, swipe, or drag the finger on the screen of the device to navigate or use an application, see Figure 3.6.

The goal of the application is to mimic the YouTube's 360-degree click and drag interaction



Figure 3.6: Click and drag interaction technique is used by swiping and moving a finger on the screen. Application calculates the movement with the finger and rotates the camera accordingly.

technique to the point that it is fairly similar to it. However, building this application required the "Easy Movie Texture" plugin from Section 3.2.1 to work properly. As earlier mentioned it required the 3D sphere with the UV map for the spherical texture purpose, in this case it was easier to use the demo scene sphere as it was provided ready with the plugin. However, the most important part lies in the script that utilizes the touch screen dragging for the first application.

The dragging effect is the effect that makes a user feel like that they are dragging the video, or the content, in an application by swiping a finger/mouse(input technique) on a screen that is suitable for a certain interaction technique. In this case, the input technique is the finger on a touch screen device. Figure 3.7 demonstrates how the dragging works in the click and drag technique. When the input finger is swiped equivalent to the length of the red arrow on the screen, inside Unity, the camera in the center of the sphere rotates to an inverse direction to a certain length depending on the sensitivity set in the script for this movement.

For the click and drag interaction, it required a script that allowed a similar effect as mouselook with inverted directions. However the script used first ended up being too inaccurate so it was required to look for a new script fast. The script by Andre Odendaal [19] was found to be all that was needed for the application and it fixes all the errors that existed with the old script. The script provided a few settings that could be edited to the developers liking, like sensitivity of Y and X axis and if the navigation was inverse on X and Y axis, as seen in the picture below. The picture also states the settings used for the application. With inverted Y axis navigation the sensitivity for X axis is 20 and for Y is 10. These settings were done by trial and error until they quality of the settings were in a satisfactory level. The application settings were tested on few people and tried to mimic the Youtube's click and drag feeling.

While making this application, an annoying problem occurred as the new version, Unity 5.0, included in hidden to its new engine for Android devices. This immersive mode when enabled

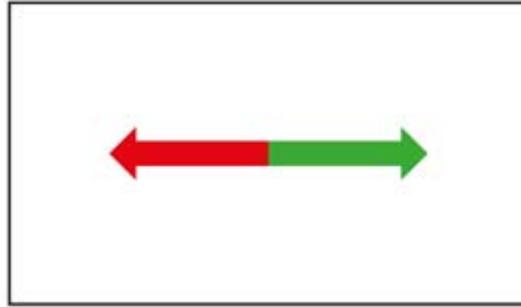


Figure 3.7: Example how the click and drag works. Red arrow indicating the direction of finger swipe and green arrow indicate which direction the content on the input device move when the red arrow interaction occurs.

automatically hides the navigation bar from Android so the application can utilize more screen space for its purposes. However, in this case it caused a problem that was rather hard to fix without a lot of knowledge in this area. While building the application, the problem surfaced when the tap on the screen happened and especially when a person wanted to drag a finger through the screen to rotate around in the 360-video. However, the first input worked in the beginning with dragging but when the input happened the input brings the navigation bar from the immersive mode to screen which is only registers that input for bringing the navigation bar and does not allow to rotate in the video. It required a second tap right after bringing that menu to navigate again, which basically meant that every second click/tap on the screen worked properly on the screen.

```
<meta-data android:name="IMMERSIVE_MODE" android:value="false" />
```

Figure 3.8: The line that was causing problems on the click and drag application. Problem was resolved when immersive mode was set to "false".

In order to fix this quickly, the immersive mode needed to be disabled so the application can work the experiment as intended. The immersive mode settings were found in the Android-Manifest.xml inside of the unity's assets folder, and there you can find a row that states if the immersive mode is enabled, see Figure 3.8.

### 3.3.2 IMU-based Rotation

The IMU-based rotation is an inertial measurement unit based rotation. This piece of technology has become a very important and well-known technology in a consumer market. It is sensory technology that makes it possible to movement on devices. According to Kai Oberländer [18], IMU combined with magnetometer, is capable of determining the orientation of the device from the gravity and earth's magnetic field. Therefore, the technology can measure the orientation and rotational movement of the device.

The IMU-based rotation is quite new and unknown in the consumer market. However, many

games and applications use the built-in gyroscope for additional features inside the application like racing with a phone. Furthermore, YouTube has recently added a feature that allows users to navigate in 360-degree videos by rotating the mobile device around. The gyroscope inside of the device reacts to this movement accordingly and makes the users feel like they are actually inside the 360-degree video. Instead of being stereoscopic view it is a mono view with only one camera in the Unity.

A perfect example of how the application works is a mobile device's camera mode. It resembles a lot when a user applies a camera in an environment. However, in this case, it allows feedback from another 360-degree environment. The person can rotate and move around with the tablet in the environment as seen in the Figure 3.9. When the person rotates the tablet upwards, the application calculates the movement done with the tablet, and rotates the Unity camera accordingly. This to the same distance as the upwards movement did. Therefore, creating a "window" to a 360-degree video.



Figure 3.9: IMU-based rotation, and how the orientation and rotation is calculated with this technique. Tablet sensors calculate the relative movement done in the application and moves the camera in the application accordingly.

While building the application, a simple way was found to create the IMU-based rotation. The Google Cardboard script has a row that actually determines if the Virtual reality mode is enabled. This includes the stereoscopic view and other features that the VR mode is using. When changing the virtual reality mode to "false" the application technically becomes the IMU-based rotation method. The script seen in Figure 3.10 was changed in the script "Cardboard.cs" which is included in the Google Cardboard Unity development package. This action changed the application from virtual reality stereo-view to mono-view.

```
private bool vrModeEnabled = false;
```

Figure 3.10: The script that changes the virtual reality stereo-view to mono-view.

### 3.3.3 virtual reality

Googles own approach to budget virtual reality with smartphone devices is Google Cardboard which allows smartphone users to use their phones as a virtual reality headset. The headset's original design is a low budget cardboard box with cheap components inside but as the Google has shared to schematics of the headset to public, countless third party versions for the Google cardboard exist. Google Cardboard uses the gyroscope inside the mobile device to rotate accurately when the mobile device is rotated. Basically, it calculates the orientation and rotation so the developer can use that data for the movement required applications like virtual reality (see the picture in Figure 3.11). In other words, Google Cardboard technology uses the same IMU-based technology to rotate around in the virtual environments.

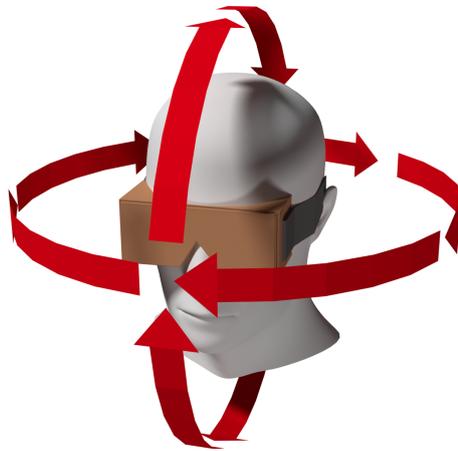


Figure 3.11: Example how the mobile device's gyroscope calculates orientation and rotation, and how free the head movement can be with this interaction technique.

Building the application was simple, due to the Google's ready made Unity package. The Google Cardboard Unity package included everything to make the Google Cardboard virtual reality working. It provided all the scripts necessary for the application. However, building the application, again required the earlier mentioned sphere object and the "Easy Movie Texture" Unity plugin. The Google Cardboard camera prefab from the Google Unity package needed to be positioned in the center of the sphere object which was included in the demo scene of the plugin or self created.

Google Cardboard also features stereo rendering which uses the cameras in Unity for each eye. In Unity, the distances between these cameras can be set to the developer's preference but in this case, the distance was left to default settings. Google Cardboard automatically tries to define the Android screen size to adjust the eyes for the mobile device. Unity allows users to preview the virtual reality application with different device settings, Google Cardboard Unity package included the most popular devices with their setup for the preview mode. However, there were no setups specifically made for the Samsung S2 so it was let to be on the default mode.

## 4 | Limitations of Design

During the project, a few problems and limitations on hardware occurred. Especially with the mobile devices and their old Android operating system versions. Both of them, smartphone (Samsung S2) and tablet (Toshiba Excite 7.7) had Android version 4.1.1 which is the lowest version that Google Cardboard SDK support on android device. This means that some features are not available for the devices due to their old firmware.

By receiving a lot of feedback about the video and screen quality, it is important to issue it here. Due to being old devices for especially virtual reality purposes, the quality of the screen might not have been the best as its resolution was 800x480 pixels. Along with the low resolution screen, the resolution of the camera was only 1080p which means, that the whole sphere object around you is in 1080p quality. Thus, when the user is experiencing 360-degree video, they are only seeing a small portion of the 1080p video in their field of view, and therefore, for the users, the quality is much lower than the 1080p. If the screen quality was bad on the mobile device, then it does not matter if the quality of the video was good because the bad screen resolution would lower the quality of the video as well. However, the participants also gave feedback that the video quality was not the greatest on the tablet either. This means that in this section there is room for some improvements. However, this study is still valuable as it provides important information on deictic pointing.

360-degree video viewing could have been done in YouTube or any 360-degree consumer market streamer but experiment's Android devices did not support these applications. This resulted on the Unity approach and building the applications for the devices. The applications were not as feature rich as the consumer streaming applications are.

During the end of the experiment, a few hiccups occurred on the IMU-based rotation application. During the experiment, participants are required to rotate around with a chair to navigate in the 360-degree video. However, even when participants were holding the tablet stable without moving at all, the tablet registered movement from somewhere. Reason for this could not be verified but chance on gyroscope failing during experiment is quite possible. Restarting the application made the application work properly again and the error did not occur after that. Error happened only few times randomly before the actual testing started and therefore the experiment did not lose any data due to this error. In some cases the gyro failed and the image started falling on its own. Basically this meant, that gyroscope has a slight drift that could have affected the results, but serious draft only occurred few times.



## 5 | Experiment

As mentioned in the Motivation in Chapter 2, this is a subject that has not been investigated much. This experiment's purpose is to see if there are performance differences in detecting the deictic pointing directed outside of the user's field of vision in 360-degree videos.

The aim of the project was to have four different tests in the experiment where a participant was tested in each test. The four tests are, click and drag, IMU-based rotation, a virtual reality and a real world test. The test would follow within subject design. The tests had three different applications that used different interaction techniques for the 360-degree videos. Along with these three applications, a real world test was held first as a comparison to the other three applications. The experiment had a group of 24 participants that said out loud which number the host was pointing at. After the test, a questionnaire with demographic questions was answered. The gathered data provided information of how the test results were different from each other. The logic behind the experiments was that if there was no significant difference between the three applications, then it meant, that they were equally similar to each other. Thus, detecting a deictic pointing towards outside of the field vision in 360-degree videos in any of these three interaction methods did not have an advantage.

	Number Sequences									
<i>Video1</i>	5	9	12	6	8	3	8	4	1	10
<i>Video2</i>	11	5	7	3	10	5	8	2	4	12
<i>Video3</i>	2	7	11	6	9	7	1	4	8	9
<i>Video4</i>	8	2	10	8	6	4	1	3	10	5

Figure 5.1: Numbers used for each number sequences.

The experiment design had a test order that rotates with multiple number sequences, see Figures 5.1 and 5.2. For example, if participant A experienced the test order 1 "TO1", seen in Figure 5.2, then the participant B in the next test had the test order 2 "TO2". These test orders determine which number sequence is to be shown for the participant in each test. For example, test order 3 "TO3" had number sequences in the following order: sequence 3, sequence 4, sequence 1 and

sequence 2. The reason for number sequence rotation was because some number sequences could experience more bias from users than other number sequences. For example, individual could detect the deictic pointing much better in number sequence 1 than the sequence 2. Therefore it was necessary that all four tests had four different number sequences rotating around so they had a more diverse results. Basically, in the experiment the user experienced a different number sequence in every test. The next participant then experienced the second test order where all the same number sequences were as well, but in different tests. This was done so the the all tests had a equal amount of results from all the number sequences and therefore, avoided the problem of number bias.

Test order for participants

TO1	Video1	CND2	RN3	VR4
TO2	Video2	CND3	RN4	VR1
TO3	Video3	CND4	RN1	VR2
TO4	Video4	CND1	RN2	VR3

Figure 5.2: This picture demonstrates the test order used in the filming and the experiment.

## 5.1 Hypothesis

The research will be focusing on the three interaction techniques, click and drag, IMU-based rotation, and virtual reality. The project will compare these methods to the real world situation results to investigate if gesture pointing towards non-seen-target is reacted and perceived differently in these three interaction methods. The following null hypothesis is:

There are no differences between the real world and the three interaction techniques, click and drag, IMU-based rotation, and virtual reality.

## 5.2 Participants

In total, 24 Aalborg citizens were used as test subjects in a within subject experiment design. However, a majority of the participants were students from Aalborg University. With ages ranging from 20 to 38 and 8 out of those 24 participants were female. Participants were informed, that the experiment was about 360-degree videos and gesture pointing. The entire experiment was carried out in English.

### 5.3 Equipment

A limited amount of equipment for the implementation and experiment was applied. Due high demand of up to date devices in storages of Aalborg University, older devices had to be used for the experiment. However, all of them supported the interaction technique implemented for the experiment. The camera was Ricoh Theta S with a normal tripod which was only needed for the filming part of the project. However, the phone for the experiment was Samsung S2, it is a dated phone but it was capable of running these application if the applications were made for that specific android version. For the click and drag and IMU-based rotation, the tablet was acquired, and it had the same problems as the phone which was the old Android version. A list of the equipment is listed in the Table 5.1 and can be seen in Figure 5.3



Figure 5.3: The equipment used for the experiment. Durovis Dive with the Samsung S2 in the headset, and next to headset is the Toshiba 7.7 tablet.

To hold the Samsung S2 as a virtual reality tool, a Durovis Dive [4] headset was acquired from the University. As seen in Figure 5.3, the headset allows users to adjust the focus lenses for the device to their liking. The Google’s Cardboard version does not include this feature, therefore, the Dive headset was a little more advanced. Basically, Durovis Dive is a high quality version of the Google Cardboard and provides good a virtual reality experience if the device is right for it. The headset is not made specifically for a certain phone and therefore, multiple mobile device can fit in it.

Equipment	Model
Camera	Ricoh Theta S with tripod
Tablet	Toshiba Excite 7.7
Phone	Samsung Galaxy S2
Virtual reality headset	Durovis Dive

Table 5.1: Tools required for the experiment and the implementation.

## 5.4 Experiment Apparatus

The experiment was held in one of Aalborg University's classrooms, see Figure 5.4(a). The setup included a table with a chair, a chair between the table and the wall, and numbers going from 1 to 12 on the wall and the mobile devices; The tablet and the smartphone, including headphones and the Durovis Dive Virtual reality headset (can be seen in the Table 5.1 on page 25).

Everything done for the filming in the project also needed to be identical in the experiment itself. This included the position of the host, the camera, and numbers at the wall. During the filming of the videos for the applications the host was using the right main hand for pointing.

It was also rather important to do a precise replica of the videos on each test so it was important in the real world test to remember to point at the numbers with the same hand (right). This was due to the chance of participants to perceive it differently and cause a bias on the dominant handedness and therefore, participants would react differently if the hand was changed during the test to the other one. It was also to keep the host more focused on pointing, and not to think about what hand to use.

During the experiment, it was important to keep up with test numbers for each participant as every fourth participant had the same experiment test order. As mentioned earlier, to reduce the number of bias, multiple videos were made to cycle through the applications and therefore, every fourth test the participant had the same test order as demonstrated in the figure 5.2. An application was made for the each of the number sequences and for each of the interaction techniques. That resulted in 12 different applications for the devices, and therefore, in each test, the host needed to be very careful which application was required for each of the participants.

The classroom needed to be the same one as the environment where the videos were made, but unfortunately the room(3.529) was occupied during the experiments. Luckily, Aalborg University had multiple identical classrooms on the same spot in building in each floor and therefore the test could be done in the identical room (3.429). Although the room had a different order of tables, participants did not notice the difference. Furthermore, the environment was a open room where there is enough space on the wall for the numbers to put there. The numbers were in order from 1 to 12 and were the size of a A5 paper and having 15cm space between them on the wall as seen in the picture below. The number papers were 158 cm from the floor and participants position was centered in between number 6 and 7, see Figure 5.4 (b).

The setup was tested multiple times to make sure that the camera, wall and the host was in comfortable distance from each other. The safe distance between host and the wall was three meters and the camera/participants sitting in the middle on the point of 1.5 meters from the wall and the host. This setup of the filming was necessary to remember because in the experiment the real world test is part of the experiment and requires the same identical environment from the video to make results that can be compared to the other three tests, see Figure 5.4(a).

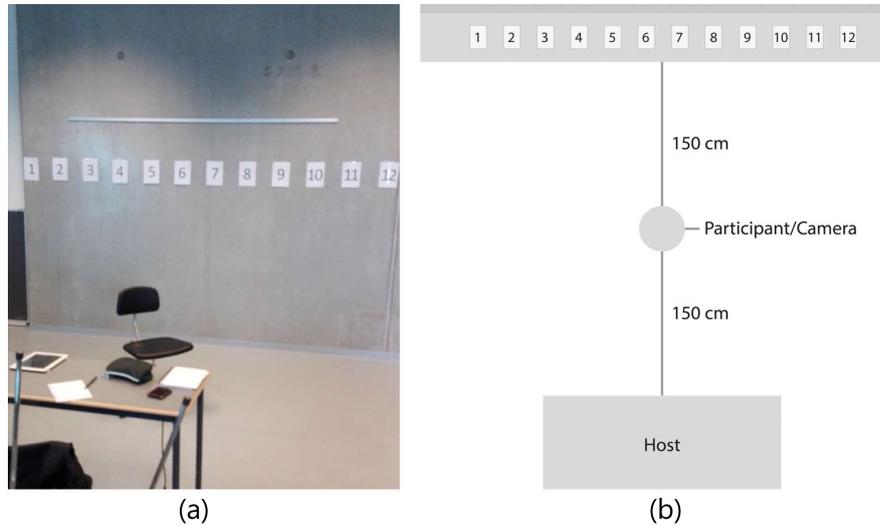


Figure 5.4: (a) A picture from the experiment showing how it looked (b) This sketch demonstrates how the setup was in the classroom. This setup was the same for the filming and the experiment, only the camera changes places with the participant.

## 5.5 Procedure

Each of the participants were guided to the classroom where the experiment was happening (Aalborg University’s classroom in the Create building). The participants were instructed to sit down on a chair and not move away from the spot(piece of tape) marked on the floor and were ordered stay in that position during the whole experiment. Then, the participant was to face towards the host so they quickly could hear what the experiment was about. The participants were instructed as follows;

“This test is about gesture pointing. The test will be a 4-part test where we start with a real world test and from there we will continue with other electronic devices followed by a similar test, like the first one. Behind you, you can see numbers on the wall. I will be pointing at these numbers and you are allowed to move with the chair to see and figure out where I am pointing at. Please say out loud which number I am pointing at each time. There is a chance, that I will be pointing at a number more than once during the test. Are there any questions?”

The instruction was given, to make sure that the participants were sure what to do and what the test was about. After explaining how the experiment works, the testing started. During testing, each number the participant said out loud, was written in a notebook. After the first test, we continued towards the second test which included the same kind of instruction. However, this time they were told that; “in order to move the camera around in the 360-degree video, touch the screen and drag the finger around” while putting the headphones on. When the test starts again, they were instructed to say the numbers in the same way as in the last test.

Right after the second test, the experiment continued with a third test. Here the participant was told the following; “instead of using the finger, this time you are going to rotate around

with tablet. You can use your hands to rotate and move the tablet to see more of the 360-degree video. You can use the chair to rotate around more easily.” The participants again got the tablet with an other application that was meant for the IMU-based rotation and again the test could be started. The same procedure was followed again. After the last number, new instructions were given; “Now to the last part of the experiment, which includes virtual reality”. The last part of the experiment included a virtual reality on a Samsung S2 mobile device. With the help of the host, the headset was easily set on the participant. To start the test, an adjustment of the headset was necessary, so it fit the participant’s needs. The lenses in the headset were adjustable which allowed the users to set the lens focus for their own liking. After adjusting the headset to be comfortable on the head, and it was ensured that the video had the best possible quality (no blurriness), and then the test was ready to start. By manually starting a new video in the mobile phone, the last part of the experiment started. Doing the same procedure again with the virtual reality device and headphones on, it was yet again asked for those ten numbers for the data.

During the test of the three interaction techniques, the participants were asked only to focus on what happened in the application of the three devices. In that way, they did not have to pay attention to the host sitting at the table and writing the numbers down. After the tests, the participants were asked to fill out a quick questionnaire with quick a instruction of what to do. When the participants finally finished the questionnaire, the participants were free to leave but sometimes, they wanted to start a small conversation and give feedback on the experiment.

## 5.6 Questionnaire

The questionnaire had questions asking the participant to rank the quality of pointing and his or her preferred 360-degree video interaction method. The ranking system consisted of numbers to rank the quality of the pointing in each test. The ranking questions were a Likert scale where number 1 meant that the quality of pointing was hardest and 5 was the easiest. These results can provide information about differences between the questionnaire and the test results. Assumptions and comparisons of statistics, gathered from the questionnaires, could provide information that would assist in understanding the results gathered from the experiment. The questionnaire also asked other questions as well but due to a misunderstanding of some of the questions, the data was not usable for the project. To see all the questions from the questionnaire. (Appendix, p:47)

## 6 | Results

The gathered data had 10 results from each tested participant. However, one of the participant's had a missing value in one of the tests thus, a row of data from each test needed to be removed. Therefore, instead of 240 values for each test, there was now 239 values. Afterwards, the averages of each participant's results were calculated so the comparison of the data would be simpler. Instead of the original 239 values in each test, each test now had 24 values. However, one of the participant's data has one value less (9) than the rest of the participants' (10) due to a missing value during the tests. Therefore, one of the test values out of 24, has 9 values instead of 10. This data was not left out as it still provided data equally to all tests. Therefore, the data all the tests received, had one value that was an average of 9 values instead of 10.

Since there are multiple groups in this data, it was required to test for differences in the data. In order to see if there was any significant difference in the data, the Friedman's test was used. The results of The Friedman's test revealed that there was a significant difference in the data with the p-value of  $7.118e-07$  which was below the test's significance level of 0.05, seen in Table 6.1 (b). With help of the Friedman's test, it was known that there was a group, or maybe groups, that were significantly different from each other. Therefore, further comparison between the tests was required in order identify which test or tests caused this p-value.

To identify which groups in the data were responsible for significance difference, a post hoc analysis was performed on the results of the experiment. In order to find the tests with significant differences, all the tests needed to do a Mann-Whitney's U test between all the other tests. Therefore, the number of comparisons between the tests was six, see the Table 6.2.

Tests	Mean ranks	Methods	Value
Real world test	0.411574	Friedman's test	$7.118e-07$
Click and drag	0.765278	Bonferonni correction	0.00833
IMU-based rotation	0.794907		
Virtual reality	0.769907		

(a) (b)

Table 6.1: (a) Table holds the mean rank values of each experiment test results. (b) Friedman's test p-value and the new significance from Bonferonni correction.

Doing the Mann-Whitney's U test showed that there were at least three of the comparisons

that were significantly different. In order to identify which of the tests in the comparisons were significantly different, it was required to have mean ranks of each of the tests, see Table 6.1. With mean ranks the difference of the mean ranks were calculated. The real world test compared to all three interaction techniques was significantly different with p-values of 8.782e-06 for click and drag, 4.259e-06 for IMU-based rotation and 1.125e-05 for virtual reality. As seen in Table 6.1 (a) and Table 6.2, the mean rank of the real world is much smaller than the mean ranks of the three techniques. Therefore, the difference of the mean ranks indicated that the real world test was significantly better than the three interaction techniques. Since, this was a test that compared multiple groups, it will increase the likelihood of receiving significant results and therefore, some of the significant results could be a result of pure chance. Therefore, the Bonferroni correction was done to ensure that the Mann-Whitney's U test did not cause "false" significant values. The Bonferroni correction divides the originally used (0.05) significance level by a number of comparisons made in the test. Basically, as seen in Table 6.2, there are 6 comparisons in total, and therefore, the significance value of 0.05 is to be divided by 6. The Bonferroni correction resulted in a new significance level of 0.00833 which is seen in Table 6.1 (b). Even though, the Bonferroni correction was done for the results, the results remained the same and the real world test compared to the three interaction techniques remained significantly different. All the p-values of the real world tests were below the significance level of 0.00833.

Compared techniques	Difference of mean ranks	p-value
Real world test & Click and drag	-0.354	<b>(8.782e-06)</b>
Real world test & IMU-based rotation	-0.383	<b>(4.259e-06)</b>
Real world test & Virtual Reality	-0.358	<b>(1.125e-05)</b>
Click and Drag & IMU-based rotation	-0.030	(0.9917)
Click and Drag & Virtual Reality	-0.005	(0.6930)
IMU-based rotation & Virtual Reality	0.025	(0.6777)

Table 6.2: Table with the difference of mean errors and p-values of the tests after Mann-Whitney tests. The difference of mean errors is negative if the second test received a higher (i.e. worse) mean error.

The results from the test showed, that there was no significant difference between the three interaction methods. However, there seems to be a difference between these three methods when compared to the real world test. As seen in Table 6.2, the results are close to each other's p-values between the three interaction techniques. As seen in Table 6.2, the p-values between all of the three interaction techniques were above the value of 0.67 which indicated that they were not significantly different from each other. However, the results show that they were slightly different from each other. The click and drag and the IMU-based rotation technique received a p-value of 0.9917 which indicated that these two techniques were close to each other. However, the difference of the mean ranks in Table 6.2 showed, the value of -0.030 which proved that the click and drag received better results than the IMU-based rotation. However, virtual reality seemed to be most different from the other two interaction techniques as it showed a p-value of 0.6930 for click and drag, and 0.6777 for the IMU-based rotation. However, the differences in the mean ranks tell that with a difference value of -0.005, the click and drag received slightly better results than virtual reality. Furthermore, the differences of mean ranks

between IMU-based rotation and virtual reality showed a value of 0.025. This indicated, that virtual reality received better results than the IMU-based rotation.



Figure 6.1: This histogram demonstrates the distribution of the results in each test and platform. Errors indicate the difference from the correct number. For example Error +2 means that the amount of the bars distribution had an error 2 numbers from the correct one.)

To further understand the results, a diagram was made to show the errors of each test in total. It showed how many correct numbers the participants got, and how many of the numbers were incorrect of 1 or more numbers from the original number, see Figure 6.1. In the histogram, it is clear that the click and drag and virtual reality techniques were very close to each other but they seem to have a little difference from the IMU-based rotation at it has more correct numbers but also numbers that are further away from the correct one than other methods. For example, on rotational there is a slight tab error of +4 from the correct number. However, the real world test information on the histogram, confirmed the results of the Mann Whitney U test as it demonstrated a clear difference from all the other three methods.

These results indicated, that the real world results were significantly different from all three interaction techniques; the click and drag, the IMU-based rotation, and the virtual reality. Therefore, the results from the main experiment rejects the null hypothesis. The three interaction techniques are very similar to each other and do not differ significantly.

## 6.1 Results from The Questionnaires

Along with the results of the actual test, there were also the results from the questionnaire which revolved mostly around the test participant's preferences and opinions than actual solid data that was recorded from the tests. However, the questionnaire results holds interesting data that is important to discuss here. As seen in the chart on figure 6.2, there are differences in the data when the questionnaire data are compared to the gathered data from the tests.

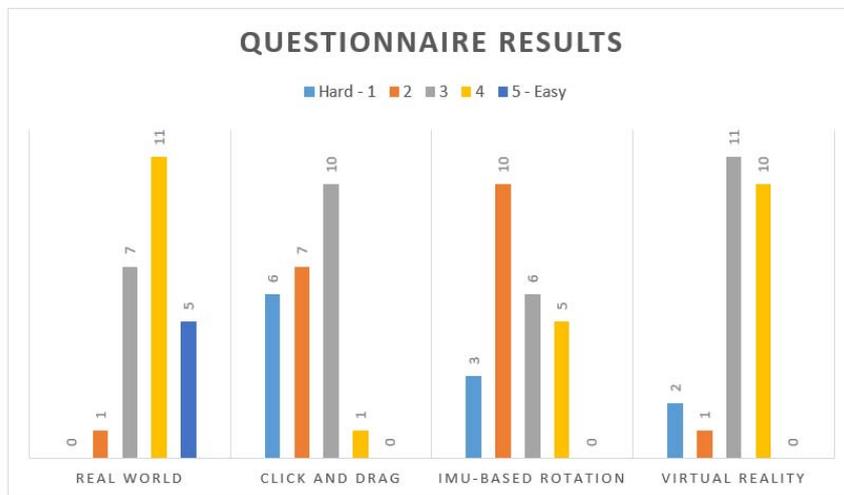


Figure 6.2: a chart for quality of pointing asked in the questionnaire. The ranking system works by number 1 being the hardest and number 5 the easiest.

Questions 5 to 8 (See Appendices) ask questions about the quality of the pointing in each test. Each test was graded with the 5 step Likert scale where number 1 was hard and 5 meant easy. Using this data, the chart of Figure 6.2 was made and when quickly looked, it showed that the virtual reality seems to have better results than the click and drag and the IMU-based rotation. The results of the questionnaire caused an interesting situation as it conflicts with the results from the experiment. The experiment results indicated, that the three interaction techniques, click and drag, IMU-based rotation and virtual reality, did not differ significantly from each other. However, there was a significant difference between these three tests in the real world test. According to the chart, the virtual reality test results from questionnaire seems to be much closer to the real world test results than the click and drag and the IMU-based rotation. In the chart, the real world seems to be the only test that was rated to be easy with grade 5, and therefore, meaning that the other three were slightly harder for the participants. Especially, click and drag and IMU-based rotation showed a great amount of responses on grades 2 and 1. This meant, that the quality of the pointing was hard. However, the click and drag and the IMU-based rotation had conflict as the click and drag demonstrated much better results in grade 3, but then has much less answers on the 4th grade in Figure 6.2. Some participants seemed to perceive these tests differently, maybe it is because of the previous experience in these interaction techniques. It could also simply be because of different user preferences.

The data had 24 values in each experiment test, and therefore, a similar approach can be used to investigate the differences between the test results. To investigate if there was any significant differences in the data, the Friedman's test was again necessary. The Friedman's test proved, that there is a significant difference in the data with a p-value of 3.56e-08a, seen in Table 6.3(b). Therefore, further investigation and analysis of the data was required. In order to find the tests that caused the significant difference, all the tests needed to do a Mann-Whitney's U test between all the other tests. Afterwards, the Bonferroni correction was required once more.

Tests	Mean ranks	Methods	Value
Real world test	3.84	Friedman's test	3.56e-08
Click and drag	2.24	Bonferroni correction	0.00833
IMU-based rotation	2.48		
Virtual reality	3.2		

(a)
(b)

Table 6.3: (a) Table holds the mean rank values of each experiment test results. (b) Friedman's test p-value and the new significance from Bonferonni correction.

With the same amount of comparisons (6), the corrected significance level was 0.00833. This is seen in Table 6.3 (b).

Differences of the four test (Questionnaire)		
Materials	Difference of mean ranks	P-value
Real world test & Click and drag	1.6	<b>(5.835-07)</b>
Real world test & IMU-based rotation	1.36	<b>(2.42e-05)</b>
Real world test & Virtual Reality	0.64	(0.01391)
Click and Drag & IMU-based rotation	-0.24	(0.4706)
Click and Drag & Virtual Reality	-0.96	<b>(0.0003056)</b>
IMU-based rotation & Virtual Reality	-0.72	(0.008855)

Table 6.4: Test results from the questionnaire. Table with the difference of mean errors and p-values of the tests after Mann-Whitney tests. Differences of the mean error is negative if the mean rank of test on the right side on table is higher

The Mann-Whitney's U test right away showed, that there were significant differences in the results. The results showed, that the Click and drag compared to other three experiment tests, was significantly different – except when compared with the IMU-based rotation.

However, with the Bonferroni correction, the significance value was 0.00833, which means that three of the comparisons were not significantly different and were not below the significance value, seen on Table 6.4. The real world and virtual reality test results received a p-value of 0.01391 which indicated that their value was not above the significance value of 0.00833 and did not have any significant difference. However, the IMU-based rotation and the virtual reality received a p-value of 0.008855 which was above the significance level of 0.00833, and therefore, the data did not differ significantly from the results. However, the Bonferroni correction was a very conservative method to protect the chance of random significant values emerging from the data. Therefore, if the significance correction was done with more advanced methods, the significance level could have been on a level where the IMU-based rotation and the virtual reality could have been significantly different from the results.

However, the test also had results that were significantly different from each other. The real world test compared to the click and drag and the IMU-based rotation received p-values of 5.835-07(Click and drag) and 2.42e-05(IMU-based rotation). The higher the numbers in questionnaire mean ranks were, the better the results were. The differences of the mean ranks for the click

and drag was 1.6 and for the IMU-based rotation it was 1.36. This proved, that the real world test was significantly better than the click and drag and the IMU-based interaction techniques, see Table 6.4 and Table 6.3 (a). Furthermore, the p-value of the click and drag and the virtual reality stated that there was a significant difference between them. With a mean ranks of 2.24 and 3.2, the mean rank difference was -0,96 which indicated that the virtual reality was, in the participants' opinion, the easiest platform to perceive the deictic pointing.

It is interesting, that the participants' opinion of the test as well as the interaction techniques were completely different from the actual results from the test. The participants graded the virtual reality to be easiest to use for experiencing 360-degree videos. However, the virtual reality results show, that it is significantly different from the click and drag and slightly different from the IMU-based rotation. Furthermore, according to the participants' opinions, the virtual reality was comparable to the real world results. Therefore, it is safe to say the participants prefer virtual reality, followed by the IMU-based rotation, and the click and drag. The IMU-based rotation and the click and drag test results were very close to each other but the IMU-based rotation was slightly better.

## 6.2 Observations

The observations made during the experiment indicate that the users had some problems with the hardware, especially, the virtual reality. The users complained about the low video quality of the 360-degree video which could be the camera's or the mobile device's fault. A low resolution screen combined with a low resolution video might lower the immersion of the virtual reality – as the experiment could have been carried out with better equipment and obtained a better result. Even though, the experience was not the best for most of the participants, the majority of the participants still chose virtual reality as their preferred interaction technique for experiencing 360-degree videos, in the questionnaire.

The 360-degree video as well had a slight mistake which I did not notice myself. After hearing some feedback from the participants, it became evident, that there was a disturbing effect in the video. It occurred during the experiment that the host in a couple of the videos changed places in the middle of the experiment after the instructions clip. The reason for this to happen, is due to the orientation of the camera itself. While filming, I did not have any knowledge of this or that it was important to keep the same orientation of the camera. However, during the experiment some participants started noting that it was a little weird that the host in the video was in a different location than in the real world. This caused some confusion on the participants, but as the experiment went on, they easily adapted to it. See the image in Figure 6.3 to see the difference between the orientations, the left picture is when the camera's front with the record button is facing the opposite direction from the host and the right picture is the correct orientation when the record button is facing towards the host.

The orientation changes during the video because the filming was done in two parts. The first recording was the introduction part and the second part was the test itself. Between recording, the battery died on the camera and therefore, it required charging. During the process of

charging and putting the device back on the tripod the orientation of the camera changed. The camera's orientation was only opposite during the introduction part of the filming. Therefore, in the experiment, the host only changed place when the transition between the intro and the experiment video took place.

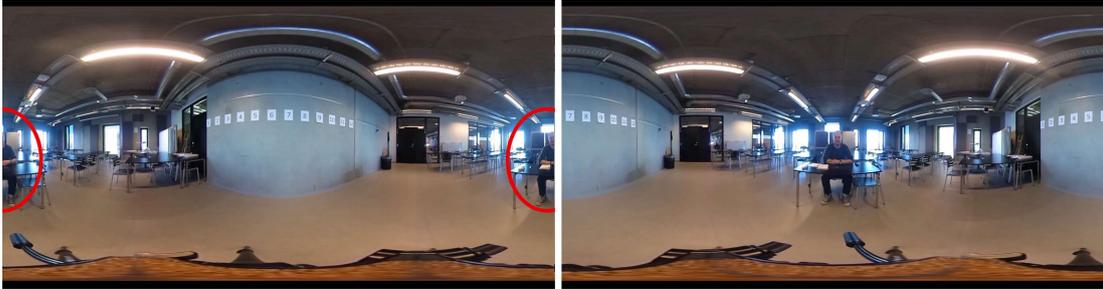


Figure 6.3: Error observed from the experiment. Mistake had happened during the filming when the camera's orientation had changed.

During the experiment, by observing the participants, many of them used the existing room as a tool to investigate where the host was pointing in the video. Especially, using the rooms and objects as a tool to indicate where each number could be located. However, it is unsure if the participants did this consciously because it seems that they did not notice it themselves.

However, this may have been connected to Gibson's theory of affordances [6], which could explain the participants' assumptions on not needing to turn or rotate the chair to see the numbers on the wall. A study by Gibson and Pick [7], investigated eye gaze direction and accuracy. They came to the conclusion, that to determine the gaze of the eyes, they used the hosts' head direction to help detect the direction of the gaze. In this case, the participants used the room and objects to navigate and determine the pointing target.

Furthermore, during the experiment, it occurred that the peripheral numbers on the wall seemed to be most difficult to the participants. As an observation and a feedback from many of the participants, the numbers on the wall, that were behind the participants, were easiest for them to detect. It seems, that the participants perceive point towards them easier than the ones with a long distance to the host.



## 7 | Discussion

The experiment's results showed that there is a significant difference between the real world test and the three interaction techniques. However, the questionnaire results revealed interesting facts as the participants rated the quality of pointing in the virtual reality test to be significantly better than in the click and drag or the IMU-based rotation tests. It should be kept in mind, that the data from the experiment is more objective than the self reported data. However, together with the experiment data, the questionnaire data could reveal further information and contribute towards the already received data.

To fully understand these results, it is necessary to discuss what kind of decisions have led to these results. Furthermore, as the results indicated that there were differences between the real world test and the interaction techniques, it is interesting to know what could have influenced these results. The following scenarios and causes are a result of a process that tried to identify possible reasons for the results and therefore, they may, or may not, be connected to the rejection of the null hypothesis.

(1) The devices were too old and could not provide the needed quality to complete the experiment. Better equipment could have a better support for consumer market applications. Therefore, the quality of the application could possibly increase the quality of the experiment. With a higher image quality and sensory equipment, the experience could become more natural than with the equipment used for this project's experiment. Therefore, it presents a possible reason for future implementation and repetition of the experiment.

(2) As mentioned in State of the art, telepresence is a big factor in the virtual reality and 360-degree videos which could have an effect in the experience. Marvin Minsky[15] stated, that in order to create true telepresence, it requires all of the senses like smell, haptics and auditory senses and visual senses are required in that virtual environment. Therefore, it brings an interesting question about telepresence on about possible reasons on the results of the experiment. What if there were more critical elements of the telepresence present in the application and experiment? Would the results have been different?

(3) The possibility of the participants to learn where the numbers are by memorizing in which direction the numbers are located. With these, they also estimate the locations of the numbers on the wall by perceiving the geometry of the room and the environment itself and then using that unconsciously the environment as a navigation tool to determine the direction of the gesture pointing. Even the smallest unconscious movements and actions can cause a person to

make unconscious decisions. For example, where the host is pointing and if the host is pointing towards the subject and had the oriented to right, this might affect the recipient's opinion of the pointing target. As mentioned in the observations, the study by Gibson and Pick [7] was investigating how a person can read another person's eye-gaze and a direction. They stated, that the perception of eye-pointing is influenced by the direction of the head.

Furthermore, Gibson's Affordances [6] could explain the behaviour of using the environment as a tool to determine the direction of deictic pointing. According to Gibson, the affordances of an environment is what it can offer to the individual and what it provides or furnishes, even good or ill. According to Gibson, the perceived surfaces and layouts are separated in two parts; horizontal and vertical.

By horizontal surface Gibson means the earth – the ground that affords support. He says, that the ground is the basis of visual perception and the behaviour of all land animals, including humans. The ground is the earth that we perceive, a flat ground that reaches the horizon and beyond. The horizontal flat is the ground for everything that lies on the horizon. For example, trees, hills, mountains, and houses. Therefore, anything that we sense and perceive is connected to the horizontal plane and even though, the horizon sometimes is not visible due to objects on the horizontal surface, we always know that the horizon continues behind the objects.

A vertical surface is a solid surface such as a wall or a cliff that is according to Gibson, a "barrier" to human movement. The only way to overcome these "walls" is to have an object that affords climbing e.g. a ladder.[6] The only other way to go through these vertical surfaces is to find an open area between these vertical surfaces. Thus, depending on the steepness of the open area between the vertical ares, the open area can afford for an action. For example, a corridor's walls are the vertical surfaces and the open gap between the walls form the floor that affords walking.

It would be interesting to know how these Gibson's affordances were present in the experiment and if they play a big part in perceiving the deictic references. As mentioned in the observations, some participants were not fully interacting with the application. They knew the numbers even without looking around to see where the deictic pointing was directed at. It seemed like the participants used the environment to find the direction of the pointing. If they were indeed perceiving these affordances in the environment, they could have used all the objects and flats in the surroundings to identify what the target of the pointing was. Gibson mentioned, that objects can have countless qualities and properties like color, texture, composition, size, shape, mass etc. The properties, in this case the size, could help the participants to determine the sizes of tables, the room, and the wall where the numbers are standing. [6]

Based on these theoretical assumptions, it can be interpreted that Gibson's theory of affordances could have played a bigger part in the experiment. Even though, there are no clear tendencies stating this phenomenon in the data. However, analysing the observations opened this possibility of affordances affecting the results of the experiment. The real world test was held in the same room as the other three tests with the same environment and with numbers on a wall. Therefore, there is a chance that the real life environment may have affected how the participants interpreted the virtual environment. This means, that if the participants used their assumptions of objects' and surfaces' sizes of the real world, they could have used this

information to determine the real target of the deictic pointing. Therefore, Gibson's theory could be one of the influencers for the experiment's results.

Therefore, it is clear that a person can perceive a lot of information from body language, eye-gaze, and direction of the head. Studies have even proven that some of these even reveal information about where a person is looking or pointing. However, combined with the information gathered from the questionnaire, it seems that the participants experienced the virtual reality technique as the easiest one. Moreover, the questionnaire data stated that the virtual reality technique was the only technique that was comparable to the real world situation. However, it is interesting that people felt that the virtual reality experience was much easier than the click and drag, they still received the same results from the experiment. Maybe it was more challenging to them to determine the target of the deictic pointing, but without further investigation in these earlier mentioned possibilities, it is not possible to conclude on these assumptions.

## **7.1 Possible Future Implementations**

As mentioned in the discussion, possible reasons that could have influenced the results provided ideas that might lead to different results if the experiment was done again in the future. The experiment needs to be done again in order to verify the data received from the current experiment. As the possibilities mentioned multiple approaches for the future experiment, they are separated in three parts.

First possible future implementation would be a test in an actual location where this deictic pointing could occur, e.g. a museum, a real-estate tour, or even a virtual tourist tour in the city. Therefore, this kind of test could provide more information about the motivation of the project and provide real data from a real scenario. Something that must be payed attention to in future implementations are the eye-gaze and body language of the host. These were not considered at the time of the experiment as big influencers, but multiple studies seemed to support that eye-gaze and body language easily can influence the direction of the deictic pointing.

### **7.1.1 Better Equipment**

The second future implementation that would interesting, would be to enhance the equipment applied in the experiment. It is stated in the limitation of the design and multiple other locations that the equipment was outdated and faulty, and therefore, it caused some unnecessary problems for the project which also could have caused a lower quality for the experience in the experiment. Upgrading the equipment to up-to-date technology, that is meant for virtual reality and 360-degree video experiences, for example, a smartphone and a tablet that provides screen resolution above full hd, and a 360-degree camera that can record higher than full HD. These improvements only, could improve the device resolution problem which was much complained by participants.

Furthermore, it could also be interesting to see how the experience differ in the mobile virtual reality to the desktop computer virtual reality like Oculus Rift. As the advanced virtual reality devices are released on the consumer market this year(2016), it becomes relevant as the their popularity becomes higher.

If an upgrade of the equipment is done right with all the bug fixes and application fixes, the existing experiment results could possibly provide a completely different experience in the experiment and result in completely different results from the current experiment.

Along with gathering better equipment for a retest, another test in the field of telepresence would be interesting. It is known, that the more senses the virtual experience occupies, the more immersed the user is in the virtual environment.[15] However, it is not well known how the user's experience and behaviour changes, especially when detecting deictic pointing, in virtual environment e.g. virtual reality when there are more senses included in the experiment. Basically, a test to investigate if and how a level of telepresence can affect the experience of interpreting deictic pointing in mediated environments.

### 7.1.2 Gibson's Affordances

As mentioned in the discussion, Gibson's affordances could have been a possibility for the null hypothesis, however, in order ot test this, it is required to create an environment that does not provide too much information that could alter the results of an experiment. He mentioned, that we humans perceive all the flat horizontal and vertical surfaces and their properties alongside all the objects with their own various properties. Therefore, instead of having multiple objects around in a rectangular room, the room should be empty and cylindrical so the experience would be something that the participants to do not experience on a daily basis.

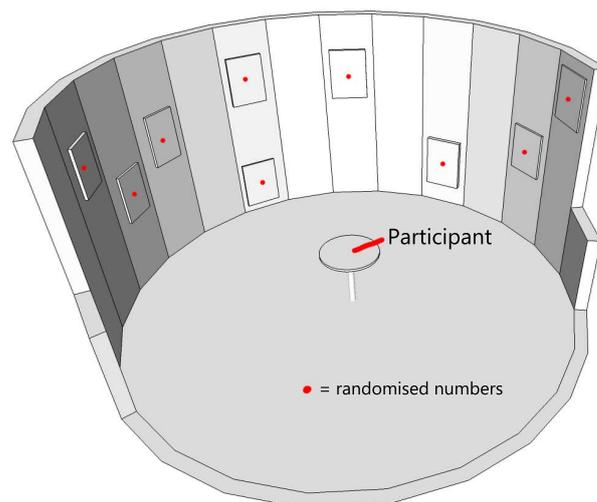


Figure 7.1: Future implementation to investigate how a round room could influence deictic pointing. Sketch's wall is open for demonstrative purposes, in reality the room would be closed fully.

However, in the future implementation, if the room is a cylindrical small room or space, a person cannot perceive its measures as well as in a rectangular room. It is harder to perceive the dimensions of round rooms because people are more used to rectangular rooms which dimensions are easier to estimate. The implementation would have randomised numbers around the round room or space which would minimize the learning effect because the memorization of the numbers would become harder. As seen in Figure 7.1, the room would have locations on the wall where the numbers change randomly during the test. The red dots indicate the possible places of these numbers on the wall. This implementation would provide information on how important the environment actually is for the deictic pointing.



## 8 | Conclusion

This project compares different navigational styles that can be used for exploring 360-degree videos, especially focusing on deictic pointing in the videos. During the experiment, the participants were asked to detect and identify the direction of deictic pointing to outside the field of vision in 360-degree videos. The experiment compares the three interaction techniques; click and drag, IMU-based rotation and virtual reality with real world test results.

According to the results, it does not matter whether one is using a virtual reality or a tablet to navigate through 360-degree videos because the users can detect these gestures as well with either technique that was tested in the experiment. No significant difference in this experiment between click and drag, IMU-based rotation and virtual reality was found. However, there was a significant difference between the real world results and these three interaction technique results, as the real world results were significantly better than the other three.

A questionnaire asked the participants to rate the quality of the pointing in each test, and therefore, the tests were comparable to each other. The participants preferred the virtual reality interaction technique while the quality of the click and drag technique was the least preferred. However, the main experiment results showed that both interaction techniques resulted in a very similar performance. Overall, the users seem to prefer and experience the virtual reality to be the best interaction technique out of the three tested techniques even though, it received the same results as the other two techniques in the main experiment.

Overall, this study provided important information on how users determine the target of deictic pointing. However, instead of merely focusing on only the deictic pointing, the focus was also put on the environment. As mentioned in Chapter 6.2, it was observed during the experiment that the participants used the environment as a tool to determine the target of the deictic pointing. Furthermore, as discussed in Chapter 7, it is important to state here that the environmental cues can provide a lot of information when determining the deictic pointing direction. According to Gibson[6], the environment we perceive is full of information on properties of the objects which include sizes, measurements, colors, textures, weight, shapes etc. Therefore, according to Gibson [6], we can use this perceived information to help us, for instance, to determine the pointing direction. These observations could lead to a very relevant information for the actual applications that were mentioned in the Motivation section in Chapter 2. A possible application for this information could be the virtual tours made in museums, and real-estate and tourism locations. However, these results are only relevant if the environmental cues and Gibson's[6]

affordances are indeed influencing the results.

However, the experiment results showed, that when people experience 360-degree videos with tour guides where the guide may point at objects outside the field of vision, it does not matter which mobile interaction type is applied to experience 360-degree video. Even though, the virtual reality was the preferred interaction technique, the actual data from the experiment indicated that virtual reality was not significantly different from the other two interaction methods. However, there is no doubt that the experiment should be repeated with better equipment in order to see if the results would be the same, after fixing these issues.

# Bibliography

- [1] Adrian Bangerter and Daniel M Oppenheimer. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture*, 6(1):85–102, 2006.
- [2] D.A. Bowman. *3D User Interfaces: Theory and Practice*. Pearson Education. Addison-Wesley, 2005.
- [3] Tara Callaghan and John Corbit. *The Development of Symbolic Representation*. Wiley, 2015.
- [4] Durovis. Durovis dive. <https://www.durovis.com/dive.html>, Accessed 2016 May 16th.
- [5] Caroline Gale and Andrew F Monk. Where am i looking? the accuracy of video-mediated gaze awareness. *Perception & psychophysics*, 62(3):586–595, 2000.
- [6] James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.
- [7] James J Gibson and Anne D Pick. Perception of another person’s looking behavior. *The American journal of psychology*, 76(3):386–394, 1963.
- [8] Lee JaeYun. Easy movie texture plugin. <https://www.assetstore.unity3d.com/en/!/content/10032>, Accessed 2016 May 12th.
- [9] MerryAnn Jancovic, Shannon Devoe, and Morton Wiener. Age-related changes in hand and arm movements as nonverbal communication: Some conceptualizations and an empirical exploration. *Child Development*, pages 922–928, 1975.
- [10] Sotaro Kita. *Pointing: Where Language, Culture, and Cognition Meet*. Psychology Press, 2013.
- [11] Jacques D Lempers. Young children’s production and comprehension of nonverbal deictic behaviors. *The Journal of Genetic Psychology*, 135(1):93–102, 1979.
- [12] Ché Lucero, Holly Zaharchuk, and Daniel Casasanto. Beat gestures facilitate speech production. pages 1029–1038, Unknown.
- [13] David McNeill. So you think gestures are nonverbal? *Psychological review*, 92(3):350, 1985.

- [14] David McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University Of Chicago Press, 1992.
- [15] Marvin Minsky. Telepresence. 1980.
- [16] Robert J Moore, Nicolas Ducheneaut, and Eric Nickell. Doing virtually nothing: Awareness and accountability in massively multiplayer online worlds. *Computer Supported Cooperative Work (CSCW)*, 16(3):265–305, 2007.
- [17] Nokia. Nokia ozo. <https://ozo.nokia.com/eu/>, Accessed 2016 May 22th.
- [18] Kai Daniel Oberländer. Inverse kinematics: Joint considerations and the maths for deriving anatomical angles. 2015.
- [19] Andre Odendaal. Touchlook. <http://wiki.unity3d.com/index.php?title=TouchLook>, Accessed 2016 May 5th.
- [20] Ricoh. Ricoh theta s. <https://theta360.com/en/about/theta/s.html>, Accessed 2016 May 12th.
- [21] Michelle L Rusch, Mark C Schall, Patrick Gavin, John D Lee, Jeffrey D Dawson, Shaun Vecera, and Matthew Rizzo. Directing driver attention with augmented reality cues. *Transportation research part F: traffic psychology and behaviour*, 16:127–137, 2013.
- [22] Allison Sauppé and Bilge Mutlu. Robot deictics: How gesture and context shape referential communication. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 342–349. ACM, 2014.
- [23] Chris L Schmidt. Adult understanding of spontaneous attention-directing events: What does gesture contribute? *Ecological Psychology*, 11(2):139–174, 1999.
- [24] Transworld. 360fly closes 40 million dollars in series c round of funding, with hydra ventures. <http://business.transworld.net/news/360fly-closes-40-million-in-series-c-round-of-funding-with-hydra-ventures-as-newest-investor/KXI6yj9bMIF7z66w.97>, Accessed 2016 May 16th.
- [25] Unity. Movie texture. <http://docs.unity3d.com/Manual/class-MovieTexture.html>, Accessed 2016 May 12th.
- [26] Venturebeat. Matterport raises 30m for virtual reality real estate tours. <http://venturebeat.com/2015/06/25/matterport-raises-30m-for-virtual-reality-real-estate-tours-and-more/>, Accessed 2016 May 12th.
- [27] Nelson Wong and Carl Gutwin. Where are you pointing? the accuracy of deictic pointing in cves. pages 1029–1038, 2010.

# A | Experiment Questionnaire

# Nonverbal pointing Questionnaire

1. Your age?

.....

2. Gender

Mark only one oval.

Female

Male

3. Are you right or left handed?

Mark only one oval.

Right handed

Left handed

4. Which of the platforms feel most natural to use for you?

Mark only one oval per row.

	Best	OK	Worst
Click and Drag	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Move Around with tablet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
VR	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. How did you consider the quality of pointing in the real world test? (1 = Hard, 5 = Easy)

Mark only one oval.

	1	2	3	4	5	
Hard	<input type="radio"/>	Easy				

6. How did you consider the quality of pointing in test2? (1 = Hard, 5 = Easy)

Mark only one oval.

	1	2	3	4	5	
Hard	<input type="radio"/>	Easy				

7. How did you consider the quality of pointing in test3? (1 = Hard, 5 = Easy)

Mark only one oval.

	1	2	3	4	5	
Hard	<input type="radio"/>	Easy				

**8. How did you consider the quality of pointing in test4? (1 = Hard, 5 = Easy)***Mark only one oval.*

	1	2	3	4	5	
Hard	<input type="radio"/>	Easy				

---

**9. If you were to experience a 360 degree video, what would be your personal choice?***Mark only one oval.*

- Click and Drag navigation
  - Tablet rotational navigation
  - Virtual Reality Mobile
- 

Powered by

