Universiteit Utrecht

Finger-tracking based interactions in Augmented Reality

Exploring the usability of finger-tracking-based interactions and the effects of multimodal feedback in mobile augmented reality applications

Author:
Kevin Vriens
Thesisnumber:
ICA-3570479

Supervisor:
Wolfgang Hürst

Master Thesis for Game and Media Technology

October 30, 2013

Abstract

Recent years have seen a rise of mobile Augmented Reality applications. The increasingly powerful mobile phones have not only brought us practical applications, such as direction overlays on a map, but also made way for mobile AR gaming. However, many of the current applications still make use of touchscreen gestures for interaction. In this thesis we delve into finger-tracking-based gestures with the purpose of delivering a more pleasurable and immersive experience to mobile phone users.
In our first experiment we compare a finger-based implementation to a touchscreen-based implementation in a mobile AR board-game, featuring both physical and virtual objects, to test both performance and enjoyability. Based on the findings and issues that emerged from this experiments we decided to take a closer look at the intricacies of finger-tracking and touchscreen interactions in an attempt to enhance the performance of our finger-tracking based system. The outcome of this second experiment suggested that our finger-tracking based system, and others, could be further improved by adding additional feedback. A third experiment was therefor designed to study the effects of multimodal feedback on performance and user perception. With the phone as our only source of feedback, we tested the the combinations of visual, audible and remote haptic feedback, with constant and temporary intervals. The results showed that multimodal feedback in general, and constant visual cues combined with temporary haptic cues especially, can increase user responsiveness when transitioning between interactions.

Acknowledgements

I would like to express my gratitude to dr. Wolfgang Hürst for his ideas, insights and feedback during the course of this thesis. Under his supervision, parts of this thesis have also been published to the MobileHCI 2013 workshop: Designing Mobile Augmented Reality. A special thanks goes out to Joris Dekker and Casper van Wezel, who both inspired me and greatly helped me understand and setup the software needed for this project. I would also like to thank all the people that participated in the user-studies for their cooperation and feedback.
Finally, I would like to thank my family, friends and girlfriend for being patient and supporting me till the very end.

Contents

Chapter 1

Introduction

1.1 General introduction

Recent years have seen a huge increase in the percentage of the population that uses mobile phones. An increasing number of these phones are the so called smartphones, which are mobile phones build on a mobile operating system, enabling the manufacturers to create more advanced capabilities and add features such as media players, cameras, GPS and mobile internet. In early 2013 the worldwide sales of smartphones exceeded those of 'normal' mobile phones [16] and as of July 18th 2013, 90 percent of global handset sales are attributed to the purchase of Android and iPhone smartphones[17]. Seeing the countless possibilities in terms of applications, social media and gaming, it is of no surprise that most of the early adopters that embraced smartphones were young adults.

During the emergence of smartphones, the high-resolution touchscreen made its appearance. Even though multi-touch human-computer interaction had already been described by Sears et al. in the 1990's [18], it was not until the smartphone that the general public came in contact with touchscreens and started using a new way of interacting with their phones, in the form of touchscreen gestures.

Due to the growth in computational power and the inclusion of a camera, Augmented Reality found its way to smartphones. Augmented Reality(AR) is a live, direct or indirect, view of a physical, real-world environment whose elements are augmented (or supplemented) by computer-generated sensory input such as sound, video, graphics or GPS data[21] and according to the widely acknowledged definition by Azuma[19] it has to combine the real and virtual world, be interactive in real time and be registered in 3D. For a long time, augmented reality existed mostly by combining the computational power of a stationary or wearable PC or PDA combined with a webcam, camera or headmounted display, but in 2003 Siemens released the SX1, which came with the first commercial mobile phone AR game called Mozzies.

Here, mosquitoes are superimposed on a live video feed from the camera. Aiming is done by moving the phone around so that the cross-hair points at the mosquitoes, and firing is done by pressing on the touchscreen or pressing a button[25].

Figure 1.1: Mozzies

Fast-forward to 2013 and we see a plethora of AR applications being developed for mobile phones, such as Wikitude, an AR-based GPS navigation systems which overlays directions onto the real world images, or Arhrrrr!, a game which displays highly detailed content on a game-board which functions as a marker that can be tracked by the phone. In the game ARhrrrr!, the users control what happens in the virtual world by making use of the buttons and touchscreen of their mobile phone, but they have an additional way of interacting with the virtual world and that is by placing certain physical objects on the game-board, which can also be tracked.

Figure 1.2: Wikitude

Figure 1.3: ARhrrrr!

This is a novel way of interacting, which bridges the physical and virtual world and increases immersion. However, because the users still control some parts of the virtual world by buttons on the phone, they constantly have to switch between the board on the table, and the phone in their hands, a direct effect of having two distinct places to deliver input to the system, which we perceive as a negative influence on immersion and performance.

In this thesis we will explore the use of hand-gestures performed in front of, and tracked by, a smartphone's camera in an attempt to find a more immersive and enjoyable way to interact with AR environments. From this point forward, we may refer to these gestures and the interactions one can perform with them as Finger-Tracking-based interactions or FT interactions for short.

1.2 Goals

The main goal of this thesis is to:

- Explore the usability of finger-tracking-based interactions in mobile augmented-reality environments and improve them

This main goal will be divided into sub-goals:

- Determine the performance and user experience of finger-tracking-based interactions and compare them to touchscreen interactions
- Examine the finger-tracking and touchscreen interactions and compare them to identify the strengths and weaknesses with regards to performance
- Determine the effects of multimodal feedback on finger-tracking-based interactions

1.3 Overview

As the initial sub-goal, we investigated:

1. Determine the performance and user experience of finger-tracking-based interactions and compare them to touchscreen interactions

The results of this study, which was done prior to this thesis and is summarized in chapter 2, suggested that while the users experienced the FT interactions as positive, the performance differences between FT and touchscreen interactions were further apart than expected, which leads us to the second sub-goal:

2. Examine the finger-tracking and touchscreen interactions and compare them to identify the strengths and weaknesses with regards to performance

To identify the performance issues discovered after investigating sub-goal 1, we did a followup experiment, which is described in Chapter 3, that delves deeper into the performance aspect of the interactions. In order to overcome the issues identified in this second study, we aimed at improving feedback to the users during and after interactions, leading us to the third sub-goal:

3. Determine the effects of multimodal feedback on finger-tracking-based interactions

Chapter 4 describes an experiment in which we explored different kinds of feedback for FT interactions in an attempt to solve the issues described in Chapter 3. Finally, Chapter 5 will list our conclusions and also our recommendations for future work.

1.4 Software

At the start of this thesis a number of students at Utrecht University were working on projects involving augmented reality on mobile phones and as such some of the groundwork had already been done on the Android OIS. The most powerful phone available at that time also ran Android and therefore the decision was made to use Android.

The code for the gesture-based interactions is based on [6] and is written using the Android SDK. Specifically the code for retrieving and decoding images, and for detecting and tracking the finger-markers was used, albeit modified for better performance specific to the phone's camera. For tracking the game-board we used the QualComm AR SDK. This toolkit uses natural features to keep track of an image's position relative to the phone. Not only was it faster than other tracking systems at the time, it was also able to track

a range of images whereas the others were only able to track square matrix markers. However, the biggest benefit of this toolkit is that, if supplied with an image with enough high contrast points, the marker-tracking even works when the image is partially occluded. This made it by far the most robust tracker and as such the best option for us.

Using these SDK's also had some (undocumented) limitations; we could only use a limited amount of textures on planes, and there was no option to change textures or colors of loaded objects during runtime, which prevented us from having arguably the most optimal visual feedback during the final experiment.

1.5 Hardware and attributes

The device used in all experiments described in this thesis is a Samsung Nexus S with a 5 megapixel camera and a 1 GHz Cortex-A8 processor. The physical game-board we used consists of two A4 pieces of paper taped together at the long ends, depicting a black and white graphic of rocks, which makes it a picture with a very good distribution of high contrast points. This graphic is used by the Qualcomm AR toolkit as reference point for the AR world. The markers we used are made from green and red plastic or paper. For the last experiment we used a blue medical glove to further enhance the robustness of the finger-tracking.

Chapter 2

Experiment 0: Touchscreen versus Finger-tracking

2.1 Related work

The first part of this thesis summarizes an experiment, which conclusions became the motivation for this thesis. This experiment, which we will call Experiment 0, builds on the thesis of C. van Wezel, in which he designed, implemented, evaluated and discussed several different interaction concepts for mobile AR applications to determine the feasibility of finger-based gestures[6]. In the first half, he compared three ways to do interactions: by using the touchscreen, by using the accelerometer and compass of the device, and by using gestures in mid-air in front of the device's camera. While the performance of the mid-air gestures was the worst of the three, users rated it highly in terms of fun, engagement and entertainment. He also calculated the optimum distance for interactions to occur in front of the camera, and while most of the participants did not perform interactions within this range, he showed that the ones that did had better control. These results led him to to a follow-up study where he made use of a game-board to compare different forms of finger-based interactions in front of the camera. He found that the accuracy of finger-based gestures was good, but when it comes to fast interactions where high accuracy is demanded, it is not as good as touchscreen or device gestures. Nevertheless the feedback from participants was very positive with regards to fun and engagement and as such he concluded that finger-based gestures would be suited for entertainment and leisure applications. With regards to AR in combination with the game-board, he mentions that users were not only more positive about this setup, because of the connection to the real world and the context it gave to the interactions, but that it also has the potential to increase the robustness of the interactions, by giving the users a plain to perform the interactions, such that the hand is farther away from the camera.

2.2 Goals and expectations

For Experiment 0, we designed and implemented one of the future work recommendations of [6]. We created a system using color tracking by implementing one of his interaction techniques and we set up a comparative user study which we tested in an Augmented Reality board-game setting (a game) featuring both virtual and physical objects. In one part of the study the virtual objects were manipulated by the touchscreen and in the other part by gestures in front of the camera. The interactions include picking up/selecting, moving and dropping/deselecting. Sub-goal one of this thesis reads:

- Determine the performance and user experience of finger-tracking based interactions and compare them to touchscreen interactions

The goal of this experiment was to evaluate performance, usability and enjoyment of both FT-gestures and touchscreen gestures in a game-like setting, instead of task-based setting, and compare them to each other. In both parts of the study users also had to manipulate physical objects, while looking at them with the camera. Adding these physical objects was done for a multitude of reasons. First, when we envisioned an AR board-game we saw it as a typical physical game enhanced by augmented reality, not a purely virtual game. Second, by adding physical objects on the game board, users have an extra reason to interact as close to the game board as possible, which should make the interactions more robust.

We expected that doing both virtual and physical interactions close to the game-board, as is the case with FT-gestures, would be faster than interacting with physical objects close to the board and interacting with virtual objects on the touchscreen of the phone, as users would not have to switch between the two plains constantly. With regards to enjoyment, we expected the FT-gestures to outperform the touchscreen gestures, as FT-gestures have the potential to enhance immersion by bridging the virtual and physical world and for many users it would be a novel way of interacting.

2.3 Setup

In this experiment we decided to limit ourselves to translation(selecting, moving and deselecting) as it is the most used form of interaction with objects in board-games. While we would have liked to study all thee in a game setting, it is important to note that each interaction brings its own unique set of problems with it. For instance, scaling objects in an intuitive manner uses the same kind of finger-gestures as dropping an object does

and because we do not get any depth information from the phones camera it would only be possible to rotate objects on one plane, which would in turn not make for an interesting game.

For the physical object described in the previous section we chose a blue coin. While for instance a pawn would be easier to pick up we would have ran into the problem that virtual objects can only be drawn on top of the real-world camera image. A virtual ball rolling behind a physical pawn will be drawn on top of it (1 in Figure 2.1 depending on the pawns size, instead of behind it (2 in Figure 2.1). Because we didn't want such graphical anomalies we chose the coin.

Figure 2.1: Graphical behavior

The game-world has as its center a virtual game-board consisting of 8 by 6 tiles projected onto the physical game-board as can be seen in Figure 2.2. The upper, left and right side of the board have walls with gates on them. It is through these gates that yellow and blue balls will appear and roll to the other side of the board.

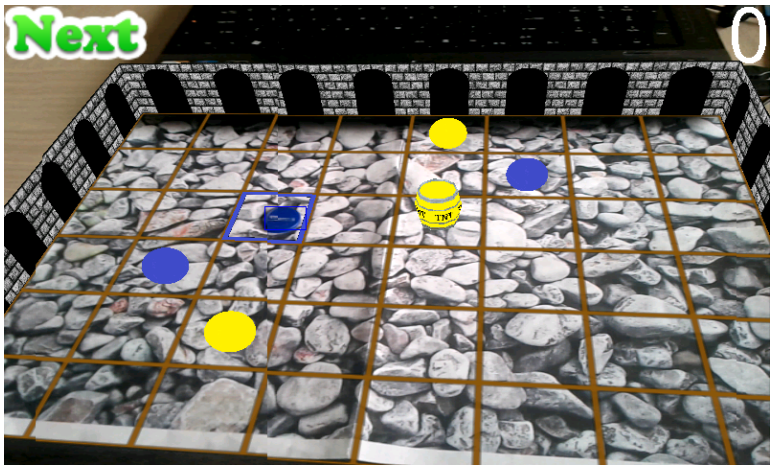


Figure 2.2: The game-world - a virtual world projected on a physical game-board

The game features a virtual yellow barrel and a physical blue coin. These objects both occupy a tile on the board and can be moved around to catch the balls. If a ball is not caught in time and rolls over the boundaries of the game-board, the ball is destroyed but no point will be added to the score. In the game the blue balls are caught when the physical blue coin is on the same tile as the balls. This coin can only be controlled in the physical world by pushing it around or picking it up and placing it somewhere else. The yellow balls are caught when the virtual yellow barrel is on the same tile as the yellow ball. In the touchscreen part of the experiment this barrel will be the object that has to be controlled on the touchscreen, while in the finger-tracking part it is controlled by FT-gestures. To translate objects we have decided to use the pick up and release method, which we will describe below, as [6] showed this was a good method for translation on a board. This method is both fast and very intuitive as it uses the same hand- and finger-movements for interaction with virtual objects as one would use for interaction with physical objects.

2.4 The 'pick up and release' method

We can differentiate three interactions while translating a virtual object. Selecting (also called grabbing or picking up) an object, moving an object and deselecting(also calling placing or dropping) an object. Interactions are performed by moving fingers in front of the camera. All of the interactions

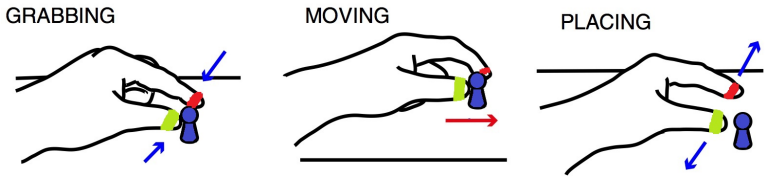


Figure 2.3: Hand gestures

make use of a green and red marker attached around the fingertips of a user's thumb and index-finger, so that the marker can be seen from the front, sides and back of the finger. It does not matter whether the thumb is red and index-finger is green or the other way around. When a marker is seen by the camera, an open square of the same color is drawn around the marker, this makes it easier for users to see if by any chance a marker is lost, for instance due to occlusion. The reason we use two colored markers is that our fingers operate in a 3-D space and the camera only shows a 2-D view of this space. Considering we are not getting any depth information, having only one marker would make it conceptually more difficult to differentiate between selecting an object or just passing over it as the system has no way to determine the intent of the user. To make this distinction clear, another

condition needs to be met, which is easiest to achieve by using an additional marker. In this case, as long as the distance between the two markers is greater than a set minimum, the object will not be picked up even when hoovered over.

2.4.1 Picking up / Selecting

To pick up an object the users have to bring their thumb and index-finger close enough to the object so that both the markers collide with the (invisible) bounding-box of the virtual object. If this is the case the object is immediately picked up and centered between the red and green marker.

2.4.2 Dropping / Deselecting

To deselect an object users are required to:

1. Move the markers a minimum distance from the bounding-box of the object
2. Move the markers a minimum distance apart from each other
3. Keep both markers visible

When these three requirements are met the object will be dropped halfway between the red and green marker. All of these requirements are implemented for specific reasons. Requirement 1 is the most basic and exist solely to notify the system that the user wants to deselect an object.

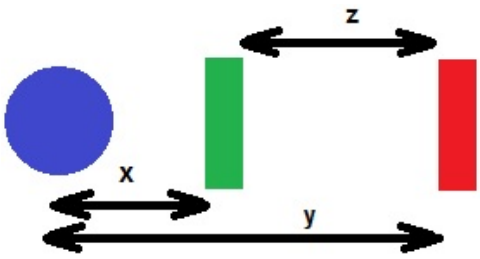


Figure 2.4: Requirement 2

Requirement 2 is implemented to circumvent a limitation that arose during translation in the original work of [6]. When moving virtual objects at high velocities, a discrepancy can occur between where the object is rendered and where it should be(halfway between the red and green marker) due to frame-rate. When this happens the system believes the markers have left the bounding-box and consequentially deselects the object at the wrong location. To prevent this from happening, we devised this additional requirement, which makes sure that as long as both markers are not too distant from

each other, the object will not be deselected, even when moving objects at high velocities. To illustrate, see Figure 2.4.In this case a user moves his hand so fast from left to right, that the red and green markers are both detected at the right side of the blue object, as the object still has to be rendered in between the markers. In this case however, the distance y (red marker) and the distance x (green marker) both meet requirement 1, in that they are greater than the minimum distance from the object's bounding-box, which would cause this object to be deselected. By adding the requirement that the markers have to be more than distance z apart, the object will not be deselected. The only thing that had to be taken into consideration when implementing this was that the actual distance to meet requirement 2 would not make it too difficult to meet requirement 1.
The reason for requirement 3 is to prevent an object from deselecting when a marker is accidentally lost for instance due to lighting issues or by being outside the field of view of the camera, for instance when trying to deselect an object at the edge of the camera view. In these fringe cases it is up to the user to operate the device such that the markers stay in the field of view.

2.4.3 Moving

To move a virtual object it has to be selected first. Moving the fingers in front of the camera moves the object along, always in the center of the red and green marker. An object can be moved as long as it is selected. When one of the markers is lost, the object will stay at it's current position.

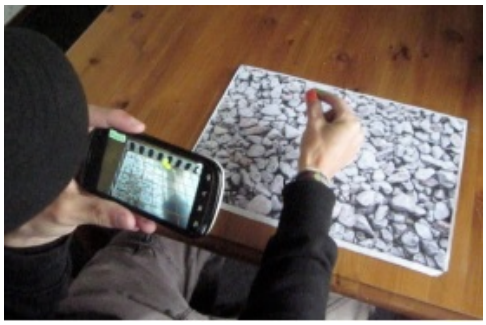


Figure 2.5: Photo from the actual experiment

2.5 Discussion

The sub-goal we addressed in this experiment was:

- Determine the performance and user experience of finger-tracking-based interactions and compare them to touchscreen interactions

The experiment showed that the performance of FT interactions was sub-par to that of touchscreen interactions as users were faster and more accurate with the touchscreen, while we expected the users to be faster with gestures. We identified a number of reasons that could cause this. Conceptually, moving the hand from one side of the game-board to the other takes a bit longer when not doing it on the touchscreen, as users have to cross bigger distances in the real world(the distance of the whole game-board versus the size of touchscreen). This however is only an issue when comparing to the touchscreen; if instead of a virtual object, an actual object was being used to interact with, like in a typical board-game, the distance traveled would be the same. Another reason could be that users are not accustomed to these kinds of interactions. During touch-based interactions, if users touch the screen where an object is they intuitively know the object has been selected, as they have been selecting elements on touchscreens a large number of times during their lives, and as such they start to move the object immediately. Note that this is not just caused by the tactile feedback of touching the screen, as touching the screen at a position without an object gives the same feedback. During the gesture-based interactions, when users place their fingers in such a way that the object should be picked up, instead of already starting to move the object they wait until they have visual conformation that the object is picked up and then start to move. The same goes for when they want to drop an object. From here on out we will refer to this potential delay as wait-time.

User feedback shows us that users experienced switching between the game-board and touchscreen as confusing, annoying and unwanted. FT-gestures however were experienced as significantly more fun than touchscreen gestures and users described the FT-gestures as 'immersive', 'intuitive' and 'cool'.

One thing we noticed was that between the first and second part of the experiment users vastly improved their speed. The most likely reason for this is that, because this way of operating and interacting with a system is so new, users get significantly better within a short period of time during their first use.

A problem brought to our attention by the users was the difficulty of

picking up the physical object while looking through the device. The reason for this is that the screen of the device is unable to provide depth information to the users. Gauging the depth by the size of the board and your hand is possible but is complicated by the fact that the size depends on how far the device is held from the board and the eyes. The users who suffered from this the least were the users that held their hands closest to the board, subsequently these were also the users who scored the best in the experiment. This again shows the importance of having, and forcing, the users' hand close to the game-board.

Chapter 3

Experiment 1: Examining interactions

3.1 Goals and expectations

As we've seen in the previous experiment finger-tracking based interactions are not as fast as touchscreen interactions. While the reasons given for this discrepancy hold true, the magnitude of the difference between touchscreen and gestures and the large difference between individual users were still surprising to us and therefor we decided to further investigate what causes this. With this goal in mind we set up our second sub-goal to:

- Examine the finger-tracking and touchscreen interactions and compare them to identify the strengths and weaknesses with regards to performance

By creating a test with a longer duration, we also have the options to investigate a number of the issues we discovered in Experiment 0, with regards to improvement and fatigue. Our goal can then further be specified:

1. Identify if users get significantly better within a short period of time
2. Identify bottlenecks by measuring and comparing FT-gestures and touch-based gestures
3. Identify if users experience fatigue after a longer period of using the system

Following the results from Experiment 0, we do expect that users improve the speed of their interactions within a short period of time, but we do not know if it will prove to be a significant difference. For point 2 our expectations are that moving and deselecting will take longest, because both these interactions are conceptually slower, and also because these measurements would pertain the wait-time as mentioned in the conclusions

of the previous experiment. If this wait-time does prove to exist, further research on how to diminish this duration is needed if we want the interaction times of FT interactions to come closer to those of touchscreen interactions or maybe even more important, to those of interactions with physical objects. As for point 3 we expect users to get fatigued after a while but if it is within the duration of our experiment remains to be seen. However, seeing as this experiment will take about 45 minutes consisting of constant interactions, if users do not experience fatigue after this, it stands to reason that an average length board-game incorporating the same interactions, but used less frequently, will have no negative effect either.

3.2 Related work

To avoid three separate related work sections we have combined the related work in Chapter 4.

3.3 Setup

The experiment consists of two parts with a small questionnaire at the end, containing questions about their own feel of improvement, ease of use of the interactions and fatigue (See Appendix A). In the first part of the experiment the users will play an extended version of the game created in Experiment 0. To find out if the users get better, a ratio r is calculated measuring how good a user is doing. This ratio is simply the percentage of caught balls, measured each time 8 balls have passed, so after 8, 16, 24 etc. If the users catch less than 25% of a set of 8, the speed at which the balls move will be decreased, do the users catch more than 75% of a set the speed will be increased. After multiple sets this creates a list of ratio's and speeds. By examining this list, we can see if the ratios pertaining to a certain speed improve, or if the speed at which a user manages to stay at a certain ratio improves. For example, a male user's data might like look this:

Speed	5	6	7	7	7	8	7	7	8	8
Ratio	0,88	0,75	0,5	0,63	0,75	0,13	0,63	0,75	0,25	0,38

Speed	8	7	7	8
Ratio	0,13	0,63	0,88	0,38

From this data we can see that at the start the user only improved, until he arrived at the speed setting '8', at which point the amount of balls he caught decreased substantially and he dropped back to 7, where he consistently gets over 50% of all balls. As we see he keeps making it into speed 8 but his ratio does not seem to climb much anymore. We can now determine if the user improved significantly by comparing the first set to the last or by comparing the first 2, 3 or 4 sets combined to the last 2, 3 or 4 sets combined.

In the second part of the experiment we will locate bottlenecks by measuring and comparing FT interactions and touchscreen interactions. We will do this by:

1. Measuring the time it takes to select objects
2. Measuring the time it takes to move objects
3. Measuring the time it takes to deselect objects
4. Measuring the time it takes to select, move and deselect an object
5. Compare the times measured for FT-gestures with the times measured for touchscreen gestures and identify bottlenecks

To make accurate measurements, some interactions are started when users hear a sound. The reaction time to sound is factored into the measurements as explained in the results section.

The way we plan on identify bottlenecks is twofold; One way is looking for large discrepancies between FT and touchscreen. For instance when considering a total round of interactions involving selection + moving + deselection, selection during FT-gestures might take 10% of the total time whereas selection during touchscreen gestures might take 70%. In this case there is a clear discrepancy which reasons might prove important into solving the issue of why touchscreen gestures are found to be so much faster than FT-gestures. Another way to look for bottlenecks is to look at FT or touchscreen individually. For instance, when looking at a total round of interactions involving selection + moving + deselection, selection might take 15% of the time while deselection might take 55%. Again there is a clear difference here, which reasons can give further insight into the aforementioned issue.

The total duration of the experiment is 45 - 60 minutes and is carried out by 24 volunteers. The age of the volunteers ranged between 21 and 34, with an average age of 24.3. The volunteers consisted of 1 female and 23 males. None of them have been involved in any way in the first study or had experienced our system before. Each volunteer executes part one first, as it also serves as practice, and part two second. In the second part the touch- and FT-based levels are counterbalanced.

3.4 Procedure and data

Procedure
Start test
Explanatory test:

> The goal of this test is to select, move and/or deselect objects as fast and correct as possible. The experiment begins when you press 'OK'

Level1:

> After a while an object is shown at a random position, a sounds is made and a timer starts. As soon as the object is selected the timer stops. Repeat 10 times

Level2:

> An object and a 'goal location'(tile) are shown. When the participant selects the object a sound is made and a timer starts. As soon as the object hoovers over the correct location the timer stops. Users have to hoover over a position 500 milliseconds, this time will be subtracted. Repeat 10 times. Keep track of number the of times the object is deselected before it reaches the goal location and restart the try if it has.

Level3:

> Identical to Level2, but the goal location is twice as far

Level4:

> An object has to be selected by the user. When completed, a sounds denoting that the object has to be deselected again at the same position. At that moment a timer will start. When the object is deselected correctly the timer will stop. Repeat 10 times. Keep track of number of times the object was deselected incorrectly and restart the try.

Level5:

> Do full cycle of fixed position select+move+deselect and time them individually as well as the total time of the cycle. Repeat 10 times.

End Test
Questionnaire:

> Users have to fill in a questionnaire.

Data
At the end of the experiment we will average all the data per participant of level 1, 2, 3 and 4, add them and look for discrepancies with the times of level5. We will also look at the times of touchscreen and FT for patterns and discrepancies, to see what action takes how long, in which interaction-mode.

3.5 Results

Unfortunately, due to unforeseen performance issues, brought on by an update to the phone's camera, right before user studies, the first part of the experiment did not yield any useful quantitative data, such as ratios, with regards to user improvement. However it did succeed in give users 30 minutes of time to practice interactions non-stop.

The second part of the experiment did yield some interesting quantitative results. Because the participants start their interactions based on a sound, calculations are made using an average reaction time to audio of 150ms, which falls within the range agreed on by many researchers[14].

3.5.1 Handling outliers

The size of our datasets meant we could plot our data and manually look for outliers. We did this using SPSS and after applying a Kolmogorov-Smirnov test it showed that the data in sections 3.5.2 to 3.5.7 was all normally distributed. For determining statistical significance we then used the paired t-test, the following data is all statistically significant with $p < 0.05$, unless formulated otherwise.

3.5.2 Selection

Averaged over all participants we found that selection is 2.92 times as fast with the touchscreen, with an average difference of 1587.5 ms (TS = 904.3ms, FT = 2672.2ms, effect size = 2.8, t = -12.794).

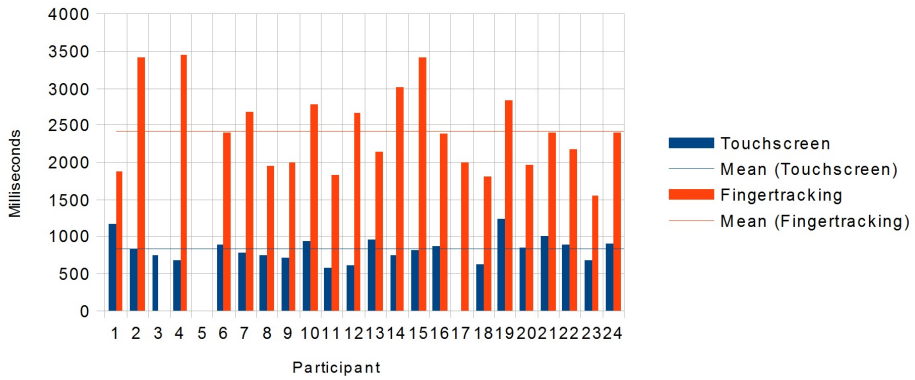


Figure 3.1: Average 'selection time' per user

3.5.3 Move

Averaged over all participants we found that moving is 4.62 times as fast with the touchscreen, with an average difference of 1584.4ms (TS = 437.2ms, FT = 2021.6ms, effect size = 2.7 t= -13.227).

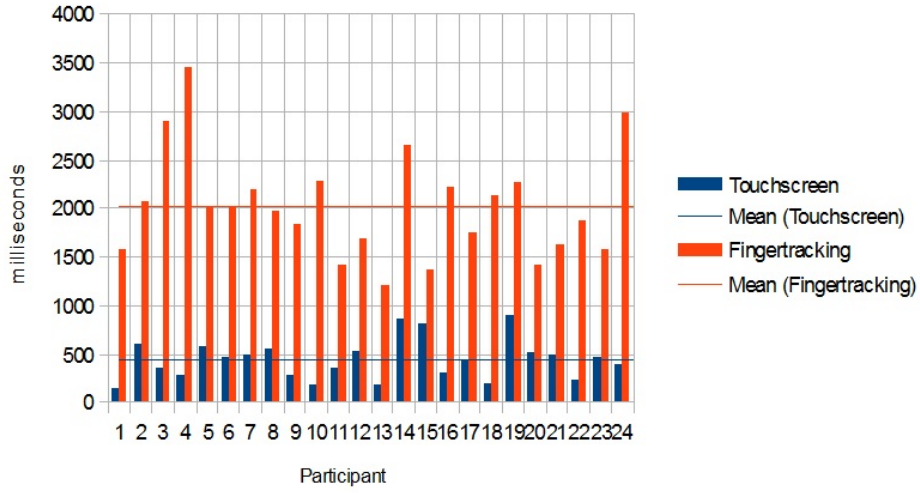


Figure 3.2: Average 'move time' per user

3.5.4 Move farther

Averaged over all participants we found that moving farther is 4.64 times as fast with the touchscreen, with an average difference of 2159.4ms (TS = 593.1ms, FT = 2752.6ms, effect size = 2.7 t = -12.565).

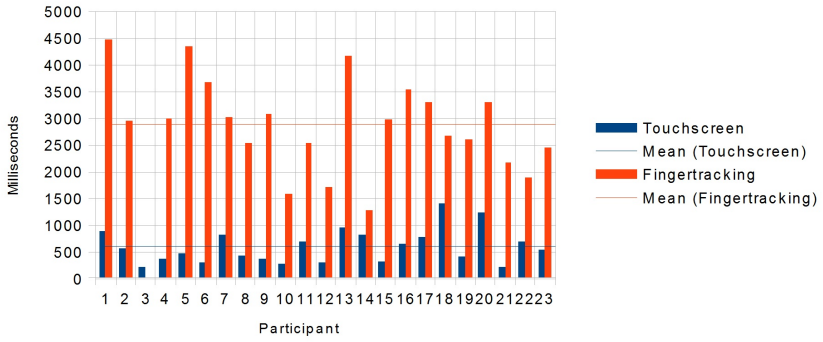


Figure 3.3: Average 'move farther time' per user

3.5.5 Deselection

Averaged over all participants we found that deselection is 2.25 times as fast with the touchscreen, with an average difference of 893.2ms (TS = 716.8ms, FT = 1610.0ms, effect size = 2.4 t= -10.541).

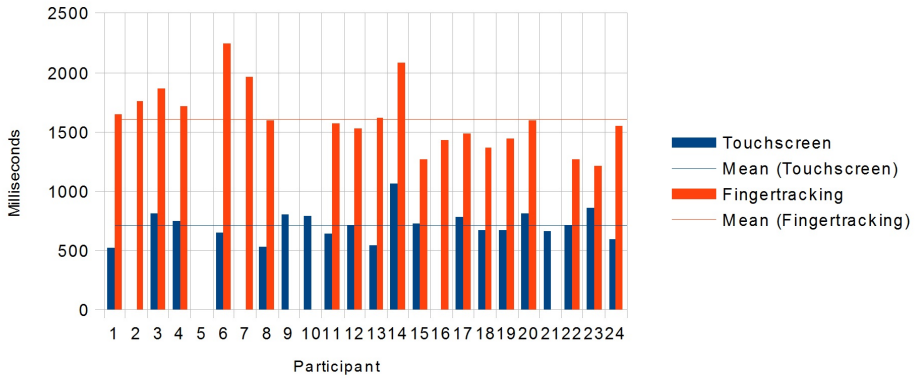


Figure 3.4: Average 'deselection time' per user

3.5.6 All-in-one

Averaged over all participants we found that all-in-one is 2.48 times as fast with the touchscreen with an average difference of 2797.8ms (TS = 1887.6 ms, FT = 4685.4 ms,effect size = 2.3, t = -10.620).

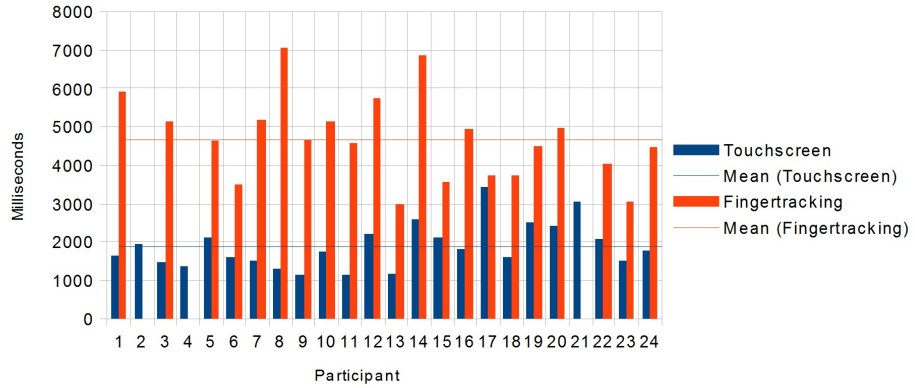


Figure 3.5: Average 'all-in-one time' per user

To determine statistical significance of the next tasks, which are not normally distributed, we used the Wilcoxon signed rank test as described in[10] and found they are all statistically significant with p <0.05.

3.5.7 All-in-one selection

Averaged over all participants we found that all-in-on selection is 1.53 times as fast with the touchscreen, with an average difference of 407.9ms (TS = 774.1ms, FT = 1182.0ms).

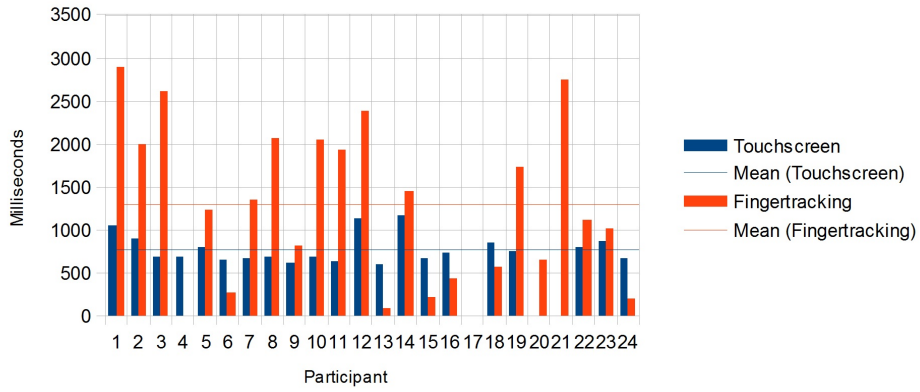


Figure 3.6: Average 'all-in-one selection time' per user

3.5.8 All-in-one move

Averaged over all participants we found that all-in-on move is 3.58 times as fast with the touchscreen, with an average difference of 1008.2ms (TS = 307.8ms, FT = 1316.0ms).

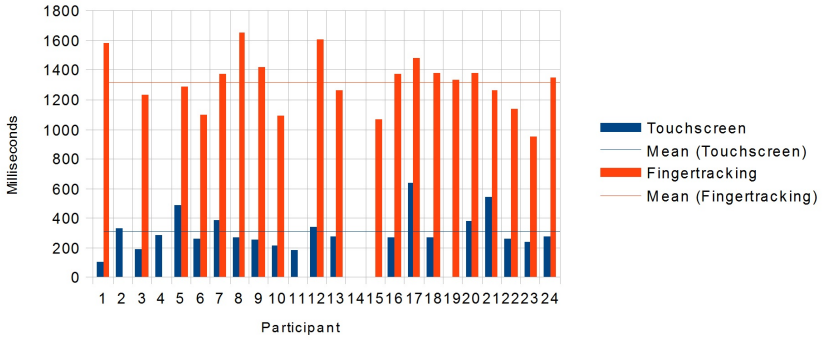


Figure 3.7: Average 'all-in-one move time' per user

3.5.9 All-in-one deselection

Averaged over all participants we found that all-in-one deselection is 3.44 times as fast with the touchscreen, with an average difference of 1585.5 (TS=649.2ms, FT=2234.7ms).

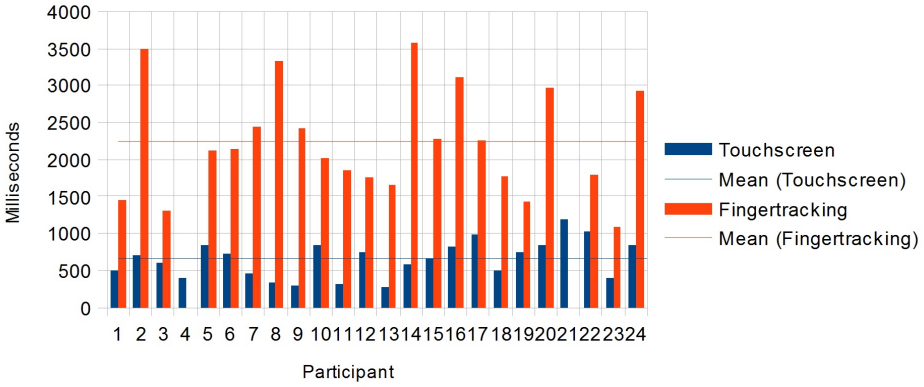


Figure 3.8: Average 'all-in-one deselection time' per user

3.5.10 All-in-one interactions versus separate interactions

The differences when adding all the separate parts of the interaction together(select,move,deselect) compared with the all-in-one-interactions are not significant for both the touchscreen and FT interactions.

3.6 Questionnaire

In the questionnaire the users were asked how they would rate their own improvement, or increased control of the interactions, over time. On average they gave themselves a 7.6 out of 10 which shows us they themselves thought that their control over the interactions increased greatly(Figure 3.9).

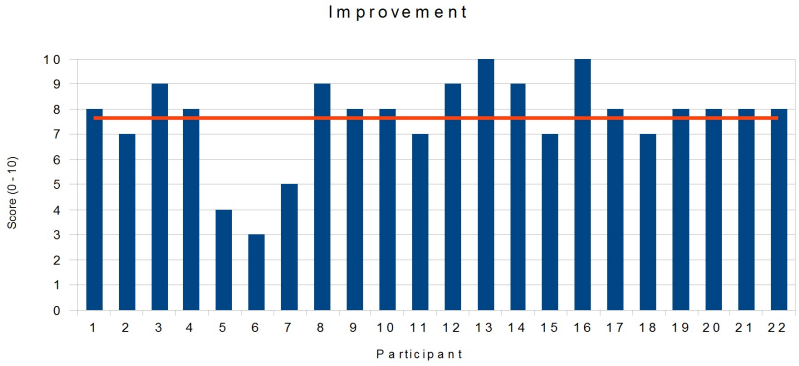


Figure 3.9: Improvement rating

When asked about the feeling of fatigue they gave themselves on average a 6.3 out of 10 after a 45 minute period of constant interactions(both touch-based and FT-based). As one can see in Figure 3.10, the scores users gave range from 1 to 9, meaning some did not experience any kind of fatigue at all and some really started to feel tired in their arms. None of the users reported feeling pain.

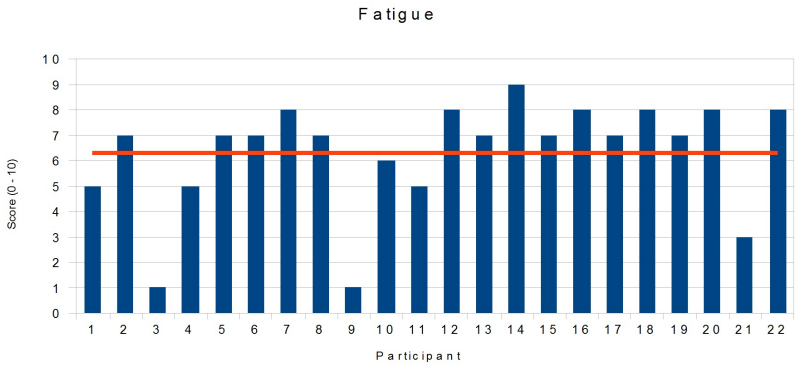


Figure 3.10: Fatigue rating

Figure 3.11 shows the difficulty rating. The users' experience of difficulty of the FT-based interactions, averaged over all users, is 5.5, just above neutral. When asked how to improve the current system most comments were aimed

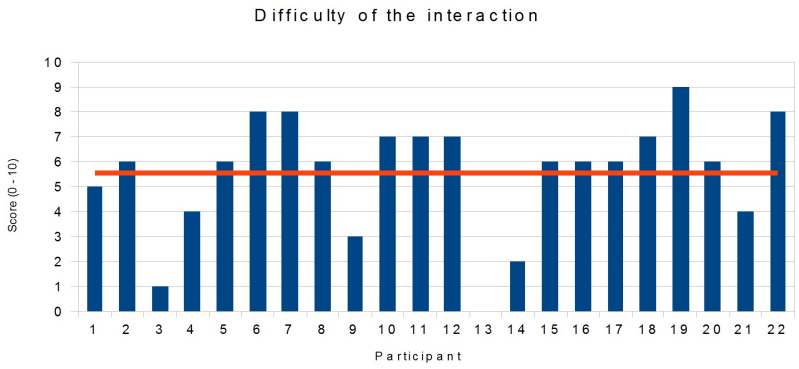


Figure 3.11: Difficulty rating

at the inconsistent frame-rate of the first part of the experiment. A number of users also pointed out that the red marker sometimes did not get tracked properly under certain lighting conditions, such as direct bright light, and how another tracking algorithm may improve the robustness of the system even more.

3.7 Observations

While observing the participants we noticed behavior similar to the Experiment 0. When doing the interactions for selecting and deselecting, most users slowed down. Instead of making smooth, swift finger-movements, users made careful, slow movements, as if they were afraid to not select or deselect the objects correctly. When they had selected or deselected the objects, they hesitated and slowly went on to the next interaction. After moving their hand slowly and noticing they did the previous interaction successfully, they accelerated again. Having observed this in both experiments, the next section tries to determine whether or not this is also evident in the data and if so, how much time is spent waiting. The only way to determine this is by approximation because there is no precise way to measure how long a user waits between interactions as the only points during the interactions we can get data is at the exact point of selection and the exact point of deselection.

3.7.1 Waiting before dropping

Even though we can not exactly calculate how long it takes between moving an object to a desired location and dropping it, we can calculate that there is a clear difference between when users act upon a sound or when user act when reaching a certain location, which does not involve sound. When users

were asked to drop an object when they heard a sound, for touch-based and FT-based interactions it took them 709.163ms and 1581.367ms respectively. When they were asked to drop it as fast as they got the object to a certain location it took them 649,229ms and 2347.512ms respectively. What's interesting here is that for the touch-based interactions users were almost as fast when acting upon the sound as when acting without hearing the sound (1,09 times faster without sound), but for the FT-based interactions users were 1.484 times slower when acting without hearing the sound. In other words when users have moved an object to a certain location using FT, it takes them 1,48 times as long to drop an object when solely relying on their own eyes and intuitiveness compared to when they get feedback in the form of a sound.

3.7.2 Waiting before moving

We take a similar approach for the wait-time between selecting an object and moving it. Again, we can not calculate the exact time but we can look at the difference between the users' speed when acting upon a sound or when acting without sound. Here we look at the time it takes to move an object one square. When users heard a sound after they had selected an object, it took them, averaged over all users, 851.5ms. When the users did not get to hear a sound, it took them 1357.2 ms to move one square. If we factor in the reaction time of sound($\sim$ 150ms) and sight($\sim$ 190ms) this becomes 701.5ms for sound and 1167.2ms without sound. By dividing these number we see that when reacting to sound users are 1.66 times as fast.

3.8 Discussion

3.8.1 Identify if users get significantly better within a short period of time

Even thought we did not get any useful quantitative data from the first experiment regarding the users improvement with FT-based interactions, we do see that users rate their own improvement as good with an average of 7.6 out of 10, suggesting that they do get better in a short ($\sim$ 30 min) period of time.

3.8.2 Identify if users experience fatigue after a longer period of using the system

From the fatigue score, with an average of 6.3 out of 10, we can see that symptoms like tiredness slowly start to manifest with most users during the 45 minutes of constant interactions. A number of users explicitly mentioned that holding the arm with the mobile phone at the same angle for too long

caused some tiredness. One user suggested how a tripod might solve this and free up an extra hand for interactions but while true, this collides with our ideas of using just a mobile phone and being able to freely move around the game-board or game-world.

3.8.3 Identify bottlenecks by measuring and comparing FT-gestures and touch-based gestures

The biggest bottleneck we can see from the observations (and approximation of the data) is the time it takes to move an object when comparing FT-gestures to touchscreen gestures. While a part of this discrepancy is the additional time it takes for the user's hand to move over the game-board instead of the touchscreen, calculations show a major factor is also the presence of feedback. Selection and deselection are very close together, though in the case of no audio feedback, we can see that deselection was slower. An alternate way of deselection was suggested to the users during the questionnaire, which would improve the robustness and give users a better idea of where an object will be deselected. If we look at Figure 3.12, we can see the current implementation at 1 and 2, where the green and red circles represent the green and red markers and where the black square represents an object. 1 Shows the configuration of markers and objects

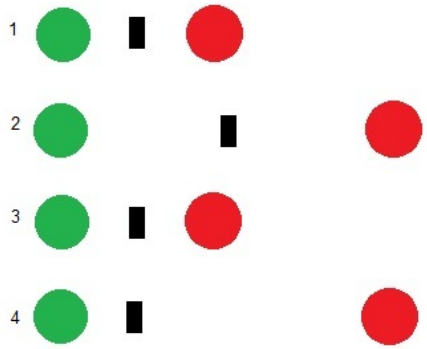


Figure 3.12: Alternative way of deselecting an object

when an object is selected and 2 shows the configuration at the moment of deselection. The experiment showed us, that when deselecting an object in this manner, a few users had to get used to the speed at which to move their fingers, such that the object would be placed exactly where they wanted. In the proposal, as can be seen in 3 (selection) and 4 (deselection), the object is stuck at a static distance from the thumb (green marker). Here, the index finger (red marker) moves away to meet the minimum distance requirements for deselection, at which point the objects is dropped at the thumb location instead of in the center of the thumb and the index-finger. However an

overwhelming majority of the participants said they liked the intuitiveness of the current implementation over the (minor) increase in robustness the alternative way of deselection potentially would provide.

3.8.4 Wait-time

The outcome described in paragraph 3.8 is in line with what we observed. Even though we were not able to calculate the wait-time exactly, by formulating the new approach as the difference in the interaction-speed of sound versus no sound we have come to the conclusion that we have actually been describing a plausible explanation for this wait-time problem, and that is the limited feedback the users get during FT interactions. During touchscreen interactions users get tactile feedback when reaching the screen. Coupled with their experience of touchscreens and basic hand-eye coordination users intuitively know when they have correctly targeted and thus selected an object. At this point users do not require or wait for additional feedback that implicitly states they have selected the object and will directly start to move their hand. The same can be said for waiting before dropping. Hand-eye coordination coupled with the fact that not releasing the hand from the touchscreen means the object is not deselected allows for users to have a good grasp on where the object is at all times and whether it is correctly deselected or not. There is a third point during an interaction with wait-time, which is the time between deselecting an object and selecting an object. This matters only when selecting a new object has to happen as quickly as possible after deselection. Deselecting an object during touchscreen interactions is simple, releasing the screen means the object will be deselected. Here, the removal of feedback from the user acts as an indication that an object has been deselected as releasing the screen is the only requirement. During FT however, no such (removal of) feedback is present and users will generally look longer to make sure they have correctly deselected an object before moving on to the next interaction.

3.8.5 Final Conclusion

In this chapter we have identified a number of issues with the current implementation as per sub-goal 2 of our thesis:

> Examine the finger-tracking and touchscreen interactions and compare them to identify the strengths and weaknesses with regards to performance

When compared to touchscreen interactions, the performance of finger-tracking interactions is worse for every interaction we described. The greatest weakness that was identified, is that the time it takes for users to react while transitioning from one interaction to the other is much larger

during FT interactions. Comments by the users in the closing informal interviews about a better reaction of the system and recommendations for improved visual feedback, coupled with our findings when looking into audio feedback versus no audio feedback, suggest how to deal with the identified performance problems. As audio has already shown to improve the responsiveness of the system, it stands to reason that different forms of visual feedback or the addition of haptic feedback can provide additional cues leading to faster response times which will have a beneficial effect on the speed at which users perform interactions. As such, evaluating these options in a comparative user study that investigates the influence of different modalities and feedback types on performance will be the next focal point of this thesis.

Chapter 4

Experiment 2: Multimodal feedback

4.1 Introduction

The results of our previous experiment indicated that for FT-based interactions to come closer to the interaction-speeds reached in touchscreen applications, a logical next step would be to diminish the time a users waits between interactions. This 'waiting time' first became evident by observing the users as they used our system and later by exploring the data we received from our test. A likely reason for this could be that the lack of haptic feedback causes the user to be uncertain whether or not they completed a gesture-based interaction in the augmented world. While during touchscreen interactions, users do not pick up an actual object either, they clearly feel when they touch the screen to pick up an object. At that moment the users are certain enough to continue with the next interaction, such as moving the object to another location. Our gesture-based interaction, and many other such implementations, currently suffer from this lack of feedback and its resulting diminished sense of confidence.

We propose to diminish the waiting time and enhance this sense of certainty by improving feedback to the users when they have 'completed' an interaction such as selecting or deselecting an object. Our next experiment tries to determine what form(s) of feedback are best suited to let the users know they have completed an interaction and can transition to the next one. We are only interested in the feedback current mobile phones can provide, which are visual feedback, audio feedback, haptic feedback(vibrations), all the possible combinations of these three and 'no feedback'. It is important to note that the hand that performs the interactions, the task-performing hand, is not the same hand that holds the mobile phone, the non-task-performing hand. One of the things we are particularly interested in is how users perceive vibrations to the non-task-performing hand.

In the previous experiments two forms of visual feedback were used to help the participants. The first was centering an object between the users thumb and index-finger when selected and or moving an object. This is purely to indicate the location of the object, but it is a form of visual feedabck. We refer to this as the default visual feedback. The second form of visual feedback was a subtly colored circle around an object denoting a user was in range of that object and about to select or deselect it. This feedback was present in Experiment 0 to help the participants see what object they were about to select or deselect when multiple (moving) objects were close to each other. While both of these visual feedbacks could be helpful during an actual game, for this experiment we decided to remove both of them as they prevent us from independently testing audio and haptic feedback and as such they could influence the results.

4.2 Related work

In recent years there has been a lot of research into gesture-based interactions for augmented reality purposes. Some of them make use of tools in the real world to manipulate the virtual or augmented world, such as MagicCup[3]. Here the researchers use a transparent physical cup to manipulate virtual objects. When the cup covers a virtual object and is put on the table the object is selected and locks into the cup. It can then be moved or rotated by moving the cup across the table. Taking the cup off the table releases the virtual object and deleting virtual objects is possible by shaking. Another system was implemented by Kato et al., which augments virtual objects on physical cards. Users can move these cards in order to manipulate the virtual objects on them[8]. These so called tangible user interfaces were designed with table-tops in mind where users have two hands free whereas our system is created for mobile phones where users only have one free hand. Others opted for interactions using vision-based finger-tracking or hand-gesture-tracking. Henrysson et al. implemented a system where a finger could be tracked in 2d without markers, and in 3d with the help of an additional marker attacked to the finger. They tested input with finger-gestures, keyboard and tilting of the phone for rotation and translation but ultimately they concluded that gestures had too many limits to be really worthwhile[4]. At a later point in time, applications were developed to track the users hands without markers. M. Lee et al created a 3D computer vision-based hand tracking system, based on four steps: Segmenting skin color, finding feature points for palm center and fingertips, finding hand direction, and simple collision detection. They tested this in three sample applications with good results but concluded that the speed did not match those of systems with data gloves or other gesture input devices[2].

None of the above papers really provide any chain of thought with regards to feedback from their system to the users, in order to improve responsiveness or immersiveness. Some researchers of course implemented visual or audio feedback into their systems but to be able to provide users with additional haptic feedback, new tools had to be developed. The most popular is the glove. Richard et al. created a glove that would create different kinds of pressure when users grabbed a hard or soft object[9]. While some gloves try to mimic the density or feel of an object, others use vibrations to let the user know they are interacting. Seo et al. developed a system using marker-less hand tracking which visualizes an entity or scene(pet, flower opening) when users have their hand opened. They also designed a prototype system using a glove for feedback. If certain motions in the hand are detected the virtual scene changes and gives tactile feedback to the hand of the user[5]. FingARtips is another system that uses a glove albeit with markers. The glove is capable of giving feedback to the fingertips when the users press or hold an object. The researchers concluded that haptic feedback worked very well and that it seemed to increase the confidence with which users interacted with the buildings[7]. A more recent project gave birth to REVEL. Here the researchers employ the principle of reverse electrovibration where they inject a weak electrical signal anywhere on the user body creating an oscillating electrical field around the users fingers. When sliding their fingers on a surface of the object, the users perceive highly distinctive tactile textures augmenting the physical object. By tracking the objects and location of the touch, they associate dynamic tactile sensations to the interaction context. The researchers list a number of AR applications for which the REVEL system could provide useful tactile feedback[10].

All of these however use additional tools, such as specially developed gloves, to give the user feedback. However, due to their cost, these are not available to everyone. Because we have stated in the beginning of this thesis that we believed FT interactions were best suited for leisure applications, we want to keep the target audience as large as possible and therefor we focus only on feedback that the mobile phone is able to provide. Because of this however, we are not able to give feedback to the task-performing hand and therefor we have to rely on feedback to the hand that holds the phone and does not perform the tasks, so called remote tactile feedback. Richter et al. already showed that, for touchscreen interactions, remote tactile feedback proved useful in decreasing task-time when compared to no tactile feedback, with 30% of participants even preferring it over direct tactile feedback[12]. Whether remote tactile feedback, in our case the vibration of the phone in the non-task-performing hand, positively influences our FT-based AR system is one of the main things we want to evaluate.

More recently, researchers have looked into the effects of modality on virtual button motion and performance[13]. They compared the possible combinations of visual, audio and haptic feedback with each other and

found a decrease in time on task completion for the VH, AH, and VAH conditions, the same conditions under which they observed a decrease in the depth of the button presses. This paper focuses on pressing(interacting with) virtual buttons on a touchscreen and the effect of modality on the amount of force and time used to press and as such it is not surprising to see haptic feedback play a major role in the outcome, albeit in combination with another feedback mode. We focus on FT-based interactions performed in air, but our goals are largely the same albeit focused on completion time and duration of the interactions.

4.3 Goal

The goal of this experiment corresponds with sub-goal 3 of our thesis, which is:

- Determine the effects of multimodal feedback on finger-tracking-based interactions.

Specifically we want to know: What is the best (combination of) type(s) of feedback during virtual object manipulation with regards to speed or responsiveness and user preference? We divide this into four sub-questions:

1. How does altering feedback modalities affect task completion time or interaction time?
2. What feedback modality is preferred by users.
3. When users complete a task, do they focus on any one particular type of feedback such as just visual cues or just haptic cues?
4. How do users regard the (remote) haptic feedback?

To answer this, we have created an experiment in which participants have to perform interactions with virtual objects during a number of tasks, each task featuring a combination of feedback types. The speed at which the participants perform these interactions will be measured and the participants' preferences will be determined at the end of the tasks by means of a questionnaire. During the experiment an objective observer will be present to write down the users' thoughts and observations and to lend assistance in the case of questions or problems.

4.4 Setup

4.4.1 Interactions

The experiment features two kinds of interactions: selecting and deselecting. These are the same interactions that were used earlier in this work. We have chosen not to incorporate moving into this experiment for a number of reasons, the most important one being that moving itself always has to make use of visual feedback and therefor testing audio and haptic feedback can not be done independently of visual feedback. Another reason is that the applications of feedback for moving are very limited. To understand this we have to look at the moments in the interaction where feedback can be applied, which is:

1. At the start of a movement
2. During the whole movement
3. When reaching a goal-position
4. At the end of a movement

Moment 1 is the same moment an object is selected and 4 is when an object is deselected. These moments of feedback are already tested during the selection and deselection tasks(see Chapter 4.5). If we imagine 1 and 4 as the temporary form of feedback for moving, feedback at the start and the end, then 2 is the constant feedback. However, constant feedback while moving is the same as constant feedback during selection, until an object is deselected, which is also already being tested during the selection and deselection tasks. This leaves 3 as the only moment of feedback left to test. While having feedback at this moment could reduce the time between users reaching a goal and starting the deselection gesture, this would only be applicable in a limited amount of scenarios, when a goal position is known to both the system and the user. This is further complicated by the fact that when reaching the edge of the goal position, the feedback has to be provided to the user as soon as possible. As we can see in Figure 4.1, feedback, whether it is a visual, audio or haptics, should be given at bor c. However, if the user immediately starts the deselect-gesture there is a reasonably high chance he will deselect the object outside of the goal. These difficulties, coupled with the fact that this kind of tasks can not be performed without the default visual feedback lead us to the decision to only test for selection and deselection.

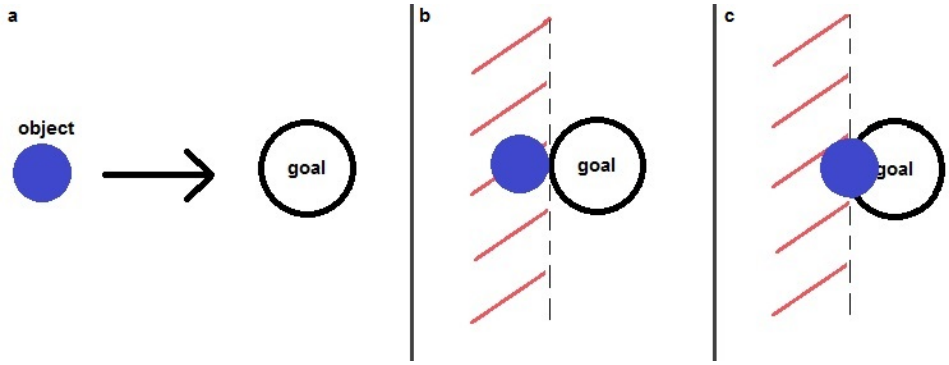


Figure 4.1:
a: Moving object to goal location
b: Edge of object reaching goal location
c: Center of object reaching goal location

4.4.2 Feedback types and how to convey them to the user

The experiment distinguishes between three main types of feedback: Haptic, Auditory and Visual. Each of these feedback types can be constant or temporary.

Haptic
At the time of writing, current mobile phones offer limited possibilities for providing tactile feedback. The only option at our disposal is vibration, with duration as a variable. Therefor, the only condition we can set for the vibration is that it has to be long enough to be felt. In the end we have chosen a vibration of 500 milliseconds for temporary feedback. For constant feedback it will constantly vibrate, between the time of selection and the time of deselection. Remember that this feedback is only present in the hand holding the phone, not the hand doing the interactions.

Auditory
For auditory feedback there is a large number of ways to relay to the user that they have performed an action. The kind of sound and the duration are variable. We defined a number of simple conditions. One, the sound has to be within the audible frequency, two it has to be loud enough to be heard and three, it has to be long enough to be heard. As there is no intuitive sound for picking up an object or selecting an object, we have chosen a simple beeping sound. This sound is not at an edge of the audible frequency so that even people unable to hear the highest frequencies, which are the frequencies where decline starts in hearing loss or the elderly, should still be able to hear the beep. In the case of temporary feedback we play it for 500 milliseconds which is three beeps, both when an object is selected

and deselected. In the case of constant feedback we loop it continuously.

Visual
Visual feedback has, like auditory, is an infinite number of options. To reach the best option we first divided visual feedback in groups. After this we identify the limitations, pro's and con's upon which we base our final decision. The groups are:

1. A visual change of the objects that is being interacted with
2. A visual change around the objects that is being interacted with
3. A visual change completely separate from the object and the close space around it
4. A combination of the above

To give an example of each, 1 could be an object changing color, 2 could be a box around the object or everything but the object changing to a shade of gray, denoting that the object that did not change is being interacted with, and an example of 3 could be a text in the corner of the screen displaying: Object x selected.
Intuition suggests that group 3 would not give sufficient information quick enough as it would take the participants' focus away from the object they are interacting with, which would result in delayed feedback. Because group 1 and 2 keep the user's focus on the object that is being interacted with, we have no reason to believe that adding 3 to 1 or 2 has an additional positive effect, whereas implementing 3 the wrong way could only negative influence it. Group 1 and 2 can give information about both which object is being interacted with and that an interaction is completed, without the need for the participant to look away from the object. As both these groups have the potential to relay the same kind of information, the only condition that has to be met is that the visual change itself is sufficiently clear. Because our framework did not lend itself to changing the texture, and therefor color, of objects at run-time, we decided to display a (bounding)box around the object that is being selected.

Out of all the feedback types, we expect any combination involving audio to be the fastest. The reason for this is that a human's reaction time to audio is faster than their reaction time to visuals or vibrations[14]. The addition of visuals to audio, can in some cases, for instance when multiple objects are present, clarify which object is being interacted with and as such be even better than purely audio feedback. It is unclear if a vibration in the non-task-performing(non-dominant) hand will have a positive effect on the users' realization that they just completed an interaction. Even though the research mentioned in the previous work section suggests remote tactile

feedback could, among other things, improve the users' task completion time, none of them perceived the virtual or augmented world through the device delivering the haptic feedback. In our case, the vibration of the phone could actually make it harder for the users to look at the augmented world and slow them down.

4.4.3 Marker tracking

In the earlier experiments some participants noted that under certain specific conditions the red marker would disappear. This could for instance happen when a bright light directly hit the marker or the see-through tape on the marker, making the colors blur or appear white to the camera due to over-saturation, which in turn caused the camera not to pick them up. Decreasing the threshold for the red color often meant the camera would identify the hand as a red marker and was therefor not an option. In this experiment, the marker-tracking is done using the same method as in the earlier experiments but we added a little tweak to make the tracking even more robust. During this experiment the participants all received a blue medical glove. This is a cheap way of enhancing the contrast between the red marker and the participants hand, which enables us to greatly decrease the threshold of the red marker, making the tracking of said marker even more robust than before, while not noticeably influencing the green marker.

4.5 Procedure

For clarity, we start by defining a number of actions and symbols:

S = Select an object

RS = Realize you have selected an object

D = Deselect an object

RD = Realize you have deselected an object

M = Move the hand to the next object

V = Visual feedback

A = Audio feedback

H = Haptic feedback

C = Constant feedback: the feedback will persist until dropped

T = Temporary: the feedback will be experienced temporary

( ) = Moment in time after an interaction or after a user realization

–>= The time it takes for an interaction or user realization

4.5.1 Task 0: Training

The first task will let participants get familiar with the interactions and feedback. In this task there are three objects placed next to each other on a virtual board. Each of these objects has their own unique feedback, from left to right: audio, visual and haptic. Participants are asked to select and deselect these objects multiple times, to get an understanding of how the interactions work and what each particular feedback encompasses. During this task an observer checks if the participants have interacted with all three objects and if they show an understanding of when a particular feedback is presented to them and how the interactions work. When in doubt the observer can ask the participant to elaborate on what is happening.

In the previous experiment we employed a training time of thirty minutes. The time was divided between understanding how the interactions work, understanding when to start moving or select and deselect and getting used to moving the camera and the objects. This time however we have static objects that do not have to be moved, a camera that does not have to be moved quickly from left to right or from top to bottom to follow objects and feedback to tell the users when to start the next interaction. By setting up the experiment in this way we are able to greatly reduce the training time by only training the interactions and exposing the users to the different kinds of feedback.

4.5.2 Task 1: Multiple object experiment - Selection

The purpose of this task is to evaluate how fast users can select multiple objects after each other., without having to deselect. The goal of this task as conveyed to the users is to select eleven objects as fast as possible.

Eleven yellow balls are shown. Because we can't measure the time it takes to select the first object the timer will start when the first object is selected. When the last object is selected the timer will stop. From this duration the average and standard deviation will be calculated. The time between each object being selected will also be measured separately. This task consists of eight rounds of 11 objects, where each round has its own feedback. The combinations of feedback are:

- V: Upon selection the object will temporarily get a white bounding-box
- A: Upon selection a tone will be heard for 500 milliseconds
- H: Upon selection the phone will vibrate for 500 milliseconds
- V + A
- V + H
- A + H
- V + A + H
- No feedback: When the object is selected nothing happens

Each of these feedback types is temporary. As this task only deals with selection, the system will automatically deselect an object and place it into a state where it can no longer be selected, after it has been selected by the user, so that the user can proceed and select the next object. Because objects get deselected automatically, having constant feedback for audio and haptic would mean it would be either very short, which would make it the same as temporary feedback, or the feedback from a previous object would still be heard or felt when the user is in the process of interacting with other objects, which would serve no purpose and would only limit the feedback for the rest of the experiment. Similarly for visual feedback, constant feedback would either be very short as the system deselects immediately after the user selects, or it would be visible after an object was selected, meaning during the whole task users would see which objects they had selected, making it obviously the best option, thus (partially) defeating the purpose of this task.

The reason we have eleven barrels is to make sure we do not make the duration of the experiment too long, but still have a big enough number to get a good representation. We also want to make sure participants have a good indication which ball to select next so they do not waste

time searching or move their hand accidentally away from the object they are looking for and as such we want to be able to have all the objects on screen at the same time. When putting the balls in a straight line it is difficult to keep all the balls in the field of vision, and putting the balls in a square or circle will make it more complicating for users to understand when they are finished. Therefor we have chosen to put the balls in a 'U'-pattern, which has a clear beginning and end, whether someone begins top-left or top-right. For each selected ball it is also clear which ball is closest and thus would be the fastest to select next.

Procedure
The process is as follows: move to and select the first object(S) → (1) → realize the object is selected(RS) → (2) → move to next object(M) → select next object(S) → (1) → repeat until all objects have been selected once.

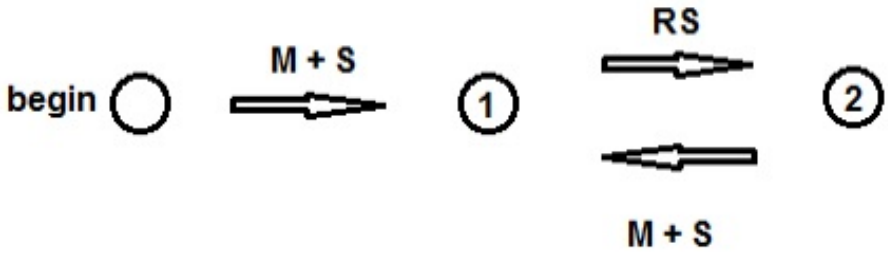


Figure 4.2: Procedure task 1

Data
As soon as the object is selected by the user the system deselects the object so another object can be selected again. The user has no knowledge of this and as such this does not affect the (inter)actions of the users. The only moments in this process where the system can actually measure time is when an object is selected, so everything from (1) through (2) back to (1) is measured: RS + M + S.

4.5.3 Task 2: Single object experiment - Selection and Deselection

The purpose of this task is to measure how fast users can select and deselect objects. The goal as explained to the users is to select and deselect one object eleven times.
One yellow barrel is shown. When the object is deselected for the first time, a timer will start until the last time the object is selected. This duration will be divided by ten to get an average. It is important to note that this average also includes the time it take for a users to realize they have selected an object. The time between select and deselection of an object, as

well as the time between the deselection of the object and the next selection will be measured separately and are the most important measurements. The feedback for selecting an object will be the same as for deselecting, for example when testing audio feedback for deselecting, audio feedback will also be used for selecting. Participants have to do this for each of the combinations of feedback, both constant and temporary, as described below:

VC	VC + AC	VC + AC + HC
VT	VC + AT	VC + AC + HT
AC	VT + AC	VC + AT + HC
AT	VT + AT	VC + AT + HT
HC	VC + HC	VT + AC + HC
HT	VC + HT	VT + AC + HT
	VT + HC	VT + AT + HC
	VT + HT	VT + AT + HT
	AC + HC	No feedback
	AC + HT	
	AT + HC	
	AT + HT	

Figure 4.3: All feedback combinations

Procedure

The process is as follows: Move to and select object → (1) → realize the object is selected → () → deselect object → (2) → realize the object has been deselected → () → select object again (1) → repeat until the task stops.

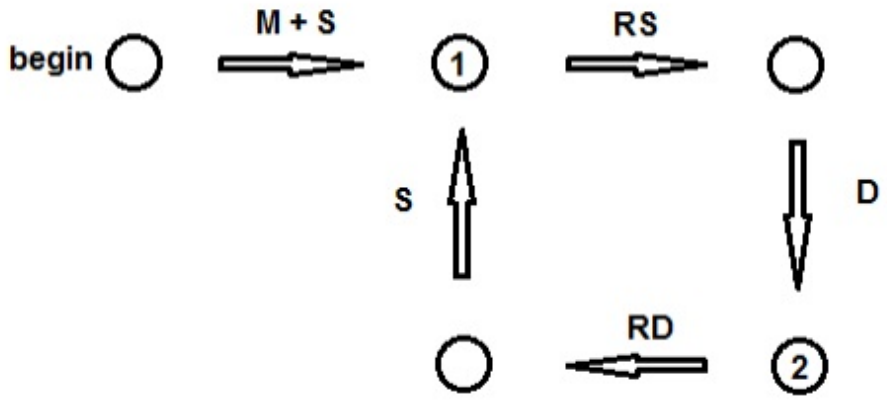


Figure 4.4: Procedure task 2

Data

The moments in this process where the system can measure time are when an object is selected and when an object has been deselected. Between (1)

and (2) we measure the realization of selecting something(RS) and the time it takes to deselect (D). Between (2) and (1) we measure the realization of having something deselected (RD) and the time it takes to select an object (S). What we are measuring is the time it takes for a full cycle of selecting and deselecting an object: RS + D + RD + S. Because the object that has to be selected and deselected is stationary, there is no need for the users to move the hand away from the object, only the finger-movements for selection and deselection are necessary.

4.5.4 Task 3: Single object experiment - Selection and Deselection with movement

Select and deselect eleven objects as fast as possible. A yellow barrel is shown. When the object is deselected, it will disappear and another yellow barrel will appear at a fixed distance (1 tile in our AR world) from the first one. When the object is deselected for the first time, a timer will start until the object is selected for the eleventh time. This duration will be divided by ten to get an average. It is important to note that this average also includes the time it take for a user to realize he has selected an object. The time between the deselection of the object and the next selection will be measured separately and is the most important measurements, as it is during this interaction that the user moves his hand from one tile to the next. This task differs from Task 2 in that the users move the position of the hand between deselecting one object and selecting the next. In doing so we can determine the time it takes to move a set distance. Participants have to do this for only one feedback, because the time it takes to move in between interactions is independent of the type of feedback:

1. Visual feedback: When the object is deselected, the object will move a tile.

Procedure
The process is as follows: move to and select object → (1) → realize the object is selected → () → deselect object → (2) → realize the object has been deselected → () → move → select next object (1) → repeat until the task stops.

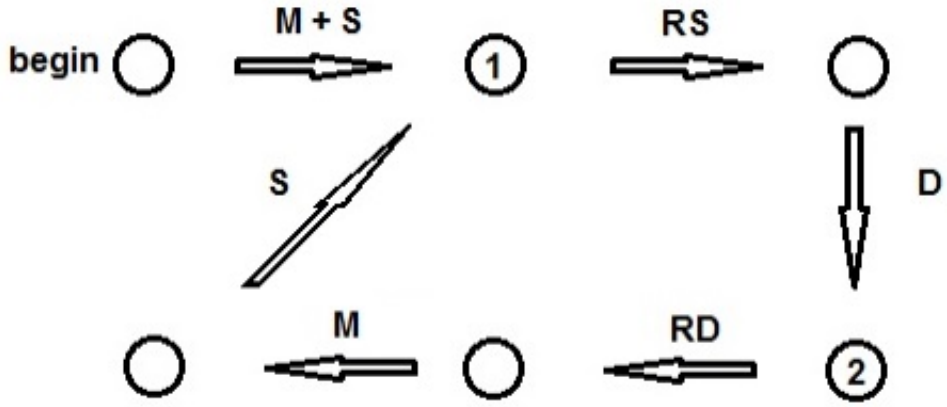


Figure 4.5: Procedure task 3

Data
The moments in this process that the system can measures time are when an object is selected and when an object has been deselected. Between (1) and (2) we measure the realization of selecting something (RS) and the time it takes to deselect (D). Between (2) and (1) we measure the realization of having something deselected (RD), the time it takes to move to the next object (M) and the time it takes to select an object (S). What we have is the time it takes for a full cycle of selecting and deselecting an object plus the time it takes to move a set distance: RS + D + RD + M + S.

Combining all the data To summarize the data we get from the separate tasks:

- Task1: RS + M + S
- Task2: RD + S, RS + D and S + RS + D + RD
- Task3: RS + D, RD + M + S and S + RS + D + RD + M

The difference between task3 and task2 is the time it takes to move to the next object. If we take the visual feedback data of one user from task2 and average it, we get the average time it takes to do one complete cycle of S + RS + D + RD. By doing the same with the data from task 3 for this user we get the average time it takes to do one complete cycle plus moving RS + D + RD + M + S. Subtracting these two will give us an estimate for the average time it takes to move one tile, this is independent of which type of feedback was used.
By keeping the distance in task1 and task3 the same we can use this average time it takes to move in task1, subtract it there from the smaller cycle RS + M + S in task1 and get the time for S + RS; selecting an object and realizing you have selected it(note that this is only for the temporary combinations

of feedback used in task1). Then we can use this again in task2 on the complete cycle RS + D + RD + S, subtract S + RS and get the time D + RD; deselect an object and realize you have deselected it.
While for comparison in our case it does not matter whether one compares, for instance, VC RD + S with HC RD + S, as would be the case in task2, or VC D + RD with HC D + RD as would be the case after doing all the aforementioned calculations, it would matter if one decided to use other feedback for selection as for deselection. In the case of S+RS+D+RD, the realization time of deselection (RD) would depend on the feedback for deselection, where the realization time for selection (RS) would depend on the feedback for selection.
Due to time constraints and after receiving user feedback with regards to different feedback for different interactions we have decided not to implement any further experiments with different feedback types for selection and deselection.

4.5.5 Participants

A total of 26 subjects participated in the experiment. These subjects are all computer science students at the University of Utrecht and range from 21 to 30 years old(average age 24, standard deviation of 2.69 years). While this group does not represent the population of a country, it does represent the expected user population of the kind of system that is being used. Out of the 26, 25 are male and 1 is female, 4 users are left-handed, 2 are ambidextrous and the remaining 20 are right-handed.

4.6 Results

4.6.1 Handling outliers

Because the data set gathered from the experiment was too large (26 participants times 11 objects times 29 tasks) to remove outliers manually, as we did in the previous experiment, we wrote a program to compute and remove them automatically. To remove outliers we used the Median Absolute Deviation method, which is a non-parametric or distribution-free approach to detect outliers based on computing medians[11], as described below:

1. Compute the median of the original input data.
2. Compute absolute value of deviations of original input data from this median.
3. Compute median of these absolute deviations.

4. Compute ratio of absolute deviation from step 2 and median from step 3.

5. If this ratio is greater than critical value consider the value as an outlier.

We used a conservative critical value at 2.5. Outlier removal methods always have a chance of classifying valid data as false positives, however by using a conservative critical value this will be kept to a minimum.

4.6.2 Multiple objects experiment

Figure 4.6 shows the time it took for users to realize they had selected an object and then select another one. As we can see, triple feedback is the fastest, followed by all the double feedback types. Last are the single feedback types and no feedback, with haptic being close to the first four, but no feedback, audio and visual much slower. A one-way repeated measures ANOVA with a Greenhouse-Geisser correction proved these results were significant ($F(2.628, 63.077) = 10{,}688$, $p < 0.05$) and Bonferroni post-hoc tests show that ATVTHT, VTHT, ATHT and ATVT are all significantly faster than 'no feedback' and VT ($p < 0.05$). HT is significantly faster than VT ($p < 0.05$).

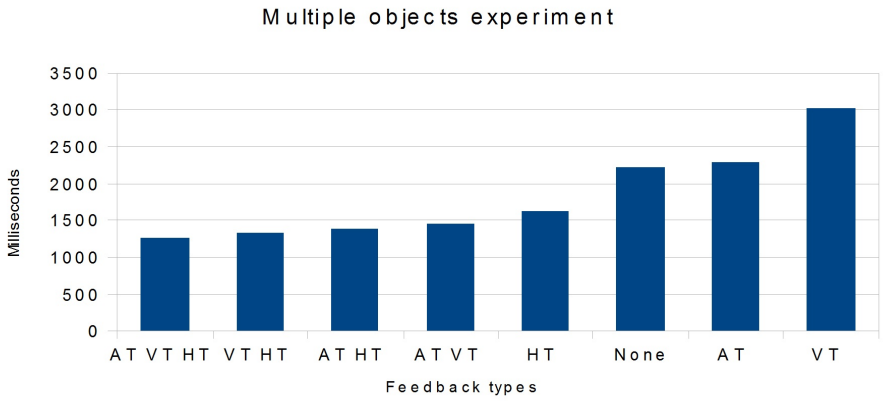


Figure 4.6: The interaction times for all temporary feedback combinations, averaged over all users

4.6.3 Single object experiment

Figure 4.7 shows the time it took users to realize they had selected an object and then deselect it, averaged over all users, for the single object task. As can be seen, the interaction times are very similar for most of the feedback types. The variation in time becomes greater when we look at the slowest feedback types(the right half of the graph), which consists of: the single feedback types, all the combinations involving only constant feedback, AC VT, AC HT and no feedback, which is at the bottom as expected. The three fastest feedback types are all combinations of triple feedback. At place six is VC HT, which, as we will shortly show, is the fastest when doing a full interaction cycle. To determine statistical significance we used one-way repeated ANOVA. A significant difference between the feedback types for select realization + deselection could not be proven.

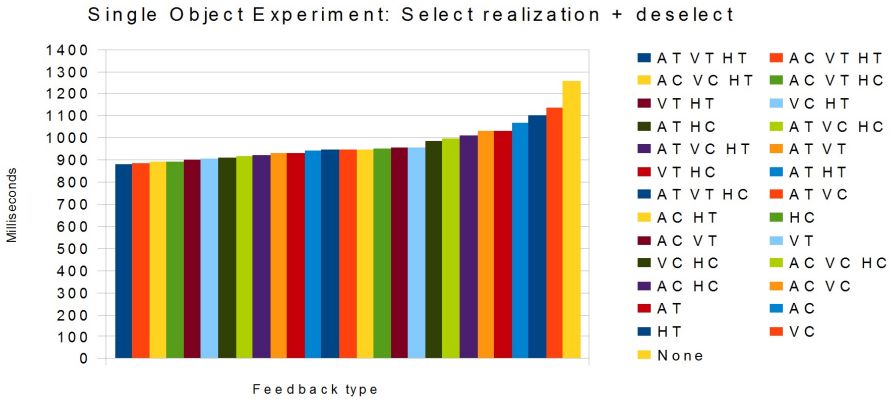


Figure 4.7: The interaction times for all feedback combinations, averaged over all users

Figure 4.8 shows the time it took for users to realize they had deselected an object and then select it again, averaged over all users, for the single object task. At first sight the order is very non-descriptive and as such it is difficult to conclude anything directly from it. However, what we can see is that all the combinations of the double feedback Visual and Haptic are in the fastest half; VC HT is 1st, VT HC is 2nd, VT HT is 7th and VC HC is 10th.
Another thing that stands out is that 'no feedback' is the 3nd fastest. While this results is counter-intuitive, it can be explained by the way users treated this task; a common mindset was If I don't have feedback, I might as well not wait for it. The entire reason for feedback was to give users more certainty when they had completed an interaction, but by applying this mindset, users gave themselves this certainty, whether it was deserved or not. When approaching it in this manner, 'no feedback' being in 3rd place confirms the intuitiveness of the interactions, as being so fast means users did the interactions right without the need for feedback.
As can be seen from the graph the time is even closer together than in figure 4.7, with the fastest and slowest having approximately 150ms difference. As selecting an object happens faster than deselecting, there is less room for deviation here.
To determine statistical significance we used one-way repeated measures ANOVA. A significant difference between the feedback types for deselect realization + selection could not be proven.

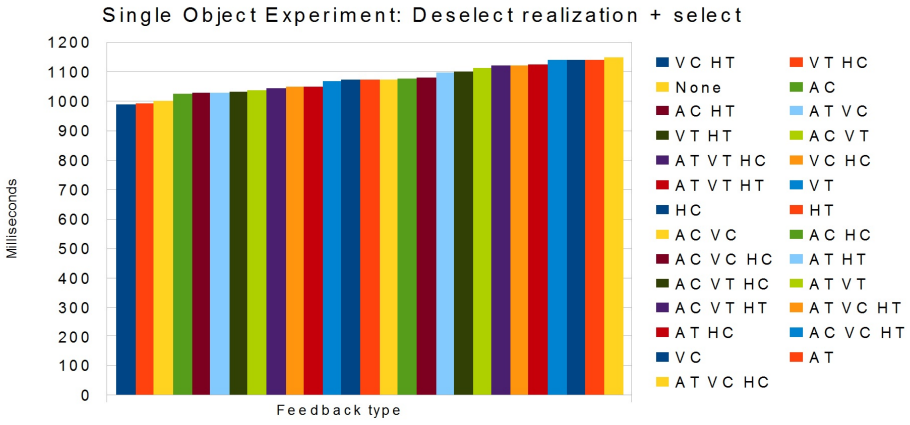


Figure 4.8: The interaction times for all feedback combinations, averaged over all users

Figure 4.9 shows the time it took for users to complete a full cycle of selecting, realizing the object is selected, deselecting it and then realize it is deselected, averaged over all users, for the single object task. The combination of Visual and Haptic comes out as the fastest, with VC HT being 1, VT HC being 2, VT HT being 4 and VC HC being 14. The slowest types are 'no feedback', 3 out of 6 of the single feedback types and some combinations with only constant feedback. To determine statistical significance we used one-way repeated measures ANOVA but no significant difference could not be proven. However, when directly using a Wilcoxon signed ranked test to test for a significance between VC HT and 'no feedback' we see that they are significantly different ($p < 0.05$). Further investigation shows that compared to 'no feedback', VCHT delivers a decrease in interaction times of 11%.

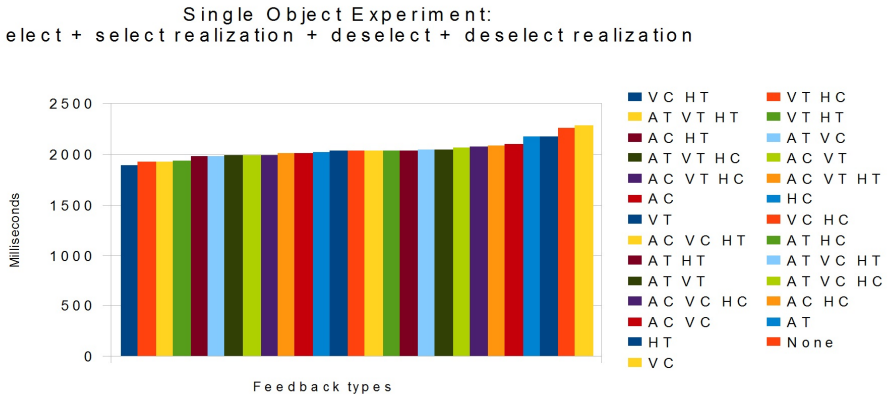


Figure 4.9: The interaction times for all feedback combinations for a full cycle, averaged over all users

4.6.4 Questionnaire and subjective user feedback

Rating multiple objects task experiment

In Figure 4.10 we can see the ratings given by the users for the multiple object selection experiment. As one would expect 'no feedback' has received the lowest rating. At 2nd to last place comes temporary audio and as can be seen, all single feedback types are lower than double and triple. The best rating was given to Visual+Haptic and Visual+Audio+Haptic.

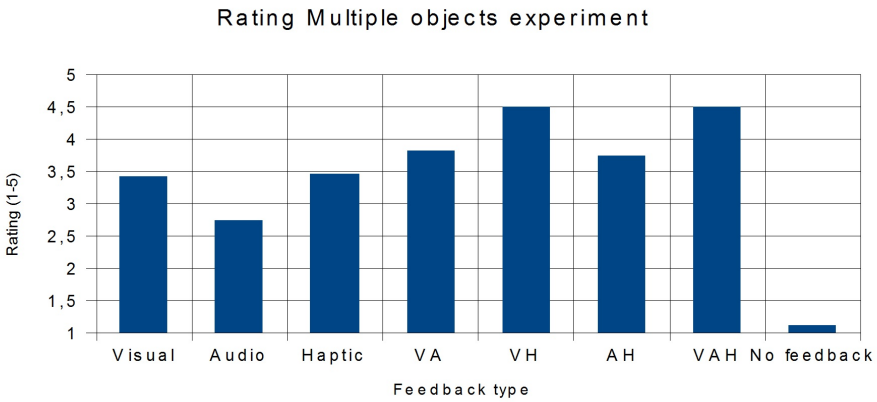


Figure 4.10: Ratings for the multiple objects experiment, averaged over all users

Interestingly, the rating for V and H is approximately the same, as well as the rating for VA and AH(albeit higher than V and H) as well as the rating for VH and VAH(albeit higher again). What we can conclude from this is that audio seems to play only a really small part during multiple object selection. Being given the lowest rating of the actual feedback types, it does seem to add something users like if it is combined with another feedback type, see VA vs V and AH vs H. Ultimately, as we see that the two highest ratings VH and VAH are equal, temporary audio does not seem to add to the user perception, nor does it seem to take away from it, making the combinations that contain at least Visual and Haptic preferable. A repeated measures ANOVA with a Greenhouse-Geisser correction proved these results were significant ($F(4.033, 100.832) = 53.269$, $p < 0.05$) and Bonferroni post-hoc tests show that both VH and VAH are significantly better than the other feedback types ($p < 0.05$). It also shows there is no significant difference between V, H, VA and AH, but they are all significantly better than audio ($p < 0.05$) and every feedback is significantly better than 'no feedback' ($p < 0.05$).

If we now look back at the interaction times of the multiple objects experiment in Figure 4.6, we see that the two combinations that are fastest are also the combination with the highest ratings, namely VAH and VH. Number 3 and 4 VA an AH are also the 3rd and 4th fastest. Looking at the bottom four, even though no feedback got the worst rating it is not the slowest as it is almost tied with audio, which had the 2nd lowest rating. Slowest however is be visual. The reason that this is even slower than 'no feedback' could be that people sometimes missed the visual feedback, therefor trying to select an object that was already selected leaving them wondering whether they had selected it or not, whereas during 'no feedback' they would just continue on hoping they had selected it.
When we asked the users for the reason of their preferences, the general consensus was that when dealing with multiple objects, one at least wants the temporary visual feedback to see which object has been selected. Additional feedback was preferred for a number of reasons, one of them being that when not looking at the objects one can still feel or hear an object was selected. Another reason was that haptic feedback made the interactions feel more real. Overall, the combination of haptic and visual was prefered over audio and haptic as some users disliked the audio, saying it did not 'fit' the interaction or not add to the 'realness' of the interaction. Finally there were some users that said that more is better, which explains the high rating for visual + audio +haptic.

Rating single objects experiment

The ratings users gave for the single object selection and deselection are less straightforward. As we can see in Figure 4.11, 'no feedback' has again the lowest rating and again audio is 2nd lowest. In this experiment, Haptic has the best scores of the single feedback, even scoring better than some of the double feedback. Visual+Haptic has the best scores of the double feedbacks and in this case also better than Visual+Audio+Haptic.

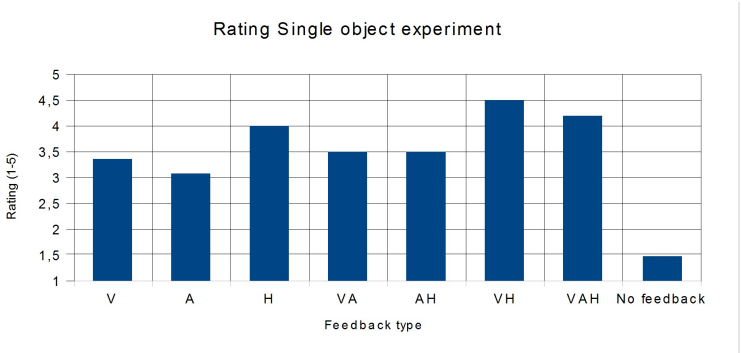


Figure 4.11: Ratings for the single object experiment, averaged over all users

There are multiple reasons why these results differ so much from the results of the multiple objects experiment. For one, the task is different and therefor one could expect different feedback to work better or worse. Another reason is that this rating actually encompasses both the temporary and continuous versions of the feedback and this leads straight into the final reason; constant audio is considered really bad, which brings the ratings of A,VA, AH and VAH down. Once again though, the combination of at least Visual + Haptic is preferred by users the most.
A repeated measures ANOVA with a Greenhouse-Geisser correction proved these results were significant ($F(3.876,\ 96.896) = 29{,}993$, $p < 0.05$) and Bonferroni post-hoc tests show that VH is significantly better than all other feedback types ($p < 0.05$) except for VAH and VAH is significantly better than the remaining ones ($p < 0.05$) except for H. All feedback types are significantly better than 'no feedback'.

Figure 4.12 shows what users thought was the best feedback. Following the table is a breakdown of all the feedback types and how many times they were named. Note that when a user states he prefers VCHT, V and H will both be counted, therefor the total number of V + H + A can surpass the amount of users.

VCHT	ATHT
VCHT/VCAT	VCHC
ACHC	(ATVT)HC
ATVCHT	VCHT
ATVCHT	ATVCHT
VCAT	VC
VCHT/VCAT	VCHT
VCHT/VCHC	ATHT
ATVCHC	ATVCHT
ACHC	VCHT
VCATHT?	VCHT
VCHT	VCATHT
VCHT	VCHC

Figure 4.12: Prefered feedback combinations

What we can take away from Figure 4.12:

- The largest preference went out to VCHT with 8, followed by VCATHT with 6
- 24 Users prefer a feedback with H, 22 with V and 15 with A
- 21 Users prefer a feedback with VC, only 1 with VT
- 13 Users prefer a feedback with AT, only 2 with AC
- 18 Users prefer a feedback with HT, 7 with HC
- 1 User prefers single feedback, namely VC
- 17 Users prefer double feedback, 8 of them VCHT
- 8 Users prefer triple feedback, 6 of them VCATHT

The general conclusion we can derive from this is that Constant Visual, Temporary Haptic and particularly the combination of both is preferred by most users.

When asked for their reasons, it became apparent that users like the constant visual feedback as an indication that they currently have an object selected and the temporary haptic feedback as an indication that they have just selected or deselected something. This was also the case for most of the users that preferred any combination of visual + audio or visual + audio + haptic, where the temporary audio replaces the temporary haptic or just adds to it as a means of indicating something is being selected or deselected. This was also explicitly stated by a number of users.
Another conclusions we can derive from this table and the users' comments is that constant audio is unwanted, to some even annoying. When asked if users thought another sound would be better none of them could come up with a serious alternative, with only one saying: Anything as long as it doesn't rattle. Most users thought that audio did add something, but could not quite figure out what or why. From the moment this phenomenon started to show itself we started debating this with some users and through this and my our knowledge and experience we came to the conclusion that from very early on, virtual environments or interactions in virtual environments have always come paired with certain audible feedback. From the sound of opening a door, or 'the reloading sound' picking up ammo makes in games (think about the sound that opening a door or picking up an object makes in real life) to typing on a keyboard or even clicking with your mouse, everything you do in virtual environments has some kind of sound associated with it. This leads me to believe users are conditioned to expect to hear some kind of audible feedback associated with the interactions that they make, even if these interactions would not make a typical sound in real life.

Whether or not this is actually the case is another research topic. One user also noted that audio might be more useful when it is used to indicate something has gone wrong instead of right.
Looking back at the interaction times in Figure 4.9 we can see that the feedback preferred the most, VCHT is also the fastest. The number two though, VCATHT, which is the same feedback with the addition of temporary audio, is number 17 when looking at the interaction time. This further indicates that the users like the audio component for reasons other that speed.
When users were asked if they focused on any one particular type of feedback such as just visual cues or just haptic cues, 11 out of 26 answered they did not. Visual and Haptc both had 7 users focusing on them more and only 1 person answered audio.

Haptic feedback

As haptic feedback was given to the non-task-performing hand we were particularly interested in how users perceived it. As such we asked users a number of questions pertaining to the speed at which they realized haptic feedback was linked to virtual interactions, the naturalness of the interactions and the feeling of immersion(see AppendixC). Figure 4.13 shows the rating users gave for each of these subjects. As can be seen by the average rating of 4.4, users were very fast to realize the feedback in the non-task-performing hand with only 1 user giving a rating below 4, stating that he would have liked feedback in the task-performing hand. For naturalness, users gave an average of 3.8, with only 1 user giving it a rating below 3 and explicitly stating he experienced it as unnatural. Other users stated that it does not feel wrong and one user even said haptic feels closest to real life. Immersion had an average score of 3.7, with only 2 users rating it below 3, meaning that users felt that even though the feedback was in the non-task-performing hand, it still added to the immersion.

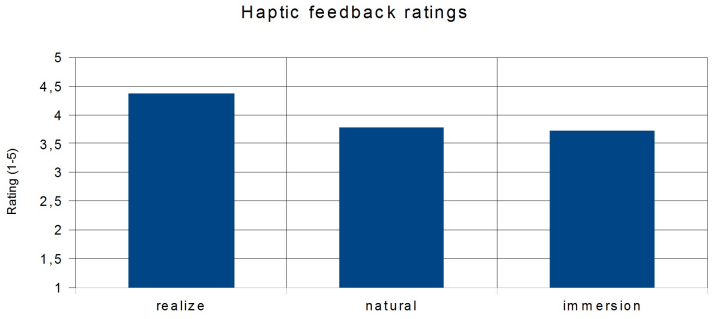


Figure 4.13: Haptic feedback ratings, averaged over all users

Chapter 5

Conclusions and future work

5.1 Conclusions

In this thesis we set out to explore and improve previously designed gesture-based interactions for mobile AR applications.
In 'Experiment 0', which directly preceded this thesis, we addressed some of the weaknesses found in earlier works and set up a comparative study in a game-like environment, with on one hand interactions on a mobile device's touchscreen and on the other hand FT interactions on a board, to test performance as well as usability and enjoyment. We found that even though the performance and usability scores for gesture-based interactions were significantly lower, user satisfaction was significantly higher. We did, however, not expect the task completion time of FT interaction to be so much slower when compared to the touchscreen.

Experiment 1 aimed at finding the bottlenecks in our FT interactions. To make sure we did not introduce any training effects users trained with our interactions for half an hour, thus giving us the possibility to determine if users got significantly better within this period and if fatigue started to set in after this many interactions. We successfully determined a number of bottlenecks within our interactions and from the data and observations it became clear that the biggest bottleneck was not present in the interactions themselves but in the time users waited between FT-based interactions, a wait-time that was not present in the touchscreen part of the experiment. The experiment also showed that after 45 minutes of constant interactions users started to get tired, although mostly in the arm holding the phone. No one reported experiencing any pain. Unfortunately, due to performance issues we were not able to determine whether or not users got significantly better in a short period of time with data gathered from the experiment. Users did however rate themselves high on improvement during the training period, implying that they believed they improved a lot.

Further analysis into the wait-time showed that a solution could lie in feedback presented to the user upon completing an interaction and as such this became the focus of our second experiment.

In Experiment 2 our goal was to find the most effective combination of feedback with regards to speed and responsiveness and user preference. In the case of selecting multiple objects after each other Visual Temporary + Haptic Temprorary (VTHT) and Visual Temporary + Audio Temporary + Haptic Temporary (VTATHT) were both the two fastest and the two with the highest ratings. It should be noted that because only temporary feedback was considered, Constant Visual feedback was not part of this experiment. However, if constant visual feedback would fit in the application, it will easily replace or add to temporary visual, making VCHT and VCATHT the best option. A prime example of this is internet links being blue when never selected, change color when hoovered over or selecting, and purple when selected before.
In the case of single object selection and deselection Visual Constant, Haptic Temporary(VC HT) was the fastest and had the best ratings. As Haptic feedback was given to the non-task-performing hand, we were particularly interesting in the ratings users gave for that particular feedback. When asked if they immediately registered our haptic feedback as feedback that an object was being interacted with, users responded overwhelmingly positive. Naturalness and immersiveness also received medium to high ratings. Overall, combinations containing Haptic did very well in both the ratings and speed and they were actually the most preferred.

As said at the beginning, the main goal of this thesis was to explore the usability of finger-tracking based interactions and improve them. We started, even before this thesis, by implementing FT-based gestures that were found to be good for translation and tested them in a board-game. Throughout our experiments we found that while FT interactions are no match for touchscreen interactions when it comes to speed, as far as enjoyment and immersiveness is concerned, they reign supreme. Our second experiment showed that to improve the speed of our interactions a good place to start was by improving the feedback from our system to the users when transitioning from one interaction to another. Therefor, in our last experiment we attempted to improve FT interactions by determining the effects of multimodal feedback on interaction times and user perception. The results show a significant difference in interaction times when selecting multiple objects after each other without having to deselect, and a limited significant difference in selecting and deselecting single objects, with feedback decreasing the interaction time of a full cycle of selection and deselection by 10% on average. User perception was significantly more positive when using certain combinations of feedback compare to 'no

feedback'. The feedback combination of constant visual and temporal haptic proved to be both the fastest and the most preferred feedback.
In summary, we have shown that when deal with FT interactions such as selection, translation and deselection, it is important to complement them with the right kind of feedback, to increase the responsiveness and speed of the interactions and to enhance the user satisfaction while using the AR application. While the speed of FT interactions might not come close to the speed of touchscreen interactions, as long as picking up an object in AR is close to the speed of picking up an object in real life and is more fun then both touchscreen and real life interactions, there will be a place for it in future applications.

5.2 Applications and future work

The FT interactions and feedback options we have studied are not only be applicable to our system but could benefit others as well, such as the new Google glass. Instead of just interacting with voice commands, Google glass could also utilize hand-tracking or finger-tracking to, for instance, select a building, do an image look-up and acquire information. The question to be answered then becomes: does remote tactile feedback to the head, or part of the head where Google glass is worn, give the same results as when it applies to the non-task-performing hand.
As Google would have us believe, the introduction of Google glass ushers in a new era, where mobile head-mounted displays(not those that require carrying a backpack with laptop) become commonplace. This also opens up the possibility of interacting with both hands, which could potentially solve some of the problems that plague FT interactions, such as switching between translation, rotation or scaling. To take a step back, in this thesis we have limited ourselves to translation(selection, moving, deselection) not only because it is the most used form of interacting with objects during games, but also because testing the atomic interactions of scaling and rotation in a game setting is no easy task, predominantly because setting up a game-like experiment that would allow for the testing of both speed and enjoyment proved difficult. Take the case of scaling for example: One would first have to select an object, then scale it by moving the fingers away from each other and then deselect it. We already run into a problem; scaling and deselecting use the same gesture, namely moving the fingers away from each other. One could work around this by having the object deselect itself after reaching the desired scale, but in a game involving all three interactions, translation, scaling and rotating, this would have to be handled differently, which brings us back to switching between gestures. A possible option could be that one hand only makes gestures to select or deselect, while the other hand deals with translation, scaling and rotation. How to specifically deal with this

issue by using a second hand could make an interesting research topic.
In our multimodal feedback experiment we have seen that feedback can improve interaction times by enabling the users to respond faster to completed interactions. Another point where interactions could be improved is accuracy. When analyzing this problem we realized that the only interaction this would be useful for is for dropping objects at a certain, predefined, location. A possible research question could be: Is it possible to add feedback at the moment a user has moved an object to a desired location to increase accuracy. To give some example implementations: audio would need a sound each time the user reaches the desired location with the object. This would result in a continuous sound when the users hoovers the objects above this location or each time they move it a bit off the location and back on again. While it may improve the accuracy it stands to reason it can become unpleasant to continuously hear a sound. A good example could be a subtle humming that become louder or deeper/lower when nearing the desired location. With visual feedback the object could, for instance, be highlighted when it is at the correct location or an emanating circle could be shown. Haptic could make the phone vibrate while the object is at the correct location.
A third area that can be influenced is user perception. Besides the user preferences from the last experiment, multimodal feedback could potentially alter user perception in another manner. We all know that sound and visually attractive images can enhance a persons senses, from watching loud scary movies to exciting videogames to sad animations, but in what manner tactile feedback can alter our perceptions is less known, and even more so for remote tactile feedback. We have just shown that remote tactile feedback has a positive influence on interactions, so it stands to reason it can have an influence on other areas as well. There are currently a lot of research topics in this direction such as can haptic feedback help avoiding collisions or can haptic feedback amplify scary situations.
Finally, to shed some light on a limitation of our study that has not had much discussion, we want to address the quality of the AR image. While the framerate was sufficient during the biggest part of the experiments, there were times when a better phone could have improved the smoothness of the interactions, therefor potentially improving the speed of the users. It would be interesting to see similar experiments performed with increasingly better hardware that comes in years to come, to see in what capacity framerate influences interaction times.

Bibliography

[1] R. Azuma, , Y.Baillot , R.Behringer, S. Feinter, S.Julier and B.Macintyre,*Recent Advances in Augmented Reality. IEEE Comput. Graph. Appl. 21, 34-47. (2001)*

[2] M. Lee, R. Green, and M. Billinghurst ,*3d natural hand interaction for ar applications. Image and Vision Computing New Zealand. (2008)*

[3] H. Kato, K. Tachibana, M. Tanabe, T. Nakajima, and Y. Fukuda ,*Magiccup: a tangible interface for virtual objects manipulation in table-top augmented reality. IEEE International, pp. 75-76. (2003)*

[4] A. Henrysson, J. Marshall, and M. Billinghurst ,*Experiments in 3d interaction for mobile phone ar. In Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia, GRAPHITE '07, (New York, NY, USA), pp. 187-194, ACM. (2007)*

[5] B.-K. Seo, J. Choi, J.-H. Han, H. Park, and J.-I. Park ,*One-handed interaction with augmented virtual objects on mobile devices. In Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI '08, (New York, NY, USA), pp. 8:1-8:6, AC.M (2008)*

[6] W. Hürst and C. van Wezel ,*Gesture-based interaction via finger tracking for mobile augmented reality. Multimedia Tools and Applications, pages 1-26. 10.1007/s11042-011-0983-y.*

[7] V. Buchmann, S. Violich, M. Billinghurst, A. Cockburn,*FingARtips:gesture based direct manipulation in*

augmented reality. GRAPHITE 04: 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South EastAsia, ACM Press, (New York, NY, USA), pp. 212221.(2004)

[8] H. Kato, M. Billinghurst, I. Poupyrev, K.Imamoto and K.Tachibana,*Virtual Object Manipulation on a Table-Top AR Environment. In Proceedings of International Symposium on Augmented Reality, ACM, 111-119. (2000)*

[9] P. Richard, G. Burdea, D. Gomez, and P. Coiffet,*A comparison of haptic, visual and auditive force feedback for deformable virtual objects. In Proceedings of the Conference on Artifical Reality and Teleexistence (ICAT94), pp. 4962. (1994)*

[10] *The Wilcoxon Matched-Pairs Signed-Ranks Test*, `www.fon.hum.uva.nl/Service/Statistics/Signed_Rank_Test.html`

[11] *Dealing with 'Outliers': Maintain Your Data's Integrity*, `http://rfd.uoregon.edu/files/rfd/StatisticalResources/outl.txt`

[12] H. Richter, S. Lhmann, F. Weinhart, and A. Butz, *Comparing direct and remote tactile feedback on interactive surfaces. In Proceedings of the International Conference on Haptics - EuroHaptics 12, EuroHaptics12, pages 301313. (2012)*

[13] A. Faeth, C. Harding *Effects of Modality on Virtual Button Motion and Performance. In Proceedings of the 14th ACM international conference on Multimodal interaction,ACM New York, NY, USA, P 117-124, (2012)*

[14] RJ.Kosinski,*A literature review on reaction time.* `http://biae.clemson.edu/bpc/bp/Lab/110/reaction.htm`

[15] *Smartphones*, `http://en.wikipedia.org/wiki/Smartphone`

[16] *Smartphones now outsell 'dumb' phones*, `http://www.3news.co.nz/Smartphones-now-outsell-dumb-phones/`

tabid/412/articleID/295878/Default.aspx

[17] *Nokia revenues slide 24% but Lumia sales rise offers hope*, Charles Arthur, The Guardian. Retrieved 19 July 2013

[18] A. Sears, C. Plaisant, B. Shneiderman,*A new era for high-precision touchscreens. Advances in Human-Computer Interaction, vol. 3, Hartson, R. Hix, D. Eds., Ablex (1992) 1-33 HCIL-90-01, CS-TR-2487, CAR-TR-506. (1990)*

[19] R. Azuma,*A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 1997, pp. 355385. (1997)*

[20] D. Wagner,*History of mobile augmented reality.* https://www.icg.tugraz.at/ daniel/HistoryOfMobileAR/

[21] *Augmented reality*, http://en.wikipedia.org/wiki/Augmented_reality

[22] W. Hürst and K. Vriens,*Mobile Augmented Reality Interaction via Finger Tracking in a Board Game Setting. MobileHCI2013_AR-workshop, Designing Mobile Augmented Reality(2013)*

[23] W. Hürst, C. van Wezel,*Multimodal interaction concepts for mobile augmented reality applications. In Proceedings of the 17th international conference on Advances in multimedia modeling - Volume Part II (2011)*

[24] Jennifer L. Burke , Matthew S. Prewett , Ashley A. Gray , Liuquin Yang , Frederick R. B. Stilson , Michael D. Coovert , Linda R. Elliot , Elizabeth Redden,*Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. Proceedings of the 8th international conference on Multimodal interfaces, November 02-04, 2006, Banff, Alberta, Canada (2006)*

[25] Mike Koenig,*Sounds used in the application* https://www.soundbible.com

Appendix A

Questionnaire Experiment 1

Participant number:

Name:

Age:

Gender:

Left/Right handed:

Previous experience with Augmented Reality on mobile:

Previous experience with mobile gaming, if so what kind:

Additional info:

Device held in which hand:

Remarks during the experiment:

Remarks after the experiment:

Questionnaire

Part 1

How do you feel about the gesture-based interactions?

Easy [0][1][2][3][4][5][6][7][8][9][10] Difficult

While the experiment progressed, did you feel like you got more in control of the gesture-based interaction?

None[0][1][2][3][4][5][6][7][8][9][10] Yes, a lot

Did you experience any kind of fatigue or discomfort?

None [0][1][2][3][4][5][6][7][8][9][10] Yes, a lot

In what kind of game would you like to see/could you see this kind of interaction work?

Additional comments?

Part 2-3

Do you think if implemented differently, the gesture-based interaction could be improved, and if so, how?

Additional comments?

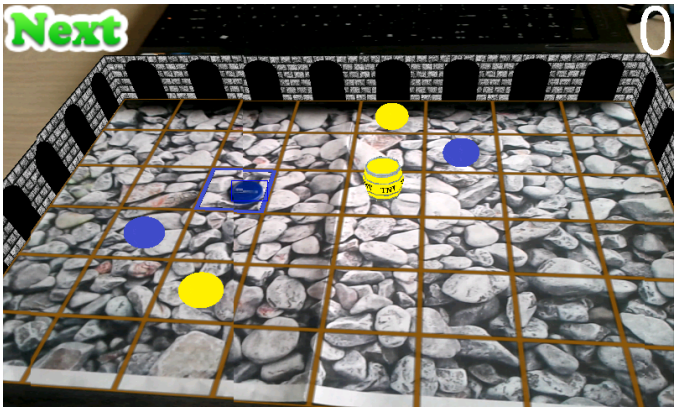
Appendix B

Introduction Experiment 1

Introduction

Thank you for participating in this experiment. The goal of this experiment is to acquire information about the usage of a number of gesture-based interactions in relation to an AR game. The experiment is divided into three parts. In the first part you will play a game and use fingermarkers to interact with virtual objects. In the second part you will use the touchscreen to perform a number of interactions on virtual objects and in the third part you will be performing the same interactions but using fingermarkers again.

The experiment consists of a small game, in which blue and yellow balls will appear and/or roll from one side of the board to the other side. It will be your job to destroy as many of the balls as possible, but be careful! If two balls are on the same square they will destroy each other and no points will be awarded.



In this screenshot you see a physical blue coin(with a virtual blue border) and a virtual yellow barrel. To destroy the blue balls you have to physically move the blue coin over the game-board to where the blue ball is or will be. To destroy the yellow balls you have to move the virtual yellow barrel to where the yellow ball is or will be. This is done by either dragging it across the touchscreen(second part of the experiment), or grabbing it with your hand and moving it in front of the mobile phone's camera(first and third part of the experiment). In the screenshot you can also see a grid. Both the blue physical object as well as the yellow virtual barrel will always belong to one square in this grid and will snap to the center of a square.

The experiment features a set of levels, all preceded by a practice level and some text explaining you what to do.

If there are any questions during the experiment or you are experiencing discomfort of any kind, you can pause the game by pressing the button circled with green below and/or ask the surveyor.

Appendix C

Questionnaire Experiment 2

Name:

Age:

Gender: Male / Female

Handedness: Left / Right / Both

Experience with natural interfaces(e.g. Kinect style gesture interaction, midair gestures via a mobiles camera, etc.): [Yes / No] If yes, what kind?

Part 1: (To be filled out by the participant after the first experiment)

In which hand did you hold the device (cross out the other one):

[Left / Right]

In this test, you experienced three types of feedback (and combinations thereof):

- (V) Visual
- (A) Audio
- (H) Haptic (vibrations)

We want to know now, which one you prefer with respect to certain criteria.

(1) Please rate the types of feedback from 5 = best to 1 = worst (identical ratings for different feedback types are possible):

- Visual Rating = [1] [2] [3] [4] [5]

- Audio Rating = [1] [2] [3] [4] [5]
- Haptic Rating = [1] [2] [3] [4] [5]
- Visual + Audio Rating = [1] [2] [3] [4] [5]
- Visual + Haptic Rating = [1] [2] [3] [4] [5]
- Audio + Haptic Rating = [1] [2] [3] [4] [5]
- Visual + Audio + Haptic Rating = [1] [2] [3] [4] [5]
- No feedback: Rating = [1] [2] [3] [4] [5]

(2) Can you explain your answers? (particularly the ones with the highest and lowest ratings) .....

Part 2:
(To be filled out by the participant after the next 2 experiments)

In this test, you experienced three types of feedback (and combinations thereof):

- (V) Visual
- (A) Audio
- (H) Haptic / vibration

In addition to this the feedback had two different durations.

- (C) Constant
- (T) Temporary

We want to know now, which one you prefer with respect to certain criteria.

1) Lets first look at individual feedback (no combinations of feedback types):

Please rate the types of feedback from 5 = best to 1 = worst (identical ratings for different feedback types are possible):

- (V) Visual Rating = [1] [2] [3] [4] [5]
- (A) Audio Rating = [1] [2] [3] [4] [5]
- (H) Haptic Rating = [1] [2] [3] [4] [5]

Which of the two implementations did you prefer (cross out the other one):

- (V) Visual : (T) temporary OR (C) constant?
- (A) Audio : (T) temporary OR (C) constant?
- (H) Haptic : (T) temporary OR (C) constant?

2) Now lets look at combinations of two types of feedback:

Please rate the types of feedback from 5 = best to 1 = worst (identical ratings for different feedback types are possible):

- (A-V) Audio and Visual Rating = [1] [2] [3] [4] [5]
- (A-H) Audio and Haptic Rating = [1] [2] [3] [4] [5]
- (V-H) Visual and Haptic Rating = [1] [2] [3] [4] [5]

For each combination, what implementation did you prefer (cross out the other ones):

- (A-V) Audio: [T] temporal or [C] constant feedback?
- (A-V) Visual: [T] temporal or [C] constant feedback?
- (A-H) Audio: [T] temporal or [C] constant feedback?
- (A-H) Haptic: [T] temporal or [C] constant feedback?
- (V-H) Visual: [T] temporal or [C] constant feedback?
- (V-H) Haptic: [T] temporal or [C] constant feedback?

Did you focus on one type of feedback in particular? .....

3) Now lets look at combinations of three types of feedback and no feedback at all:

Please rate the types of feedback from 5 = best to 1 = worst (identical ratings for different feedback types are possible):

- (A-V-H) Audio, Visual and Haptic Grade = [1] [2] [3] [4] [5]
- No feedback -Grade = [1] [2] [3] [4] [5]

What implementation did you prefer (mark with a cross):

- (A-V- H)Audio: [T] temporary or [C] constant?

- (A-V- H) Visual: [T] temporary or [C] constant?
- (A-V- H) Haptic: [T] temporary or [C] constant?

Did you focus on one type of feedback in particular?
.....

Looking back at the best of the single, double and triple feedback types, which one has your preference and why?
.....

Which one did you like least of all and why?
.....

When the mobile phone vibrated, did you realize it was because you selected or deselected an object.
No, not at all [1] [2] [3] [4] [5] Yes, immediately

How would you rate the haptic feedback in the hand holding the device, with regards to naturalness when selecting or deselecting objects.
It felt wrong [1] [2] [3] [4] [5] It felt natural

How would you rate the haptic feedback in the hand holding the device, with regards to immersiveness when selecting or deselecting objects.
I felt less immersed [1] [2] [3] [4] [5] I felt more immersed

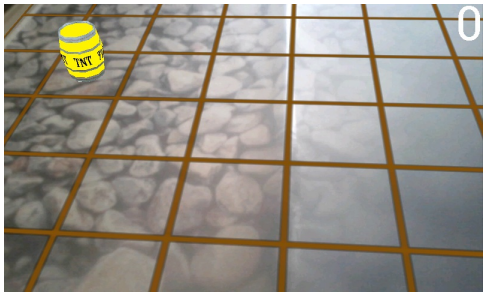
Do you miss anything/what would you have changed?
.....

Appendix D

Introduction Experiment 2

Introduction

Thank you for participating in this evaluation. The goal of this experiment is to acquire information about multimodal feedback for gesture-based interactions in an AR setting.



In the above screenshot you see a virtual yellow barrel, which can be manipulated by selecting it. To do this you have to move your thumb and index-finger together at the place of the object, like you would if you would grab a physical object. When both the markers on your fingers connect with the object it will be selected. Once selected, it will follow your hand until deselected. This is done by moving the fingers away from each other. Once the markers on the fingers are a set distance away from each other the object will be deselected.

The Experiment

The experiment is divided into three tasks. Preceding the experiment is a practice level where you can practice the interactions described above. Every task starts with a simple text explaining what to do. Finishing the

task requires you to complete a number of selection and deselection actions. Each time, different multimodal feedback will be provided. The goal is to complete the tasks as fast as possible.

DISCLAIMER: If there are any questions during the experiment or you are experiencing discomfort of any kind, you can pause the game by pressing the button circled with green as shown below and/or ask the surveyor.