NTNU
Norwegian University of
Science and Technology

Department of Electronic Systems

TFE4580

Electronic Systems Design and Innovation
Specialization Project

Measuring Quality of Experience using Augmented Reality in Training Systems

---

Peter Remøy Paulsen

December, 2020

Abstract

In this article, we took a look at measuring Quality of Experience in an mobile augmented reality application. We did this by exposing a group of participants of an experimental application, where half of the participants got to experience an environment with occlusion enabled, while the other half had no occlusion enabled. The results were gathered from a set of questions where the participants were asked regarding their own impression of emotions, usability, immersion and overall experience. There were no clear results, but there is an indication that proper occlusion will increase the Quality of Experience.

Samandrag

I denne artikkelen ser vi på måling av opplevingskvalitet i ein mobil utvida røyndapplikasjon. Dette gjer vi ved å sjå på korleis ei gruppe deltakarar reagerer på ein spesiallaga testapplikasjon. Deltakarane vart delt i to, der den eine gruppa fekk oppleve eit miljø med ein virtuell stol som forsvann bak ekte objekt i rommet, som om den virtuelle stolen skulle fulgt vanlege fysiske lover. Dette vert kalla okklusjon. I det andre miljøet forsvann ikkje den virtuelle stolen i det heile, og stolen flaut over dei ekte objekta. Resultata frå deltakarane vart samla frå eit spørjeskjema, der dei fekk spørsmål om kva dei følte, om applikasjonen var brukarvennleg, kor engasjerte dei følte seg og den generelle opplevinga deira. Det ser ut til at det ikkje var nokre klare resultat, men vi ser indikasjonar mot at okklusjon vil auke opplevingskvaliteten.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As natural hazards occur with increasing frequency, we need to be able to better prepare. A possible way of doing this is by immersive simulations and the use of these in training systems for these types of situations.

World of wild waters is an initiative put together by NTNU, which aims to research this area and find new solutions on how to better prepare. As a part of this initiative we will in this article take a look at different aspects and features of augmented reality which can affect the Quality of Experience (QoE) in a training system.

More specifically we will take a look at occlusion and how this may affect immersion and therefore QoE. We asks ourselves if occlusion, meaning virtual objects disappearing behind real objects as if they were real, will increase the immersion and make the participants feel more engaged. Will this then result in an higher overall Quality of Experience?

To research whether or not occlusion does have an effect, we designed a system with an experimental application using the game engine Unity and augmented reality software development kits, where we exposed a set of participants for two different environments. One of the environments had an occlusion algorithm working, while the other environment did not have any occlusion applied whatsoever. Before the participants tried the experimental application, they filled out a questionnaire to map their background and previous experience and after completing the experiment, they filled out a questionnaire to map their experience of using the experimental application.

We present the findings from the experiment and further discuss the results. Furthermore we present some suggestions for work that needs to be further explored and improved. In the end we come to a final conclusion from our research and try to answer our own queries.

Chapter 2

Background

2.1 Motivation

As extreme weather and natural disasters are happening with an increasing frequency, it has become one of the largest global impact problems with the highest likelihood of occurrence [1]. Natural hazards have cost Norway alone 27 Billion NOK, and in an effort to learn more about how to handle such disasters, NTNU has initiated the World of Wild Waters initiative. This initiative is set out to increase the understanding of natural hazards by utilizing immersive user experiences based on real data, realistic scenarios and simulations [1].

2.2 Immersion

Immersion is, as according to C. Jennet et al. [2], described as an experience in one moment in time and that can be graded (i.e. engagement, engrossment, total immersion). Immersion involves a lack of awareness of time, a loss of awareness of the real world, involvement and a sense of being in the task environment. Most importantly, immersion is the result of a good gaming experience.

Levels of immersion according to C. Jennet et al. [2] can be found in table 2.1.

Table 2.1: Three distinct levels of immersion. Collected from [2]

Level	Explanation
Engagement	Overcoming the barrier of gamer preference and invest time, effort and attention in learning how to play the game and getting to grips with the controls
Engrossment	Overcoming the barrier of game construction, where the game features are combined in such a way that the gamer's emotions are directly affected by the game and the controls are "invisible". The gamer is now less aware of their surroundings and less self-aware than previously
Total immersion	A sense of presence, being cut off from reality to such an extent that the game was all that mattered. Requires the highest level of attention and can be rare and rather fleeting experience

2.3 Realism in Augmented Reality

Google has stated some of their design guide lines for augmented reality [3] with the key takeaways presented in table 2.2.

2.4 Quality of Experience

QUALINET white papers defines Quality of Experience as following [4]

> The degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state.

Quality of Experience is a field which is based on multiple disciplines such as social psychology, cognitive science, economics and engineering service with a focus on understanding overall human quality requirements.

There are a number of influencing factors in regards of the general Quality of Experience, namely human, system and contextual influencing factors [5].

Table 2.2: Design guidelines for AR. Collected from [3]

	Explanation
Modeling sizes	The virtual objects should appear to be life sized and correctly scaled according to its environment
Texturing	Visual noise in the texture makes it look more lifelike
Physically Based Rendering	Multiple textures creates an illusion of depth and texture
Normal Maps	Giving virtual objects more photorealistic appearance by adding details to its appearance without adding additional geometry
Ambient Occlusion	Rendering technique where the amount of shadow on the objects is modulated accordingly
Shadow Planes	Virtual surfaces that sit beneath the virtual objects to appear as real shadows and give the objects a more grounded perception
Lighting	Virtual objects must be lit in the same manner as real objects in the environment to not break immersion
Depth	By including shadow planes, occlusion and textures
Presence	The virtual objects need to apply by the same physical properties as in the real world

- **Human influencing factors** (HIF) can be divided into two parts — low level and high level. Low level factors are factors such as age, physical form, emotions and mental constitution, while high level are factors such as previous knowledge regarding the matter.
- **System influencing factors** (SIF) which is the technical elements in role. The type of content being consumed, what kind of media (meaning factors such as encoding, resolution, sample rate), network constrains (e.g. bandwidth, delay and jitter) and device differences (e.g. different screen sizes, resolutions, frame rate and audio quality)
- **Context influencing factors** (CIF) are the surrounding factors which affects the user. The physical location (e.g. lighting and surrounding space), social relationships (e.g. inter-personal relationships), type of task, interruptions, time of day, how many times the user has been using these type of systems before and more technical contextual challenges (e.g. a system which has to work together with other separate systems) are all influencing aspects.

2.4.1 Measurement

Since Quality of Experience is such a multidisciplinary field, there is no simple way of putting a simple measurement method to it which gives clear and definite answers. Section 2.4.1.1 and 2.4.1.2 gives an overview over popular techniques which often is used hand in hand.

2.4.1.1 Mean Opinion Score

The Mean Opinion Score (MOS) is a widely used measurement for media signals and quality. While it is a popular method, its usefulness is often debated due to inherent limitations of measurements in a single scalar value [6].

The subjective quality evaluation requires a lot of human resources and can be time consuming, while the objective evaluation methods gives much quicker result. On the other hand, the objective evaluations requires dedicated computing resources.

The mean opinion score method is otherwise prone to misuse or misinterpretation, as the design of the subjective experiments have an important influence. The objective media quality metrics do also rely on data from the subjective experiments for tuning and validation, and it can therefore be challenging to make meaningful measurements and interpret the resulting findings correctly [7].

2.4.1.2 Self Assessment Manikin

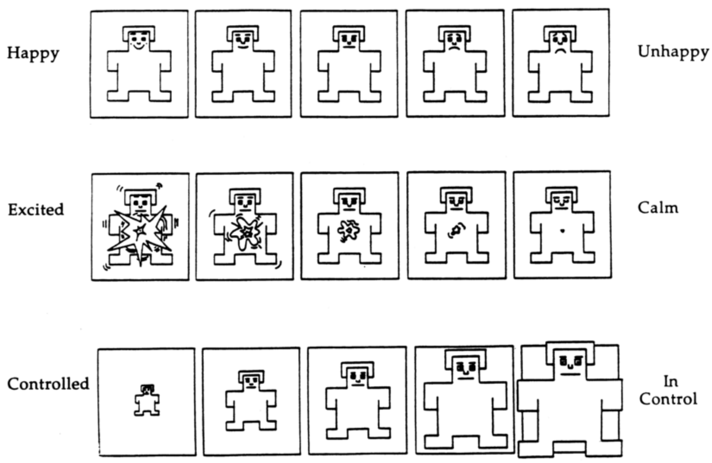


Figure 2.1: Self assessment manikin (SAM) illustrations. Reprinted from [8].

The self assessment manikin (SAM), illustrated in figure 2.1, is a non-verbal pictorial assessment technique [8]. This technique aims to map personal affective reactions, such as pleasure, arousal and dominance, meaning the self perceived level of control.

The technique is a quite effective type of measurement, as it only has 3 ratings. The Semantic Differential (SD) scale, another popular scale, has 18 ratings [8]. SAM scale is in comparison to the SD scale easy to conduct, while being inexpensive. The SAM scale still provides similar results for pleasure and arousal, and may even provide better track of the personal dominance response.

2.5 Related Work

Haynes and Lange [9] has already published some of their findings regarding an augmented reality application for flood visualization. In their experimental design they used in situ modeling of buildings to occlude flood planes in the user's scene. They found that the in situ modeling is a versatile method, but this method is sacrificing the accuracy of the modeling, concluding that there should be done further work to improve on the occlusion geometry modeling methods.

As opposed to only doing a explicit analysis of a user's Quality of Experience using augmented reality and other similar environments, Keighrey et. al. [10] conducted a physiology-based comparison of Quality of Experience. In addition to looking at the explicit answers from a post-test experience questionnaire, they also looked at the implicit measures, such as heart rate and electrodermal activities. Their findings suggested that the explicit and implicit responses seem to result in somewhat equal results regarding the augmented reality experience.

Berge researched the impact of Quality of Experience by comparing virtual reality to a desktop environment [11]. By designing an experimental prototype for both a virtual environment and a desktop environment, the experimenter let the participants solve some tasks comparing weather data and afterwards analyze them. The findings suggest that there is no significant difference to the users' level of understanding in either of the environments, but it seems like the immersion aspect is significantly higher in the virtual reality environment compared to the desktop.

Virtual and augmented reality may enable humans to learn and understand at a greater rate. It is said that the human brain is enable to remember and learn around 20 percent of what it reads after two weeks, while the brain is able to remember and learn as much as 90 percent of what it actually does physically [12]. This is derived from Edgar Dales famous, but often criticized, cone of experience [13][14]. Augmented reality brings great promises by being more human centered. Mullins states that he thinks augmented reality will excel human learning, by utilizing human's natural talent for pattern matching leading to a more efficient knowledge transfer [15]. This should further introduce a bridge between the physical world and more abstract ideas.

Chapter 3

System Design

Based on our findings in chapter 2, we want to research different aspects of Quality of Experience and immersion. At firstm we take a look at occlusion and how this might impact a user's quality of experience using a system.

We have formed the following research question:

> **RQ:** How will the level of occlusion affect the QoE in a Mobile Augmented Reality application?

To help us investigate this matter we form some supporting hypotheses:

> $\mathbf{H_0}$**:** Proper occlusion will increase the QoE in an Mobile Augmented Reality application

> $\mathbf{H_1}$**:** Proper occlusion will not increase the QoE in an Mobile Augmented Reality application

In the coming sections we go through how we designed a system which tried to test this.

3.1 Unity

The system was developed in Unity. Unity is a cross platform game engine which is developed by Unity Technologies [16]. The platform was first released in 2005 as an Mac OS X-exclusive game engine, but has since its initial release grown to become one of the largest and most popular game engines [17] supporting more than 25 platforms [16]. Unity is widely flexible and can be used to create both 2D, 3D, virtual reality and augmented reality experiences.

3.2 Software Development Kits

3.2.1 AR Foundation

AR Foundation is an augmented reality platform extension developed by Unity Technologies. The SDK provides a cross platform developing experience and gives an interface for Unity developers. AR Foundation itself does not implement many augmented reality features and must be complemented by a target platform package. AR Foundation supports the following:

Table 3.1: Target platforms used in AR Foundation

Package	Platform
ARCore XR Plugin	Android
ARKit XR Plugin	iOS
Magic Leap XR Plugin	Magic Leap
Windows XR Plugin	HoloLens

3.2.2 ARCore

ARCore is a software development kit (SDK) made by Google. The SDK enables devices to sense its environment, understand the world and interact with this information [18].

The SDK works on a great number of devices running Android 7.0 or later, and some of the APIs even work across both Android and iOS, which makes ARCore quite ideal to use for developing cross platform augmented reality applications. One should note that some of the APIs in ARCore has its demands, which results in a smaller number of devices supporting all features [19].

ARCore supports several features, such as motion tracking, environmental understanding, light estimation and lots more [20]. ARCore uses the devices camera and inertial sensors. The camera is used to identify interesting feature points (e.g. as shown in figure 3.1 as white circles on the couch) and in the combination with the inertial sensor, the device is able to determine its position and orientation. This enables the device to further understand the world around it and for example place virtual objects, annotation and other information into the real world. The virtual objects can because of this be rendered from the correct perspective and therefore appear as part of the real world.

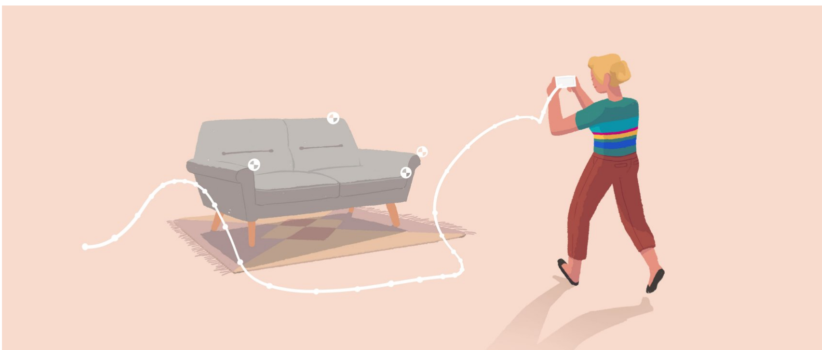


Figure 3.1: Motion tracking. Reprinted from [20].

The motion tracking feature of ARCore uses a simultaneous localization and mapping (SLAM) algorithm, which constructs a mapping of an unknown environment while keeping track of the agent's location within it [21]. The algorithm looks for clusters of feature points that appear to lie on a common horizontal or vertical surface. These surfaces can then construct planes the device can keep track of, since it is also able to find the boundaries of the planes. Virtual objects can for example be rest on a table, as shown in figure 3.2.

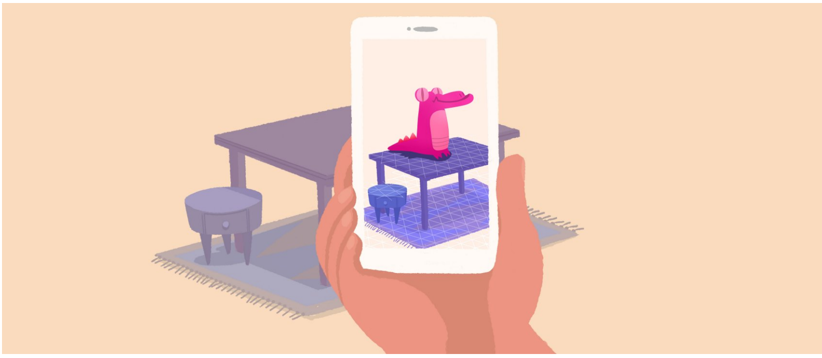


Figure 3.2: Environmental understanding. Reprinted from [20].

One of ARCore biggest features is the depth mapping capabilities. The API uses a depth-from-motion algorithm to create depth maps [22] of the space. The data is written to the depth buffer and can be further utilized. Figure 3.3 shows an illustration of how the depth mapping works in an hallway.

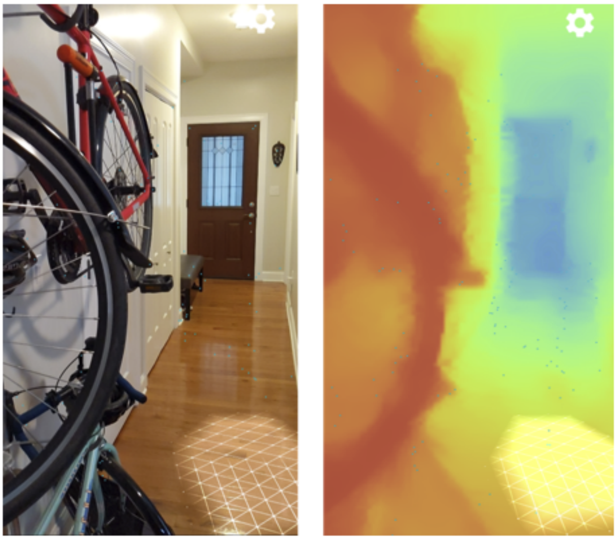


Figure 3.3: Depth map visualization. Reprinted from [23].

The algorithm works by taking multiple images from different angles and afterwards compare them to estimate the distance to every pixel as the user moves the device through the space. Pairs of camera images are treated as two observations of an assumed static scene. Therefore, non-stationary object parts within the scene, will help map the static elements in the scene and we will get a more accurate depth mapping. Unfortunately, the non-stationary objects will be weakened as a result. The algorithm has problems with surfaces without features because of the lack of feature points and is also dependent of the device being moved a few centimeters to collect different observations. The camera data can also be helped by additional depth sensors, if the utilized device has such sensory (e.g. time-of-flight sensory). This will improve the depth data and will vastly improve depth data for surfaces without feature points.

The Depth API can provide data of the depth from 0 to 8 meters, while working the best between 0.5 meters and 5 meters. The error of the depth data increases quadratically with the distance from the camera.

By utilizing the mapping data, we can accomplish a great number of things. Virtual objects can realistically collide with real world objects. When you throw a virtual ball at a wall, it can for example bounce back. The depth data also enables us to measure distances, making our handheld devices measuring tools. Somewhat more psychedelic features from the depth data enables us to manipulate the entire environmental scene around us, by for example manipulate the lighting, re-texture real world objects, manipulate the user's depth of field and numerous other environmental effects.

As some of the algorithms can be quite computationally challenging and with a certain quality requirement in regards of the camera sensor, a smaller number of devices are supporting the Depth API provided in the ARCore software development kit [19].

3.3 Experimental Design

The experimental application has been specifically built for the Google Pixel 5 [24], the latest phone (as of December 2020) from the company providing the ARCore APIs itself. This should ideally be the best device for Googles own software development kits.

3.3.1 Application Walkthrough

The participant was first met with a main introduction scene, which can be seen in figure 3.4a. They received a randomly generated participant number, which was used in both questionnaires they had to fill out (discussed further in section 3.3.2). The randomly generated participant numbers were used to bind the pre-testing questionnaire to the correct post-testing questionnaire, while keeping the participants anonymous.

Further, the participant got a general idea of what they will do in the experimental application, before they will be told by the experimenter to choose either environment 1 or 2 can be found. The difference between these two environments is explained in section 3.3.1.2.

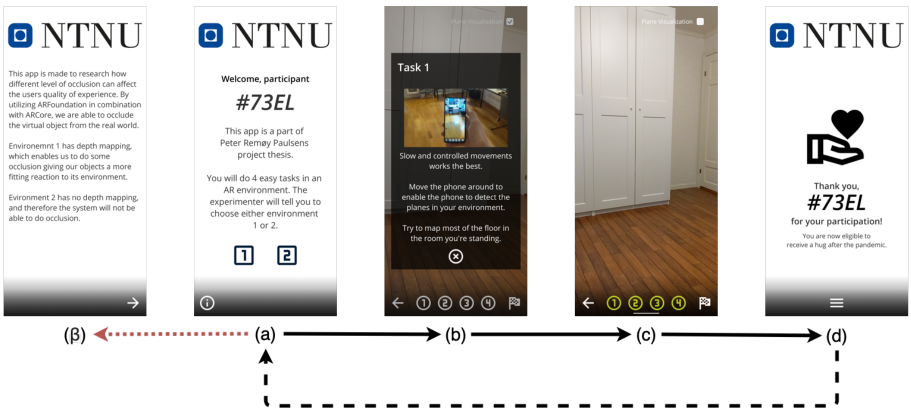


Figure 3.4: The flow through the application.

Down at the bottom of the main introduction scene, shown in figure 3.4a, a information button were presented. The button took the user to the scene shown in figure 3.4β, where a brief explanation of the difference between environment 1 and 2. The scene were not meant to be seen by the participant before the experiment but were made as a precaution for when the experimental application is public and further explanation of the differences between the two environments were needed.

After the participant pushed either the environment 1 or 2 button as instructed by the experimenter, they were taken to the scene presented in figure 3.4b, where the first task was shown by default. The participant performed the given tasks. The tasks are discussed further in section 3.3.1.1.

As the participant read and went through the tasks, the respective task turned green, to indicate that the participant could proceed. When the tasks were completed, the participant could press the checkered flag button, shown in figure 3.4c. The participant was then taken to the final screen, shown in figure 3.4d, where they would be thanked for their participation, and afterwards be instructed to fill out the post-testing questionnaire.

3.3.1.1 Tasks

The participants of the experiment where told to execute four simple tasks. The tasks were specifically designed to try to learn the participants how to use the test setup and get them comfortable using the augmented reality application. Task 4 were the final and real task of the experiment, where the participant was going to experience some of the implemented occlusion or no occlusion at all, depending on the selected environment.

The tasks were as following in the list and how they were presented for the participant in the experimental application can be seen in figure 3.5.

1. Participant is told to move the phone around to enable the phone to detect the planes in the environment.
2. Participant is told to press a plane surface to place a virtual chair in the room, and after that turn off the *Plane Visualization* toggle.
3. Participant is told to place the chair in the room and to try to see it in its environment.
4. Participant is told to try to watch the chair from behind another furniture or walls to see how it blends in and to avoid surfaces without features. If the participant is using environment 1, the occlusion algorithm will try to occlude the virtual chair from the real environment. Environment 2 has no such algorithm activated, and

therefore, the virtual chair will not be occluded, and it will float over anything on screen.

In addition to the text, the participants were presented with an explanatory video in the task dialog box, demonstrating what the participant should do as an alternative to the text based instructions.

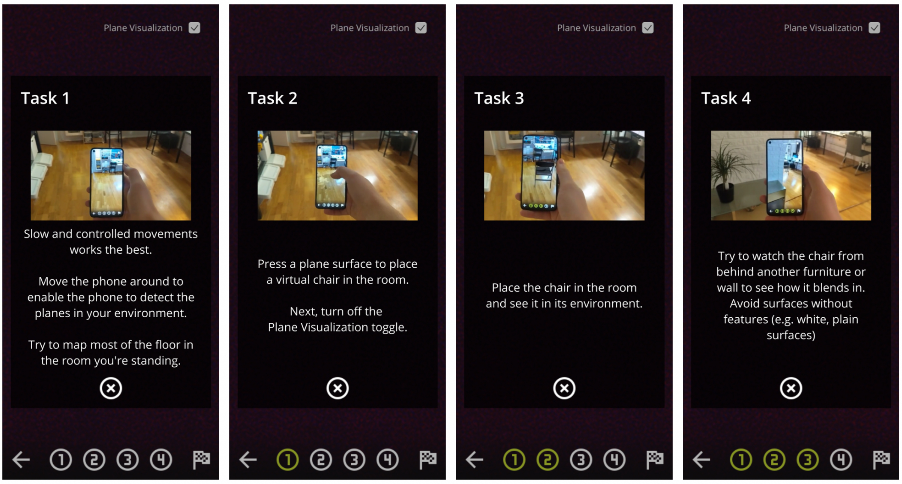


Figure 3.5: Tasks the participants had to complete

3.3.1.2 Environments

The experimental application contained two different kinds of augmented reality environments. The environments were the same, but environment 1 had an occlusion algorithm component activated. By utilizing ARCore's APIs we could enable environment 1 to map the depth and then do some occlusion giving the virtual objects a more fitting reaction to its environment. This can be seen in figure 3.6a.

In environment 2 the occlusion algorithm component was disabled. The environment was therefore not able to do depth mapping and therefore not able to do occlusion at all. This can be seen in figure 3.6b, as the virtual chair just floats above the real chair.

(a) Environment 2

(b) Environment 2

Figure 3.6: The two different environments and their occlusion.

3.3.2 Procedure

The experiments were conducted over three weeks. The experiments were conducted in different locations with the focus on getting the testing data, and ignoring the possible external environmental differences.

The experiments were conducted over three weeks. Since the experimental application works the best in good lighting conditions, it was tried to execute the experiment when natural day light was accessible. The experiments were conducted in different locations, since the experimenter actively sought out the participants. The experimental application was not especially prone to the testing environment, so it could be conducted at any location, but as mentioned, good lighting were preferred, as well as an appropriate size of the testing area. Areas and rooms containing surfaces with little features (such as white, or any other single colored, surfaces) were avoided as this was hard for the occlusion algorithm to handle. The experimenter focused more on trying to collect the participant data, rather than focusing on the possible challenges the environmental differences can introduce to the experiment.

The participants were asked to fill out a pre-experiment web based questionnaire. This was to map their demographic information, such as age, gender and previous experience with augmented reality. During the execution of the experiment the participants were isolated from other distractions. The experimenter was the only other person in the room, to possibly help if any problems or uncertainties would arise during the execution. The experimenter only gave executional help, that hopefully would not introduce biases in regards of the experience. After the participants had conducted the tasks in the experimental application, discussed in section 3.3.1.1, they were asked to fill out the post-experiment web based questionnaire. This questionnaire collected data about their experience and other feedback.

Each of the participants were explained the purpose and goal of the experimental application after they had completed the experiment. They also had the opportunity to try the respective environment which they had not tried during the experiment (the participants that tried environment 1 got to see environment 2, and vice versa).

Chapter 4

Results

This chapter will be used to compare the results from the two different environments and comparing them against each other. There was a total of 24 participants, evenly divided between the two environments (meaning 12 participants for each environment). Figure 4.1 shows the age distributions of the participants, while figure 4.2 shows the gender distribution. Figure 4.3 shows the participants own perception of their gaming experience, and lastly figure 4.4 shows their augmented reality experience.

Every measure was plotted in histograms, along with error bars giving the standard deviation for the data set. The standard deviation was used to check the spread and variability of the data.

We manually looked at the histogram data to gather information, but in addition, we tried to check for any statistical significant differences in the data. Since our data set were not normally distributed, we decided to go for a *Mann-Whitney U test* [25], where we looked at the p-value of the data.

After we had looked at the qualitative data, we took a brief look at the quantitative responses and feedback from the users.

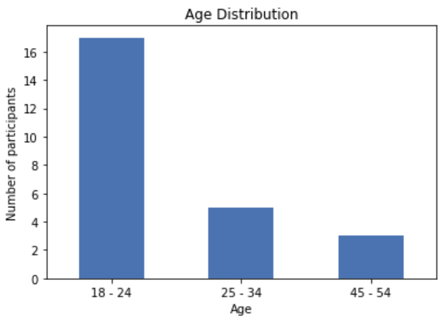


Figure 4.1: Age Distribution

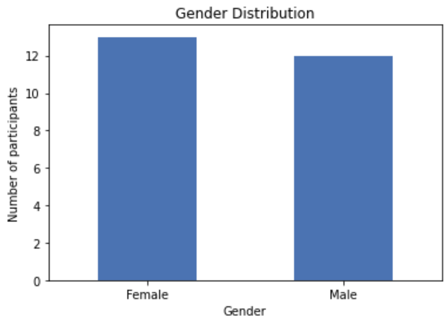


Figure 4.2: Gender Distribution

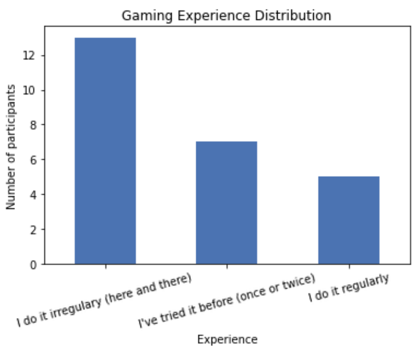


Figure 4.3: Gaming Experience Distribution

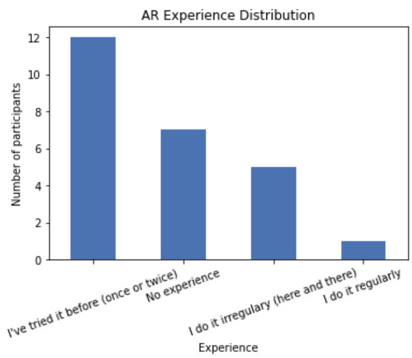


Figure 4.4: AR Experience Distribution

4.1 Questionnaire Data

The questionnaire data were collected from the post-testing questionnaire where the participants answered a total of 16 questions regarding the experimental application.

Questions regarding the emotions the participants felt were answered using a SAM-scale, discussed in the beginning of section 4.1.1, while the remaining answers could be given as Likert scale answers [26], as shown in table 4.1.

Table 4.1: 5-point Likert scale

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

4.1.1 Emotions

Questions for this part of the questionnaire are given in table 4.2 and the answers were given as a 9-point SAM scale as discussed in chapter 2, section 2.4.1.2. Note that the participants were answering a 9-point SAM scale, as opposed to the 5-point SAM scale depicted in section 2.1, as there was an additional option between each of the illustration steps.

Table 4.2: Questions regarding the participants emotions (SAM, section 2.4.1.2)

	Question
Q1	How unhappy or happy did you feel while using the application?
Q2	How calm or excited did you feel while using the application?
Q3	How little or much control did you feel you had while using the application?

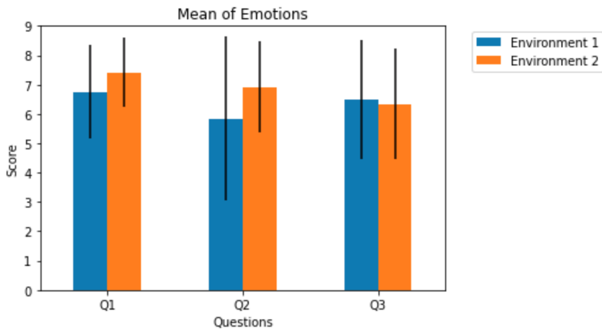


Figure 4.5: Mean score for questions regarding **emotions** with standard deviation error bars

Table 4.3: Mean average and p-values for emotions

	Q1	Q2	Q3
Environment 1 Mean	6.75	5.83	6.50
Environment 2 Mean	7.42	6.92	6.33
Delta Mean	-0.67	-1.08	0.17
p-value	0.15	0.29	0.43

Figure 4.5 and table 4.2 shows us that environment 1 scored lower than environment 2 in regards of happiness and how excited the participants were. On the other hand, the participants of environment 1 were feeling slightly more in control than the participants of environment 2. One should also note that there seems to be a pretty big standard deviation for environment 1 in regards of how excited the participants were.

Table 4.4: Questions regarding the participants feelings of usability

	Question
Q4	I thought the application was easy to use
Q5	I thought the application was inconsistent
Q6	I felt confident using the application

4.1.2 Usability

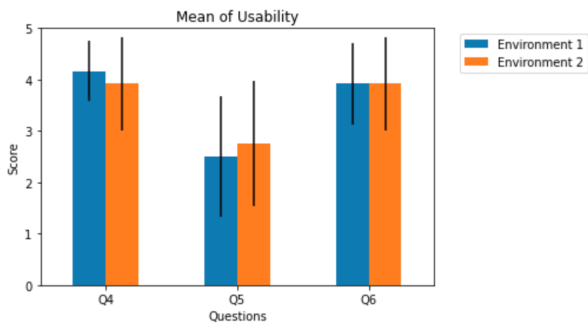


Figure 4.6: Mean score for questions regarding **usability** with standard deviation error bars

Table 4.5: Mean average and p-values for usability

	Q1	Q2	Q3
Environment 1 Mean	4.18	2.50	3.92
Environment 2 Mean	3.92	2.75	3.92
Delta Mean	0.25	−0.25	0.00
p-value	0.28	0.34	0.50

As we can see from figure 4.6 and table 4.4 there is no significant visual or statistical difference in any of the questions regarding the usability, but one can note that **Q5**, in regards of consistency, all of the participants were pretty neutral. This will be further discussed in the discussion, chapter 5.

4.1.3 Immersion

Table 4.6: Questions regarding the participants feelings of immersion

	Question
Q7	I was immersed into the mobile screen environment
Q8	I felt separated from the real-world environment
Q9	I did not feel the need to look at anything other than the mobile screen
Q10	The visuals of the virtual object were convincing
Q11	I felt I could control the virtual object
Q12	The virtual object blended nicely into the environment
Q13	The object disappeared behind other real world objects
Q14	I felt that I experienced the virtual object, rather than watching it

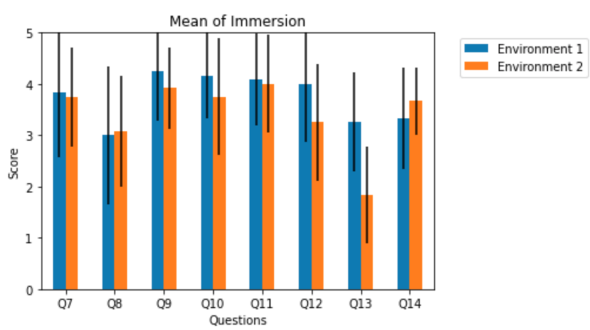


Figure 4.7: Mean score for questions regarding **immersion** with standard deviation error bars

Table 4.7: Mean average and p-values for immersion

	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
Environment 1 Mean	3.83	3.00	4.25	4.17	4.08	4.00	3.25	3.33
Environment 2 Mean	3.75	3.08	3.92	3.75	4.00	3.25	1.83	3.67
Delta Mean	0.08	−0.08	0.33	0.42	0.08	0.75	1.42	−0.33
p-value	0.32	0.41	0.11	0.20	0.43	0.06	0.001	0.22

Figure 4.7 and table 4.7 shows us some interesting results. Across the board, environment 1 scores higher than environment 2, but with quite some error bars for standard deviation, except for question **Q8** and **Q14**. One can also note that question **Q7**, **Q8** and **Q11** only has a delta of $|0.08|$. We can see that the p-value of **Q12**, $p = 0.06$, and especially **Q14**, with $p = 0.001$, gives us a statistically significant difference between the two environments.

4.1.4 Experience

Table 4.8: Questions regarding the participants experience

	Question
Q15	I enjoyed the overall experience with the application
Q16	I would like to try the application, or similar applications again

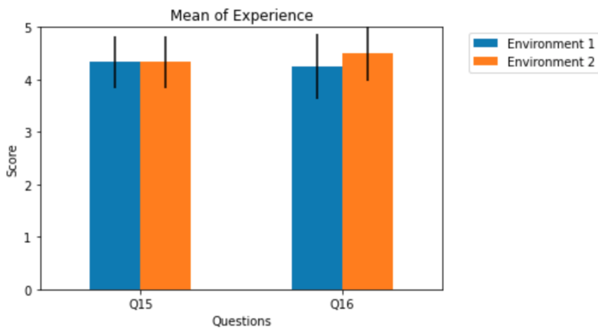


Figure 4.8: Mean score for questions regarding **experience** with standard deviation error bars

Table 4.9: Mean average and p-values for experience

	Q15	Q16
Environment 1 Mean	4.33	4.25
Environment 2 Mean	4.33	4.50
Delta Mean	0.00	−0.25
p-value	0.47	0.17

When it comes to the overall experience, figure 4.8 and table 4.9 shows us that the participants has overall been quite satisfied with the experimental application, where environment 1 and 2 were almost indivisible at **Q15** and **Q16** having a negligible difference.

4.2 Qualitative Feedback

At the end of the post-testing questionnaire, the participants were asked if they had any feedback whatsoever. There were not many, but here they are:

> The lighting/ shadows on the texture of the virtual object is the only thing that makes it look virtual instead of real. This is what reminds you that the object is not there in reality.

> It was sometimes very good, but it had glitches that ruined it from time to time. It did not like me moving fast

> Crash :(

> Good work!

In addition to these, there were some feedback regarding both the experimental application and the questionnaires being in somewhat advanced English, as some of the participants did not feel their English knowledge were quite adequate.

Chapter 5

Discussion

In this chapter we will try to evaluate and discuss our results and interpret what they really mean or at least what direction they point us in. We also discuss some of the other notable mentions, as well as going over what can be improved and what needs further work.

5.1 Results

Some of the participants had difficulties understanding what was meant with the word inconsistent in question **Q5**. It was further on getting clear that the experimenter also was quite uncertain as of what the meaning of inconsistency in this setting should be, as a variation of different explanations of the definitions of inconsistency were given when needed. There might be reason to believe that some of the participants got confused by the explanation, and mixed inconsistent up with consistent, and therefore answered the opposite of what they might really feel. This might be an explanation to why there is such a large standard deviation in the answers given for question **Q5**, as seen in figure 4.6.

When examining the answers for the questions in regards of the level of immersion, we can see that there are not any results that in themselves are too remarkable. But on the other hand, it is quite interesting to see that *environment 1* scored higher overall than *environment 2* regarding immersion. This can be an indication that the immersion is higher in *environment 1* than *environment 2*, thus indicate that the Quality of Experience might be higher when occlusion is applied, since the overall experience seems to be quite equal in both environments.

We can notice that the standard deviations in answers seem to be larger (or at least not smaller) for *environment 1*, than *environment 2*. This means that there has been a larger variation of answers regarding *environment 1*. A theory of this could be that the main feature of *environment 1*, the occlusion, might have become a victim of its own success. For the occlusion algorithm to work properly, it is dependent on quite ideal conditions. The surfaces need to have good lighting overall in the environment, and the surfaces must have texture features for the algorithm to be able compute the planes. Some of the participants were placed in non-ideal conditions and this affected the *environment 1* experience greatly.

The virtual chair from our experimental application did not have any lighting and texture reactions to its environment implemented, meaning the chair would have looked the same in a fully lit environment, as well as in a dark environment. An easy and dirty fix for this were made by initializing a directional light (within Unity) in the augmented environment to give the virtual chair some artificial and predetermined shadows and light reflections. Participants took notice of this and reported (see section 4.2) that they though this made the virtual object look virtual, and not real.

For the occlusion algorithm applied in *environment 1* to work the best, the users had to use fairly slow and controlled movements, such that the device gets time to compute and apply the occlusion effects. The participants natural speed of moving the device through the space, were a bit too quick, which led to glitches in the occlusion. Unnatural speed of movements broke immersion and the overall quality of experience. Some of the participants had to be told to slow down and give the device a chance to compute the environment, which also leads to a break in the participants natural flow and use of the experimental application.

A fun thing to observe during the experiment, was the participants natural intuitions to explore the app on their own. Some of the participants completed all tasks without even reading through them, just by their own curiosity and imagination. When they eventually read the tasks, some of them just said out loud *"but I've already done this"*. Even though some of the participants completed all of the tasks on their own, there were also a number of participants that needed help doing the tasks properly (e.g. needing more clarification in regards of what they were really supposed to do). The help was tried given in the best manner, but there are possibilities that these interferences introduced biases to the experience.

Another behavior that often occurred was the participants trying to check what happened when they put a part of their body in between the device and the virtual chair. Some took out a hand, a foot and some even tried sitting down (most probably just for the fun of it). The fact that the participants did this might break the intended immersion and experience intended for the participants using *environment 1*, as the occlusion algorithm applied in this experimental application only works for objects and planes, and not for human segmentations. If a participant did this, the virtual chair would appear on top of their limb and not give the intended experience.

As mentioned in section 4.2, some of the participants had troubles understanding some of the English words and sentences in the experimental application and the questionnaires. This led to the need of clarification from the experimenter, which then also can lead to biases and break of flow through the experience.

The duration of the task can seem to be a bit short and not pushing the participants to explore the occlusion elements of the experiment enough. If such an experiment were to be repeated, the tasks the participants must perform should be more complex and challenge them more naturally to see how the occlusion is working in the environment.

5.2 Other Notable Mentions

A factor that did not seem to give any specific or significant differences in the resulting outcome were the participants occupations and age. Most of the participants were studying electrical engineering, but there seemed to be a vast difference in the capabilities and confidence within this group alone, regardless of them all being technical students. The non-technical participants were often perceived as more skilled and performing better than the technical participants, which can indicate that the occupation and background do not necessarily give any advantages.

The experimental design was tested mainly on participants in their twenties, as these types of participants were the most readily available. One could argue that the age variations in the participant group were to narrow and that this should be extended to cover a wider range of ages.

Even though the participants previous experiences in regards of general gaming and other augmented reality applications were mapped, this data has not been further used in the post-testing analysis. It could be that the previous experience can impact the overall experience greatly, so this should be further investigated. The mapping itself is also prone to giving misleading results. The participants own perception of their gaming habits may not be giving true and accurate results, as some of the participants are regular computer gamers spending more time playing ”complicated” games, while other participants might do some ”easy” gaming such as more casual mobile games. Some of the participants thought they never had experienced augmented reality, but it turned out most of them had a lot of experience already without being aware of it, as popular mobile apps such as Snapchat utilizes augmented reality [27] in their experiences.

Due to COVID-19 still being quite the challenge, gathering participants proved to be somewhat hard. In such an experimental design, the more participants, the merrier. Ideally, we would have more participants than what we were able to collect at this point to collect more data. If we had more data, we would most probably have less deviation in our answers, and we would maybe have some more significant p-values, helping us confirm a statistical difference between *environment 1* and *environment 2*.

5.3 Further Work

Due to the limitations of the current experimental system design and other limitations in the field in general, we can suggest the following to be researched further upon if possible with today’s technology.

As Quality of Experience is difficult to measure in a meaningful way, one could try to extend the measurements to other fields, such as studying physical signs (e.g. heart rate, head rotation, facial expressions), in addition to questionnaires. Physical measurements might be less prone to biases and other influencing factors.

There is no doubt that the occlusion algorithm is in its early stages and that itself and the technology surrounding it needs further development. The algorithm is prone to lighting, which is a weakness of only using the RGB camera of the device and the device needs too much time to process the occlusion for it to be natural to use. Adding a time-of-flight sensor to help collecting depth data might ease the processing time and overall result.

Chapter 6

Conclusion

In this article we looked at several aspects that can help the Quality of Experience in a mobile augmented reality application. We were motivated by the need of creating more immersive and engaging training systems for natural hazardous situations. Natural hazards are an increasing problem for our planet, and we wish to be better prepared for these kinds of situations.

We made ourselves familiar with the term immersion, while also familiarizing ourselves with some important design elements making an augmented reality application feel more real. We looked at Quality of Experience and touched upon how this could be measured.

Further on we built a experimental system using Unity with their AR Foundation platform in conjunction with Google's ARCore. In the experiment, participants were first told to fill out a pre-testing questionnaire, mapping their demographic and previous experience using augmented reality. After this, they got to try an experimental application where they either got to try an augmented reality environment where there was an occlusion algorithm in the works, or they got to try an environment with no such algorithm. When completed, the participants filled out a post-testing questionnaire where they answered questions regarding their emotions, usability, immersion and overall experience.

The findings from our results indicate that our hypothesis $\mathbf{H_0}$, formed in chapter 4, is backed, and that we have ***indications of that proper occlusion increase Quality of Experience***. Our findings indicate this in regards of the level of immersion, where *environment 1* seemed to score a bit higher than *environment 2*, while still giving the same overall level of experience. The resulting data were not able to give us convincing results, and therefore we cannot state anything other than this may only be an indication. To get more convincing results, we need to do more elaborate testing, challenging the participants more, without the need of interference and guidance from the experimenter to avoid introducing bias. Due to COVID-19, the number of participants could have been higher. With an increased number of participants, the deviation in answers may converge to an overall smaller value.

There were several factors that broke the participants immersion, among other things the lighting of the virtual object and other glitches and weaknesses of the occlusion algorithm. The occlusion algorithm is heavily dependent on ideal surroundings to work the best, needing overall a great deal of lighting and non-homogeneous surfaces to be able to detect surfaces properly, while forcing the participant to move the device slow and controlled. This makes the application somewhat unnatural to use.

While the occlusion algorithm shows promise, it is still clear that with today's technology and processing power, it needs further work and improvement. Adding additional time of flight sensors could for example ease the strain and compliment the device's RGB camera.

References

[1] (2020). “World of wild waters: Gamification of natural hazards,” NTNU, [Online]. Available: https://www.woww.no/woww/ (visited on 12/14/2020).

[2] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, “Measuring and defining the experience of immersion in games,” *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, 2008.

[3] (2020). “Realism,” Google, [Online]. Available: https://designguidelines.withgoogle.com/ar-design/content/realism.html# (visited on 12/14/2020).

[4] K. Brunnström, K. De Moor, A. Dooms, S. Egger-Lampl, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, C. Larabi, B. Lawlor, P. Le Callet, S. Möller, F. Pereira, M. Pereira, A. Perkis, A. Pinheiro, U. Reiter, P. Reichl, R. Schatz, and A. Zgank, *Qualinet White Paper on Definitions of Quality of Experience*. Mar. 2013.

[5] U. Reiter, K. Brunnström, K. De Moor, C. Larabi, M. Pereira, A. Pinheiro, J. You, and A. Zgank, “Factors influencing quality of experience,” in. Mar. 2014, pp. 45–60.

[6] (2020). “Mean opinion score,” Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Mean_opinion_score (visited on 12/13/2020).

[7] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (mos) revisited: Methods and applications, limitations and alternatives," *Multimedia Systems*, vol. 22, no. 2, pp. 213–227, 2016.

[8] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[9] P. Haynes and E. Lange, "Mobile augmented reality for flood visualisation in urban riverside landscapes," *Digital Landscape Architecture*, May 2016.

[10] C. Keighrey, R. Flynn, S. Murray, and N. Murray, "A physiology-based qoe comparison of interactive augmented reality, virtual reality and tablet-based applications," *IEEE Transactions on Multimedia*, 2020.

[11] M. P. Berge, "Comparing weather data in display environments to measure impact of quality of experience: Virtual reality versus desktop," M.S. thesis, NTNU, Jun. 2020.

[12] (2016). "Will virtual and augmented reality move us into the knowledge age?" TED, [Online]. Available: `https://www.youtube.com/watch?v=2FA-IuDTMjE` (visited on 12/14/2020).

[13] (2020). "Edgar dale," Wikipedia, [Online]. Available: `https://en.wikipedia.org/wiki/Edgar_Dale` (visited on 12/13/2020).

[14] (2020). "Learning pyramid," Wikipedia, [Online]. Available: `https://en.wikipedia.org/wiki/Learning_pyramid` (visited on 12/13/2020).

[15] (2017). "Redefining what is humanly possible with augmented reality," TED, [Online]. Available: `https://www.ted.com/talks/brian_mullins_redefining_what_is_humanly_possible_with_augmented_reality` (visited on 12/14/2020).

[16] (2020). "Unity (game engine)," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Unity_(game_engine) (visited on 12/13/2020).

[17] E. Peckham. (2019). "How unity built the world's most popular game engine," TechCrunch, [Online]. Available: https://techcrunch.com/2019/10/17/how-unity-built-the-worlds-most-popular-game-engine/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAABFNmYnyOo5lqiu9iiAeALtQrZWaO_NzPqEIdaCnvRBoiTPOBI20AWiKg9jw9QvxwrZmPKRN1eK8DsaRCQeO2IUbfSX43vnpVky7Skq9GaNo5WXiYbPlm81nqlbOhxuCi69m6v9B5YfuXoZUDYudeZh7jlIS2vDHjxHJCdbP6U8ex (visited on 12/13/2020).

[18] (2020). "Arcore overview," Google, [Online]. Available: https://developers.google.com/ar/discover (visited on 12/12/2020).

[19] (2020). "Arcore supported devices," Google, [Online]. Available: https://developers.google.com/ar/discover/supported-devices (visited on 12/12/2020).

[20] (2020). "Fundamental concepts," Google, [Online]. Available: https://developers.google.com/ar/discover/concepts (visited on 12/12/2020).

[21] (2020). "Simultaneous localization and mapping," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Simultaneous_localization_and_mapping (visited on 12/12/2020).

[22] (2020). "Depth map," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Depth_map (visited on 12/12/2020).

[23] (2020). "Depth api overview for android," Google, [Online]. Available: https://developers.google.com/ar/develop/java/depth/overview (visited on 12/12/2020).

[24] (2020). "Pixel 5," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Pixel_5 (visited on 12/12/2020).

[25] (2020). "Mann-whitney u test," Wikipedia, [Online]. Available: `https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test` (visited on 12/16/2020).

[26] (2020). "Likert scale," Wikipedia, [Online]. Available: `https://en.wikipedia.org/wiki/Likert_scale` (visited on 12/16/2020).

[27] (2020). "Snap partner summit: Snap's augmented reality platform," Snap Inc., [Online]. Available: `https://www.snap.com/en-US/news/post/snap-augmented-reality-platform` (visited on 12/18/2020).

Appendix A

GitHub Repository

In the following GitHub repository, you can find all of the code for the Unity project making our experimental application, along with the precompiled APK used during the experimental period. It also contains the result analysis with the questionnaire answers in the form of a CSV file, which then was analyzed using a Jupyter Notebook. The Notebook can be found in the repository as well. The repository also contains a wiki with more casual notes that have been made throughout the project period leading up to this final report, along with more lively media showing off the experimental application (video/GIFs).

https://github.com/petrepa/TFE4580

Appendix B

Pre-Testing Questionnaire

Pre-Testing Questionnaire

This questionnaire is made to gather data about the participants chosen to test Peter Remøy Paulsens project thesis application before the actual test has been done.

*Må fylles ut

General information

This section will gather some info about you as a participant.

1. What is you participant number? You will get this from the experimenter *

2. How old are you? *

Markér bare én oval.

- Under 18
- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- Older than 65

3. Gender *

Markér bare én oval.

- Female
- Male
- Prefer not to say

4. What is your occupation? If you are a student, please state your study programme *

5. Do you have any previous experience with gaming? *

Markér bare én oval.

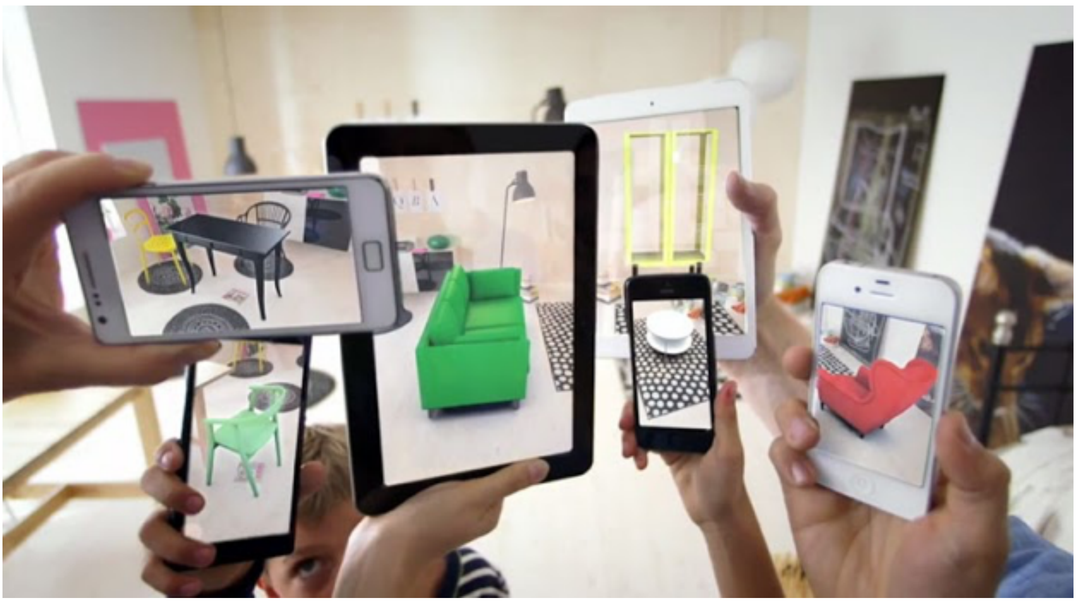
- No experience
- I've tried it before (once or twice)
- I do it irregulary (here and there)
- I do it regularly

Definition of augmented reality

Augmented reality (AR) is an enhanced version of the real physical world that is achieved through the use of digital visual elements, sound, or other sensory stimuli delivered via technology.

(Investopedia, 2020)

Mobile Augemted Reality



HoloLens Augmented Reality

6. Do you have any previous experience with augmented reality? *

Markér bare én oval.

- No experience
- I've tried it before (once or twice)
- I do it irregulary (here and there)
- I do it regularly

7. If you have any experience with augmented reality, which kind of device have you been using?

Merk av for alt som passer

- [ ] Mobile or tablet
- [ ] Head mounted unit (such as HoloLens or Magic Leap)

Appendix C

Post-Testing Questionnaire

Post-Testing Questionaire

This questionaire is meant to gather data regarding the users experience after using the application for th project thesis

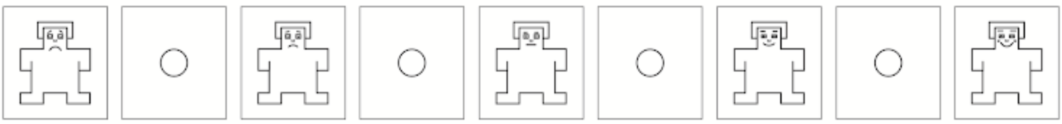
1. What is you participant number? You will get this from the experimenter

2. What environment did you use during the experiment?

Markér bare én oval.

- Environment 1
- Environment 2

3. How unhappy or happy did you feel while using the application?



Markér bare én oval.

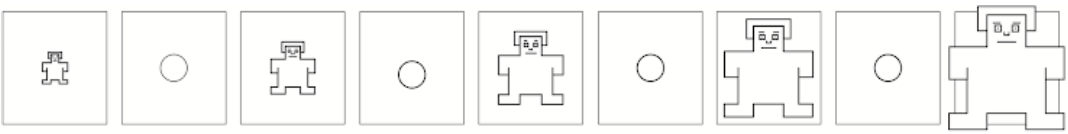
	1	2	3	4	5	6	7	8	9	
Unhappy										Happy

4. How calm or excited did you feel while using the application?

Markér bare én oval.

	1	2	3	4	5	6	7	8	9	
Calm	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Excited

5. How little or much control did you feel you had while using the application?



Markér bare én oval.

	1	2	3	4	5	6	7	8	9	
No control	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Full control

6. I thought the application was easy to use

Markér bare én oval.

- ⬭ Strongly Disagree
- ⬭ Disagree
- ⬭ Neutral
- ⬭ Agree
- ⬭ Strongly Agree

7. I thought the application was inconsistent

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

8. I felt confident using the application

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly agree

9. I was immersed into the mobile screen environment

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

10. I felt seperated from the real-world environment

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

11. I did not feel the need to look at anything other than the mobile screen

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

12. The visuals of the virtual object were convincing

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

13. I felt I could control the virtual object

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

14. The virtual object blended nicely into the environment

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

15. The object disappeared behind other real world objects

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

16. I felt that I experienced the virtual object, rather than watching it

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

17. I enjoyed the overall experience with the application

Markér bare én oval.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

18. I would like to try the application, or similar applications again

Markér bare én oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

19. Do you have any other comments? Any feedback is welcomed.