

Interactive augmented reality

Roger Moret Gabarró

Supervisor: Annika Waern

December 6, 2010

This master thesis is submitted to the Interactive System Engineering
program.

Royal Institute of Technology

20 weeks of full time work

Abstract

Augmented reality can provide a new experience to users by adding virtual objects where they are relevant in the real world. The new generation of mobile phones offers a platform to develop augmented reality application for industry as well as for the general public. Although some applications are reaching commercial viability, the technology is still limited.

The main problem designers have to face when building an augmented reality application is to implement an interaction method. Interacting through the mobile's keyboard can prevent the user from looking on the screen. Normally, mobile devices have small keyboards, which are difficult to use without looking at them. Displaying a virtual keyboard on the screen is not a good solution either as the small screen is used to display the augmented real world.

This thesis proposes a gesture-based interaction approach for this kind of applications. The idea is that by holding and moving the mobile phone in different ways, users are able to interact with virtual content. This approach combines the use of input devices as keyboards or joysticks and the detection of gestures performed with the body into one scenario: the detection of the phone's movements performed by users.

Based on an investigation of people's own preferred gestures, a repertoire of manipulations was defined and used to implement a demonstrator application running on a mobile phone. This demo was tested to evaluate the gesture-based interaction within an augmented reality application.

The experiment shows that it is possible to implement and use gesture-based interaction in augmented reality. Gestures can be designed to solve the limitations of augmented reality and offer a natural and easy to learn interaction to the user.

Acknowledgments

First of all I would like to thank my supervisor and examiner, Annika Waern, for her excellent guide, help and support during the whole project. I really appreciate the chance of working in such an interesting topic and in a very nice team. A very special thanks also to all the people in the Mobile Life center for all the memorable moments I had during these nine months.

This thesis would not have been done without all the anonymous volunteers that participated in the studies I carried out for my work. Thanks to their excellent work and their valuable input this thesis succeeded. A very special thanks to all the personnel in the Lava center, in Stockholm, for its admirable support.

I would like to thank all the friends I made in Sweden. These two years would not have been the same without all of you! Specially, I would like to thank Sergio Gayoso and Jorge Sainz for all the great moments we spent together traveling, having dinner, going around or simply talking somewhere. “Muchas gracias!”

A very special thanks to all my friends from Barcelona. Even though I am far away from home, they still supported and cared about me during these two years. Specially, I would like to thank David Martí, Eva Jimenez for all the hours we spent chatting and Elisenda Villegas for our long, long e-mails. “Moltes gràcies!”

I want to dedicate this thesis to my family and thank the unconditional support and help I always received from my parents, Francesc and Gloria, and my sister, Laia. I really appreciate her efforts for correcting my thesis, encouraging me in the bad moments to go on and always having some time to listen to me when I needed to talk. “Moltes gràcies!”

Finally, my most special thanks to my girlfriend, Marta Tibau, for her patience, kindness, unconditional support, comprehension and help. “No ho hauria aconseguit sense tu! Moltíssimes gràcies!”

Contents

1	Introduction	6
1.1	Motivation	6
1.2	Goal	6
1.3	Delimitation	6
1.4	Approach	6
1.5	Research methodology	7
1.6	Results	7
2	Background	8
2.1	User-centered design	8
2.2	Gesture-based interaction	8
2.2.1	Glove-based devices	9
2.2.2	Camera tracking systems	9
2.2.3	Detecting gestures on portable devices	9
2.3	Augmented reality	10
2.3.1	Mobile augmented reality	10
2.3.2	Interaction with AR applications	11
3	Gesture study	12
3.1	Purpose	12
3.2	Repertoire of manipulations	12
3.3	Design of the study	12
4	Design over the gesture repertoire	15
4.1	Selection criteria	15
4.1.1	Technical feasibility	15
4.1.2	Consistency	15
4.1.3	Majority's will	15
4.2	Results	15
4.2.1	Lock and unlock	16
4.2.2	Shake	16
4.2.3	Enlarge and shrink	16
4.2.4	Translate to another position	20
4.2.5	Move towards a direction	20
4.2.6	Pick up	20
4.2.7	Drop off	21
4.2.8	Place	21
4.2.9	Rotate around the X, Y or Z axis	21
4.2.10	Rotate around any axis	21
4.2.11	Rotate a specific amount of degrees around any axis	22
4.3	Resulting repertoire	22

5	Implementation	24
5.1	Platform	24
5.2	Design decisions	24
5.2.1	Manipulations	24
5.2.2	Interface	25
5.2.3	Position of the mobile	25
5.3	The application	25
5.3.1	Control of the camera	26
5.3.2	Capturing events	26
5.3.3	Marker detection	27
5.3.4	Analysis of the sensors data	27
5.3.5	Combining the gesture recognition methods	28
5.3.6	Showing the results	28
5.4	Implementation of the gestures	29
5.4.1	Lock and unlock	29
5.4.2	Enlarge and shrink	30
5.4.3	Rotate around the X axis	30
5.4.4	Rotate around the Y and the Z axis	31
6	Evaluative study	32
6.1	Purpose	32
6.2	Design of the study	32
6.3	Results	33
6.3.1	Understanding and learning to use the AR application	33
6.3.2	Usage experience	34
6.3.3	Gestures for non-implemented manipulations	34
6.4	Analysis	36
6.4.1	Performative gestures	36
6.4.2	Robustness and adaptability	36
6.4.3	Manipulations' preference for each gesture	37
6.4.4	Usability issues	37
6.4.5	Methodology used for designing the gesture repertoire	38
7	Conclusions	39
7.1	Summary	39
7.2	Discussion and conclusion	39
7.3	Future work	40
A	User study	44
A.1	Questionnaire	46
B	Evaluative study	47
B.1	Questionnaire	48

1 Introduction

1.1 Motivation

In the last few years, augmented reality (AR) has become a big field of research. Instead of involving the user in an artificial environment, as virtual reality does, augmented reality adds or removes information from the real world [1]. Being aware of the real world while interacting with virtual information offers a wide range of possibilities.

The new generation of portable devices, specially mobile phones, brings AR everywhere. Camera, sensors and compass are integrated in modern phones. There are some commercial applications which take advantage of modern phones and augmented reality. Layar¹ or Wikitude² provide information about which services are around you.

However, the main problem for augmented reality applications is how to interact with the virtual information. The examples mentioned above use buttons or the touchscreen to interact with the information displayed on the screen. Other applications could show, instead of information, 3D objects to the user. How would we interact with these objects? Is there a natural interaction technique?

1.2 Goal

The goal of this thesis is to explore the possibilities of using a gesture-based interaction with an augmented reality application. This includes an analysis of its feasibility, learnability and facility of use.

1.3 Delimitation

This thesis is focused on mobile augmented reality (mobile AR). Mobile AR brings augmented reality on portable devices such as mobile phones or PDAs.

In this thesis, an iPhone is used in the initial user study and a Nokia n900 mobile phone is used for implementing and testing a gesture repertoire which could potentially be used as a standard set of gestures for future mobile augmented reality applications.

1.4 Approach

The first step of this thesis was to define a set of manipulations with the virtual content, and to conduct a user study to get feedback on which gestures users would like to perform to interact with this virtual content. We believed that building the gesture repertoire based on user's experience was the best approach to get an intuitive, easy to learn and perform set of gestures.

¹<http://www.layar.com> - 15th of november of 2010

²<http://www.wikitude.org> - 15th of november of 2010

Once the study was done, the data collected was analyzed in order to get a consistent repertoire of gestures. According to the results of this study, a demo application was designed, implemented and evaluated in a second study. The reason for doing an evaluative study was to test the accuracy and robustness of the gestures. On the other hand, we wanted to evaluate the methodology used to define the repertoire of gestures. By comparing the results from both studies, we would verify if the results from the first study were accurate. Finally, the study also evaluated the learnability of the application, which was a secondary goal of this thesis.

1.5 Research methodology

This thesis is focused on the design study of an AR application which uses gestures as interaction method. The opinion of the users is really important to create a natural interaction with the application. Thus, iterative design [18] is an appropriate methodology to fulfill the goals of this project. Among other characteristics, iterative design motivates to get user feedback [19] in different stages of the project which is really important to get a natural and intuitive interaction with the AR object.

As explained above, users gave feedback in a user study where the application's operation was simulated according to the author's vision. A first version of the application was implemented upon the results of the user study. This prototype was tested again in a new study to check if the design worked as expected and to find usability problems.

1.6 Results

As it will be described more deeply in the coming sections, the user study succeeded, not only because participants suggested gestures for each presented manipulation, but also because the chosen methodology worked well. Users understood the task they had to do and the evaluator was able to communicate the manipulations to them.

From the collected data, a consistent repertoire of gestures was created for almost all the manipulations we had defined previously. A part of this set was implemented in a demo application which was used for the evaluative study.

The evaluative study showed the feasibility of the application, although not all the gestures were robust enough. Despite the accuracy problems, most of the participants were able to use the application themselves. Some instructions should be given to them in order to perform different gestures. The results also showed that they could guess which kind of manipulations someone was doing by just looking how s/he performed gestures.

2 Background

2.1 User-centered design

In the design of any product, from a telephone to a software for a computer, it has to be taken into account who will use it. User-centered design aims to design for the final user. In the book “The design of everyday things”, [21] Norman says that user-centered design is “a philosophy based on the needs and interests of the user, with an emphasis on making products usable and understandable.” According to Norman, user-centered design should accomplish the following principles:

- Use both knowledge in the world and knowledge in the head.
- Simplify the structure of tasks.
- Make things visible: bridge the gulfs of Execution and Evaluation.
- Get the mappings right.
- Exploit the power of constraints, both natural and artificial.
- Design for error.
- When all else fails, standardize.

These principles reinforce the use of gestures as an interaction technique as we apply them in our everyday activities to interact with the world. They simplify the interaction structure because each gesture is mapped directly to a manipulation. This interactivity is visible for the user as well as for the third parties observing him or her.

Many user interfaces in mobile devices tend to be suspensful, that is, the interaction is visible for third parties, but the effect of this interaction is not [20]. This fact imposes a limit on the learnability of the application, as people have to use it themselves in order to learn how it works. However, a gesture-based interaction could be more performative than any other interaction technique. The interaction would be visible and the effects of this manipulation partially deductible. Thus, it would be easier to learn how to use the AR application.

2.2 Gesture-based interaction

In the field of Human-computer interaction, many efforts on research have focused on implementing natural and common ways of interaction. There have been approaches in voice recognition, speech, tangible devices and gesture recognition.

A gesture recognition system aims to interpret the movements done by a person. Most of the research has focused on recognising hand gestures. There are two main research streams: the so-called glove-based devices and the use of cameras to capture movements.

2.2.1 Glove-based devices

Researchers have developed many prototypes of hardware which the user wears as a glove to recognise how the hand is moved [23, 5]. This technique uses sensors to recognise the angle of the joints and the accelerations when the fingers are moved. As Sturman and Zelter said in [5], “We perform most everyday tasks with them [our hands]. However, when we worked with a computer or computer-controlled application, we are constrained by clumsy intermediary devices such as keyboards, mice and joysticks.” Although it is a more natural interaction, it still requires the use of a glove-based device to recognise the movements. So, users are still using, or in this case, wearing this device to interact with the application. Using the movements of a mobile phone as an input reduces the number of devices to only one. Users interact with it at the same time that they observe the results of the movements on the same device. Moreover, a mobile phone is a common device that users already have, which reduces the cost of the application.

2.2.2 Camera tracking systems

Another approach is to use cameras to recognise the movements done in its viewport. These applications use algorithms that recognise the shape of a hand, for example, and by comparing its shapes in different frames, the application can determine the movement of the hand. Some applications track the hands by analysing the colors [7], while some others add a reference point in the real space[6].

2.2.3 Detecting gestures on portable devices

In the last five years, the increase of the computational power, the integration of cameras and sensors of different kinds in portable devices, have opened a wide range of possibilities. The most modern mobiles already use simple gestures, such as tilting the mobile or shaking it.

The two techniques explained above are also used in mobile phones [8, 9, 10, 11]. The difference relies on the fact that the sensors and the camera integrated in the mobile phone are used to detect the movements of the device.

There are many applications that use the camera to recognise directions or shapes. For instance, Wang, Zhai and Canny developed a software approach, implemented in different applications where they could indicate directions as if they were using the arrows of the keyboard or even write characters [8]. Other approaches divided the space in 4 directions and the combination of a set of directions permit to recognise more complex patterns like characters [9].

Accelerometers permit the detection of more precise gestures. Applications using them can recognise specific movements done with the mobile phone [10, 11]. Even though, image processing systems still have some advantages over the sensors systems. If there is an easily detectable spot on the camera’s viewport, it can simplify the recognition task [6].

2.3 Augmented reality

The concept of augmented reality was introduced by Azuma [1] in his paper “A survey of Augmented Reality”. Augmented reality is the modification of the

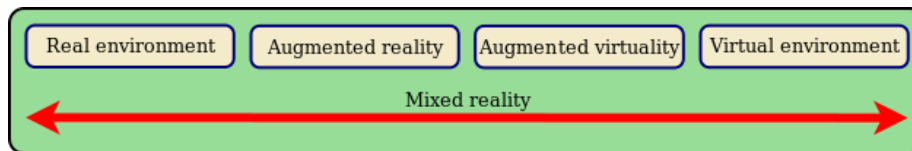


Figure 1: Classification of realities and virtualities within mixed reality

real world by adding or removing content from it. Although it is related to the visual sense, it could be applied to any other. According to Azuma [1], an AR application have the following requirements:

- Combine real and virtual objects
- Interactivity in real time
- Registered in 3D

Ideally, it should not be possible to distinguish between real and virtual elements shown on the application. This motivates the use of natural ways of interaction with these objects to make the experience as realistic as possible.

Milgram and Kishino set augmented reality as a specific case of Mixed reality [2]. According to them, mixed reality includes different kinds of realities and virtualities, as shown in the figure 1.

Virtual reality isolates the user from the real world and prevents him or her to interact with it. In an AR application, users are aware of the real world while they interact with it and the virtual content added to it.

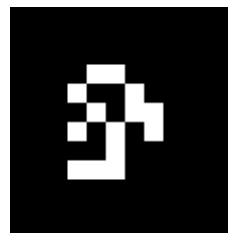


Figure 2: Fiducial marker

2.3.1 Mobile augmented reality

Rohs and Gfeller introduced the concept of mobile augmented reality [4]. Instead of using special hardware to build an AR application, they proposed to use the new generation of portable devices. The increase on the computational power, the camera’s resolution on portable devices made possible to implement these kind of applications on them.

In order to build mobile AR applications, Rohs simplified the task of recognising a spot on the image by using fiducial markers [3]. A fiducial marker (see figure 2) is 2-dimensional square composed of black and white fields. Thus, the application looks for a specific pattern on the screen. From the fiducial marker, the application can determine the position on the screen, the orientation and the scale.

Mobile augmented reality is an important field of research for its potential and feasibility to build commercial applications. It uses common hardware, which makes it cheaper for the final user.

2.3.2 Interaction with AR applications

One of the main problems that AR applications have is how to interact with the virtual information. There have been some approaches and clumsy solutions to this problem. The most common is to use buttons. The remote chinese game [12] and Bragfish [16] are two examples of this approach. In both cases, users have to use on-screen buttons to interact with the game. Other applications are designed so that they have a very low interaction. Photogeist [13] is a game about taking pictures of ghosts that appear and disappear over a matrix of markers. The game is played by clicking to take photos. This game could have a wider and more complex interaction giving more possibilities of interaction to the user.

The treasure game [14] uses a completely different approach. The game requires to pick up virtual objects from the marker. In order to perform this action, a second marker is used to indicate a pick up action. This is not feasible if the application has many means of interactions as there should be one marker for each.

The most advanced approach in terms of interaction in an AR application was done by Harvainen et al [15] who built two AR applications which used simple gestures to interact with. One application permits the user to explore a virtual model of a building. By tilting the mobile, the user can change the view mode. The other application present a simple interaction with a virtual dog. By moving the mobile closer, farther or tilting, the dog perform different actions.

This thesis does not present a solution for a specific application. Instead, it aims to define a natural, learnable and intuitive repertoire of gestures to interact with the virtual content presented in an AR application.

3 Gesture study

3.1 Purpose

This project aimed to develop an application to manipulate a virtual object through gestures. Each manipulation should be invoked by a gesture with a mobile phone.

Instead of defining the gestures for each manipulation ourselves, a user study was done in order to know how people would like to interact through gestures with the mobile phone. Thus, we assured that the gestures implemented would have a real percentage of acceptance among the potential users of the application.

3.2 Repertoire of manipulations

Before doing the study, a set of manipulations needed to be defined. The set of manipulations was inspired by previous work done in this field which accomplish the following characteristics: the manipulations should be simple and generic. This set would be used as the input in the study. Participants should suggest gestures for each manipulation.

In table 1, there is a description of the manipulations designed for the study. In order to make the descriptions more comprehensible, four coordinate-systems are used:

- GFrame: the global framework
- OFrame: the framework with origin in the virtual object
- CFrame: the framework with origin in the camera of the phone
- UFrame: the framework with origin in the user's point of view

The OFrame is fixed to another framework depending on the manipulation.

3.3 Design of the study

The repertoire of manipulations defined in the previous section was used in a qualitative study to explore which gestures users prefer to perform for each interaction with the AR object. The study did not aim to have a large group of participants (see the results in section 4.2). Instead, it should be possible to detect patterns on the gestures to know the preference of the users. Thus, a qualitative study is the most appropriate option. Participants were selected to have some experience on mobile devices, but not necessarily in AR applications.

The user study was divided in two parts. First, the manipulations were presented to the participants and they should suggest a gesture to invoke each manipulation. Secondly, they should fill in a questionnaire.

As the application was not implemented, its behavior was simulated. Participants used an iPhone with the camera enabled. Thus, they had a view of

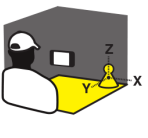
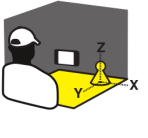
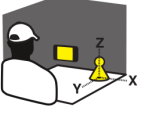
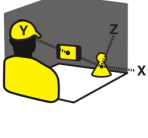
Action	Description	Reference framework
Lock / Unlock	Enables and disables the gesture interaction	
Shake	Gives a momentum to the object	
Enlarge	Makes the object bigger	
Shrink	Makes the object smaller	
Translate to another position	Moves the object from the marker to another position	
Move towards a direction	Moves the object towards a direction on the marker's plane	
Pick up	Collects an object from a marker to the phone	
Place	Places an object from the phone to a marker	
Drop off	Drops off an object from the phone to a marker	
Rotate around the X axis	Rotates around the X axis	
Rotate Around the Y axis	Rotates around the Y axis	
Rotate around the Z axis	Rotates around the Z axis	
Rotate around any axis	Rotates around any axis in the space	
Rotate XX° around any axis	Rotates an amount of degrees around any axis in the space	

Table 1: Definition of the manipulations with the virtual object

the real world on the screen while using the mobile. On the table, there was a fiducial marker. The evaluator was manipulating a real object on the marker to represent the interactions with the AR object. Figure 3 shows the set up of the study.

There were some restrictions on how users could interact with the virtual object in the study. It was as important to orientate the users on how they should interact with the simulated application as to not impose them too many limitations. Participants should focus on the marker most of the time to see what would happen to the AR object. On the other hand, keeping the marker always on the screen could exclude too many gestures. In order to balance these two premises, they were allowed to point somewhere else while performing a gesture as long as the marker was in the camera's viewport, at least, at the beginning or at the end of the performance of the gesture.



Figure 3: Set up of the user study

Users were also allowed to use the screen as a button. This was included because it could be difficult to figure out how to interact with the virtual object only with gestures. On the other hand, it was limited to be used as a button because gestures with the phone should be the main interaction.

Users should think for the best gesture for each kind of manipulation. They were not asked to create a consistent set of gestures for all the manipulations presented.

Users were asked to think aloud how they would provoke each manipulation by moving the mobile phone. They should try different options and perform the chosen one three times.

In the questionnaire, they were asked about other possible manipulations, which gestures were more and which were less natural and intuitive, which kind of information they would like to have on the screen and about having different modes. As an application with all the manipulations implemented could be difficult to use, a possibility was to divide the gestures in two subsets or modes. By switching from one mode to another, the manipulations available would change.

Each session lasted between 30 and 40 minutes and was recorded for a subsequent analysis. The outline of the study and the questionnaire is available in the appendix A.

4 Design over the gesture repertoire

4.1 Selection criteria

Before starting to analyze the data collected in the study, a list of criteria were defined to prioritize and discard the gestures.

4.1.1 Technical feasibility

The computational power and the sensors limit what could be done with the mobile. Being able to recognize a gesture with the mobile resources was the main criterion for discarding or choosing gestures.







Description	Figure
Press and hold	
Release	
Click	
Move constrained by the indicated axis	
Rotate in the indicated directions	
Hold still for a period of time	

Table 2: Icons with primitive phone movements adopted from Rhos and Zweifel [17]. Multiple arrows indicate that the gesture can be performed in any combination of the indicated directions.

with the phone for all of them.

In table 2 is defined a graphical language which will be used to describe the gestures proposed by users. This language is based on the work of Rohs and

4.1.2 Consistency

The study included 14 manipulations with the AR object presented previously in the table 1. Participants could suggest the same gesture with the phone to invoke different actions. However, the final gesture repertoire had to be consistent so that all the gestures could be implemented in the same application.

4.1.3 Majority's will

The last criterion was related to the number of participants proposing one gesture. In case of inconsistency, the largest number of people would be determinant to choose between two options.

4.2 Results

Fourteen people participated in the study, 9 women and 5 men, aged between 20 and 37. All of them were familiar with modern mobile phones and some of them knew what augmented reality was. For those who did not know it, a small introduction was given by showing videos of AR applications.

Participants understood the manipulations the evaluator was doing with the real object and they were able to suggest gestures

Zweifel [17]. As the icons represent primitive movements and some gestures are more complex, they are represented by a sequence of icons.

The following sections analyze deeply the most interesting results of the study which are summarized in tables 3, 4 and 5.

4.2.1 Lock and unlock

Ten out of the fourteen participants suggested to make a simple click on the screen to lock onto the AR object and another click to unlock it (1.1 in table 3). It is a simple interaction which does not involve gestures. In this case, a non-gesture-based interaction is acceptable as this manipulation enables or disables the gestures.

Two minor alternatives were suggested by two participants each: tapping the virtual object (1.2 in table 3) and moving closer and farther from the object (1.3 in table 3). The first one is implementable and relies on the idea of waking up the virtual object by tapping it softly. The option 1.3 in table 3 is also implementable. The option 1.1 in table 3 is chosen due to its large support.

4.2.2 Shake

In order to shake the AR object, 5 users proposed to 'tilt-tilt back' the phone around the Z axis (2.1 in table 3), while another 4 suggested the same but around the Y axis (2.2 in table 3). After a deep analysis of the videos, we realize that in both cases they imitated the shaking of the virtual object with the mobile. The difference, though, is that the first group hold the mobile on one side of the AR object and the second group hold it on top. This change on the perspective is the cause of the two different shaking. However, the idea behind those movements is the same: shake the mobile the same way you want the object to shake.

The option 2.3 in table 3 was selected by 3 users who shaked the mobile by moving it to the right and left repeatedly

4.2.3 Enlarge and shrink

There was only one main option to change the size of the object. The idea was to press the screen, change the distance between the mobile and the marker to enlarge or shrink the AR object and release to stop it. It was done by seven people. However, five of them enlarged the object while moving closer to the marker (3.1 in table 3) and shrank it while moving farther (4.1 in table 3). The other two people did the opposite (3.2 and 4.2 in table 3).

Enlarging while getting closer is more natural and intuitive. One of the participants described it as "it is a way to increase the zooming". On the other hand, this could provoke that the user would not see the whole AR object while enlarging it, as it could be out of the camera's viewport.



#	Effect	Textual description	Graphical description	No.
1.1	Lock / Unlock	Click on the screen		10
1.2		'Tap' the object		2
1.3		Move closer and further to the marker		2
2.1	Shake	Shake around the Z axis		5
2.2		Shake around the Y axis		4
2.3		Move repeatedly to the right and left		3
3.1	Enlarge	Press, move closer and release		5
3.2		Press, move further and release		2
4.1	Shrink	Press, move further and release		5
4.2		Press, move closer and release		2

Table 3: Results from the user study













#	Effect	Textual description	Graphical description	No.
5.1	Pick up	Pick up gesture	-	4
5.2		Tilt the mobile around the X axis counter clockwise		3
5.3		Move the mobile upwards	-	2
5.4	Drop off	Move the mobile towards the user		2
6.1		Shake moving closer and farther from the marker		6
6.2		Fast movement closer and farther from the marker		6
7.1	Place	Move closer to the marker		5
7.2		Tilt around the X axis clockwise		2
7.3		Slowly drop off movement	-	2
8.1	Move to another position	Get closer, mirror mobile's movement, get further	 	3
8.2		Press, mirror mobile's movement, release	 	3
8.3		Click, mirror mobile's movement, click	 	3

Table 4: Results from the user study











#	Effect	Textual description	Graphical description	No.
9.1	Move towards a direction	Rapid movement to indicate a direction		7
9.2		Tilt the mobile to indicate the direction		4
10.1	Rotate around the X axis	Tilt around the X axis		11
11.1	Rotate around the Y axis	Tilt around the Y axis		8
11.2		Tilt around the Z axis		3
12.1	Rotate around the Z axis	Tilt around the Z axis		9
12.2		Tilt around the Y axis		4
13.1	Rotate around any axis	Tilt the mobile to indicate the direction		10
13.2		Combine the rotations around X, Y and Z axis	10.1 + 11.1 + 12.1	2
14.1	Rotate XX ^o around any axis	Press, mirror the mobile's rotation, release		6
14.2		Tilt the mobile to indicate the direction		3

Table 5: Results from the user study

4.2.4 Translate to another position

Nine of the participants suggested the following structure to translate the AR object: there was an event to start the manipulation, then the AR object followed the mobile's movement and at the end there was an event to stop the manipulation. They disagreed, however, on the events to start and stop the manipulation. There were 3 propositions supported by three participants each: get closer to the marker to start and farther to stop (8.1 in table 4), press to start and release to stop (8.2 in table 4) and click to start and to stop (8.3 in table 4). All of them are easy to use, learnable and implementable. However, as the click is used in the lock/unlock manipulation (1.1 in table 3) and press and release is used in the enlarge and shrink manipulations (3.1, 3.2, 4.1 and 4.2 in table 3), the option 8.1 in table 4 is chosen.

4.2.5 Move towards a direction

Seven out of the fourteen people suggested to use the phone's plane to indicate the direction by moving the mobile rapidly in the specified direction (9.1 in table 5). This could be implemented even though it would probably have a moderate precision.

An alternative proposed by four people was to tilt the mobile to indicate the direction (9.2 in table 5). This solution would have a very low precision as it is not possible to calculate the inclination of the mobile phone. It would be a good solution if just a few directions want to be implemented.

4.2.6 Pick up

Several options came out with the picking up manipulation. Three of the users suggested to tilt the mobile around the X axis counter clockwise (5.2 in table 4). Another two people proposed to move the mobile towards the user (5.4 in table 4). These gestures were suggested for other manipulations with a larger support from the participants. So, they are discarded for consistency reasons.

A third option to pick up the AR object was to move the mobile upwards (5.3 in table 4), done by two participants. The problem is that this gesture could change depending on the perspective and position of the person and the mobile.

The last option was to make a 'scooping up' gesture (5.1 in table 4). It got more support than any of the previous options, with four people. It is a natural, easy and intuitive way to pick up an object. However, it is not technically possible to be implemented. First of all, the data provided by three accelerometers is not enough to detect such a complex gesture. The second problem is that a 'scooping up' gesture can be performed in many ways. Thus, even if this gesture could be recognized, most of the users would have to learn the exact gesture to provoke the picking up of the virtual object.

4.2.7 Drop off

Most of the people, 12 out of 14, proposed to move closer and move farther from the marker to drop the AR object off. Six of them did this movement once (6.2 in table 4), while the other six did it many times (6.1 in table 4). It is a natural, easy and intuitive gesture to perform this manipulation.

4.2.8 Place

Five out of the fourteen users suggested to move the mobile very close to the marker to place a virtual object there (7.1 in table 4). This is not technically feasible as the tracker system can not work at a very short distances. On the other hand, by doing the same gesture but keeping a distance from the marker, it may not have the same effect that they described when doing this gesture.

An alternative done by three users was to tilt the mobile clockwise around the X axis (7.2 in table 4). Despite of its feasibility, it is discarded for consistency reasons.

The last one was to make the same gesture as for dropping off but more slowly (7.3 in table 4). This is not a solution itself, but depending on the gesture done for dropping off, a slower version for placing an object on the marker could be implemented.

4.2.9 Rotate around the X, Y or Z axis

For rotating the AR object around X, Y or Z axis, participants proposed to tilt the mobile around the same axis as the one used for rotating the virtual object. More precisely, 11 people did it for rotating around the X axis (10.1 in table 5), 8 for rotating around the Y (11.1 in table 5) and 9 for the Z (12.1 in table 5).

The rotations around the Y and Z axis had a second option, supported by three and four people respectively. In this case, users switched the axis: by tilting the mobile phone around the Y axis (12.2 in table 5), the virtual object rotated around the Z axis and by tilting the mobile phone around the Z axis (11.2 in table 5), the AR object rotated around the Y axis. As it happened with the shaking, the position of the mobile in relation with the marker provoked different gestures. But they imitated the rotation of the virtual object which means that if they had hold the mobile the same way as the rest of people, they would have moved the phone like 11.1 and 12.1 in table 5 respectively.

4.2.10 Rotate around any axis

Ten people suggested to tilt the mobile to indicate the direction of the rotation (13.1 in table 5). This option is discarded for technical reasons. It would have a very low precision as it is not possible to determine accurately which rotation the user is intending to do. Even for the user it would be difficult to make the gesture

Two participants proposed to combine the three simple rotations around X, Y and Z axis to perform any kind of rotation (13.2 in table 5). This is a good solution which uses the implementation of the three simple rotations.

4.2.11 Rotate a specific amount of degrees around any axis

Six out of the fourteen people suggested that the virtual object imitated the rotation done with the mobile (14.1 in table 5). More precisely, they would press the screen to start mirroring the rotation of the mobile and release to stop it. It is technically feasible, but it should be tested to see whether it is a good solution for a rotation around 180° . Another problem is that the result of the rotation would not be visible until the gesture is finished.

Three participants suggested to tilt the mobile to indicate the rotation's direction (14.2 in table 5). This solution would not be feasible for very precise rotations.

4.3 Resulting repertoire

From the data analyzed in the previous section, the final gesture repertoire is:

- By clicking on the screen will lock or unlock the AR object (1.1 in table 3). A non-gesture-based interaction is more appropriate to enable and disable the gestures.
- By 'tilting-tilting back' the mobile repeatedly around the Z axis, will shake the virtual object (2.1 in table 3). If a different effect to the virtual object wants to be implemented, the gesture with the phone would imitate how the AR object is shaken. This gesture has a clear mapping with its effect and was suggested by many users.
- By pressing, moving closer and releasing will enlarge the object (3.1 in table 3). The opposite direction will shrink it. However, the alternatives 3.2 and 4.2 in table 3 respectively are not discarded, as we want to test them in the real application. Most of the users pointed to any of these solutions. As the results of the study are not clear, both are selected to be tested in the next study.
- By getting closer to the marker, moving the mobile and moving farther away from the marker, the users will move the AR object to another position (8.1 in table 4). Any of the suggested gestures that have the same events structure could be implemented. However, this is the only gesture consistent with the rest of the repertoire.
- By moving the mobile fast on the phone's plane it will start a motion of the object in the direction in which the mobile is moved (9.1 in table 5). The plane of the phone is mapped directly to the plane of the marker. This gesture can offer a good precision in comparison with the alternatives.

- The pick up is excluded from the gesture repertoire. The results of the study showed that there is no gesture that surpasses all the selection criteria. In this case, some screen-based interaction will be used.
- By moving the mobile closer and farther from the marker, the object will be dropped off from the phone (6.1 in table 4). If the user does the gesture more slowly, the AR object will be placed on the marker (7.3 in table 4). Both gestures could be implemented. However, this gesture allows the user to see the result as it has to move the mobile twice in two directions (6.2 in table 4), while the alternative has to move the mobile an indefinite number of times.
- By 'tilting-tilting back' the mobile in one of the three axis, the object will start rotating. By doing the same gesture but in the opposite direction, the manipulation will stop (10.1, 11.1 and 12.1 in table 5). These gestures were, according to the criteria defined in section 4.1, the only feasible among the user's suggestions and suggested by a large number of participants.
- The other rotations (13 and 14 in table 5) are discarded as results showed that the previous rotations are more understandable.

5 Implementation

Once the study was finished, its results were used to develop an AR application which would use the gestures done by the participants in the study as the main interaction method. Ideally, the application should have implemented all the manipulations from the study. However, the limited time for development forced us to narrow down the implementation to a small set of interactions. More precisely, the lock/unlock system, the rotations around the X, Y and Z axis, enlarge and shrink the virtual object were the manipulations implemented in the demo application. The lock and unlock manipulations were necessary to control the application. The rotations were chosen as they got a large support of the users and probably, it would have a larger acceptance in terms of usability and learnability. Finally, enlarge and shrink were chosen to explore why opposite gestures for the same manipulation appeared in the user study.

5.1 Platform

The application was developed for the Nokia n900. This mobile phone uses a processor with ARM architecture and a graphical card with support for OpenGL ES 2.0³. It has an integrated camera of 5.0 megapixels and 3D accelerometers.

The Nokia n900 uses Maemo 5⁴ as operative system. This OS is based on a Debian Linux distribution.

5.2 Design decisions

5.2.1 Manipulations

In the application, users are able to enable and disable the gesture-based interaction, rotate the AR object around the X, Y and Z axis, enlarge and shrink it.

The rotations are implemented in two different manners: continuously or by steps. In the first one, the gesture provokes a rotation which will only stop by doing the gesture to rotate in the opposite direction. The steps rotation means that everytime a rotation gesture is performed, the AR object is rotated a small amount of degrees. The reason to implement both options is that even if the first one is more accurate, it can be more difficult to control as there is a small delay on the detection of the gesture. On the other hand, the second option is easier to control, but it does not allow precise movements. Both options were implemented to be tested in the evaluative study.

Enlarge and shrink are implemented so that the gestures to perform these manipulations can be switched. The user study showed that some of the participants did a gesture to enlarge, and some others did the same to shrink (see 3.1, 3.2, 4.1 and 4.2 in table 3). Both options are implemented to verify the results gotten in the first study.

³<http://www.khronos.org/opengles/>

⁴<http://maemo.org/>

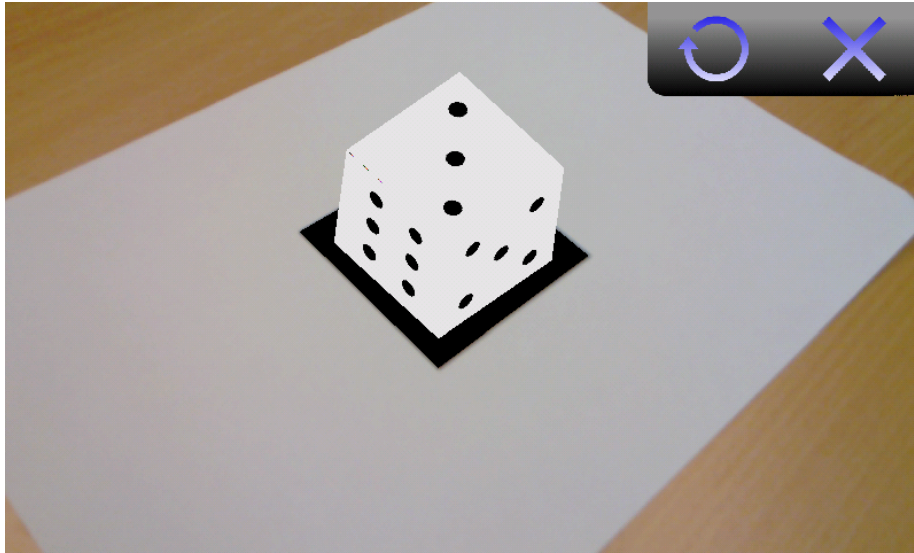


Figure 4: Screenshot of the application. The user interface has two buttons on the top right corner.

5.2.2 Interface

The graphic interface is reduced to two buttons on the screen. One of them is to reset the object to the original position and size and the other one to quit the application. The application is focused on the interaction with a virtual object in the real world. The screen is used to show the 'augmented' real world, so the interface should be as simple as possible. Figure 4 shows the application interface.

5.2.3 Position of the mobile

The position of the mobile is important to detect the gestures correctly. In the application, the mobile should be held horizontally with an angle between 25° and 75° with the plane of the marker, as shown in figure 5. Smaller or bigger angles could provoke problems in the detection of the gestures which use the data from the accelerometers.

5.3 The application

The application has, as shown in figure 6, the following functionalities:

- Capture the events on the keyboard and the screen
- Detect a marker on the frames
- Analyze the sensors data

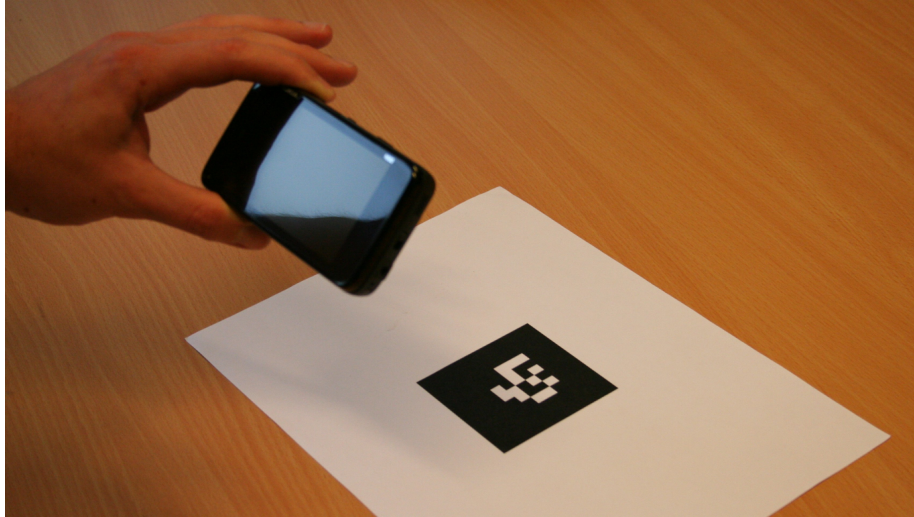


Figure 5: This image shows the appropriate angle of the mobile to detect the gestures correctly

- Generate the output

5.3.1 Control of the camera

Maemo 5 uses the library GStreamer⁵ to access and control the camera. The camera is initialized in the application, and every new frame available is used to detect a marker and shown on the screen as an output together with the AR object, if it is visible.

5.3.2 Capturing events

There are two kinds of events to be captured in the application: screen events and keyboard events.

The screen is used as a help to manipulate the AR object through gestures. The application distinguishes between three kinds of events on the screen: click, press and hold, and release. When a click is done over the area of the buttons, the manipulation with the AR object is ignored because the buttons have preference.

The keyboard is used to change some configuration parameters of the application, such as switching the effect on the AR object induced by a gesture with the phone.

⁵<http://www.gstreamer.net/> - 15th of November of 2010

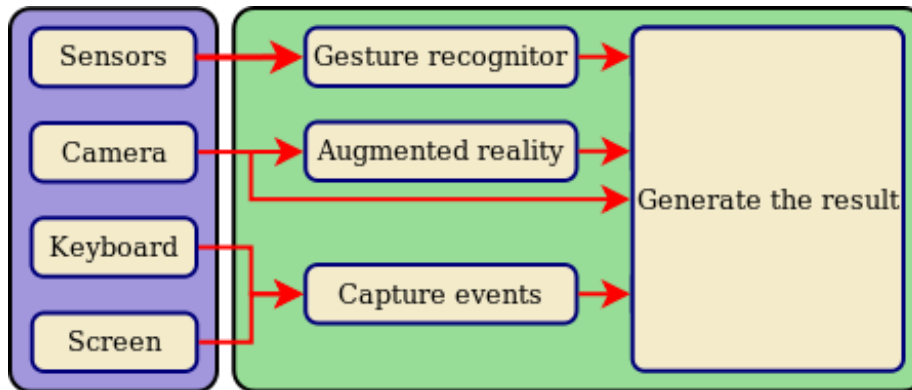


Figure 6: Schema of the application

5.3.3 Marker detection

An important design decision was to choose between marker-based augmented reality and markerless tracking. Marker-based augmented reality has the advantage that it is easier to put an AR object in a specific place in the real world. In the project, augmented reality is used as a tool and, thus, marker-based augmented reality allows to focus all the efforts on the interaction with the AR object.

The library ARToolKitPlus 2.2.0⁶, which is available in the repositories for the maemo 5 platform, is an extended version of ARToolKit being written in C++. Given a camera frame, the library returns a struct with some data regarding the marker, such as the size in pixels, the coordinates of the center and the corners of the marker, etc. This data is not only used to locate the position of the AR object, but also to detect partially or totally some of the gestures implemented in the application.

5.3.4 Analysis of the sensors data

The Nokia n900 has 3D accelerometers which are used to determine the position of the mobile as well as the movements done by the user.

The data from the sensors is read, filtered to delete part of the noise, discretized and then processed by an algorithm to determine how the mobile was moved.

A very simple but effective filter is applied to the raw data gotten from the accelerometers. The last sample gotten from the sensors while no gesture is detected is subtracted to the current value. The result of this operation is the variation between both samples for each axis.

Once the data is filtered, it is classified in four states:

- Increase: the value of the sensor has increased since the last sample

⁶<https://launchpad.net/artoolkitplus> 15th of November of 2010

- Decrease: the value of the sensor has decreased since the last sample
- Stays in the original position: the value of the sensor has no significant change. While it remains in this state, the initial position is updated with the last sample from the accelerometers.
- Stays in the same position: after the mobile was moved, which means that the previous states were increase or decrease, the value of the sensor has no significant change, but it is still different from the position before the gesture was detected.

The combination of the four states for each axis results in a set of events used in the algorithm to determine which gesture is performed. The Viterbi algorithm [22] is used to do this task. It is a dynamic programming algorithm used to define a path of states according to the observed events. The states are the results of the discretization of the data from the accelerometers. A gesture with the mobile phone is divided as a sequence of states. Some of the states are transitional, that is, they are a part of a possible gesture and the others are final states in which a gesture has been performed.

5.3.5 Combining the gesture recognition methods

The techniques used to recognize the different gestures should work as a unique gesture recognition system to avoid consistency problems.

As it can be seen in the figure 7, the application has two states: locked and unlocked. When the application is in the unlocked state, that is, the gesture-based interaction is disabled, the gesture recognition system updates the current values of the accelerometers as the default position of the mobile.

When the user locks into the AR object, the gesture recognition system begins to analyze the input to detect the gestures. The data from the sensors and the events on the screen is used in this process.

As it will be explained in coming sections, gestures are detected through events or with the data from the accelerometers. The marker information is used to calculate the results of the manipulation or to distinguish between similar gestures. Thus, the application first check if there is any event. Then, it analyzes the values from the accelerometers to detect possible gestures. Depending on the gesture or possible gestures detected, it uses some of the data from the marker to confirm which gesture it is.

5.3.6 Showing the results

The application processes the data from the camera, the sensors and the screen to generate the current state of the AR object.

OpenGL ES 2.0 is used in the mobile as it is supported by the mobile phone. The 3D model used as an AR object is manipulated accordingly and painted over the camera frame.

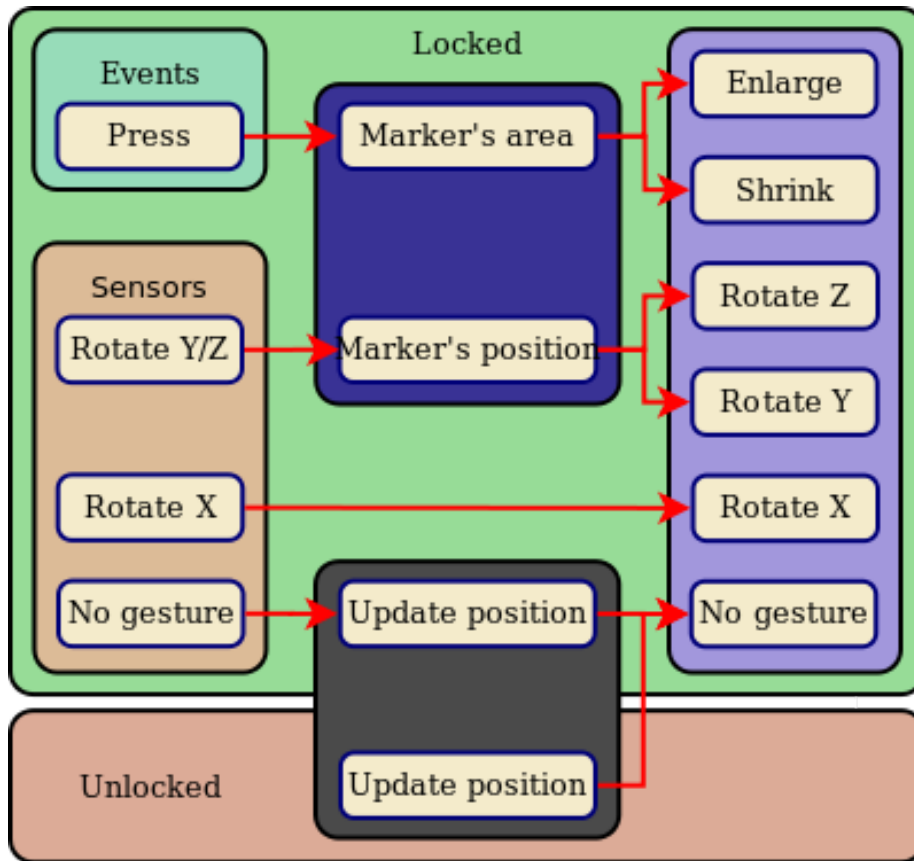


Figure 7: Internal structure of the gesture recognition system

5.4 Implementation of the gestures

As explained above, there are two technics to implement gestures: by using the accelerometers data or by using the data from the marker. Due to each gesture's characteristics, they are implemented using different methods. This makes the implementation easier and the detection of the gestures more precise and robust. In the following sections, the implementation of each gesture is described.

5.4.1 Lock and unlock

Gestures are enabled or disabled by clicking on the screen (see table 3). While the gestures are disabled, the application works as any other AR application where you can only observe the virtual object. By enabling the gestures, users can rotate, enlarge and shrink the AR object.

In order to know if the gesture interaction is enabled or disabled, the marker is painted with two colors. As shown in figure 8, when the marker is black, the

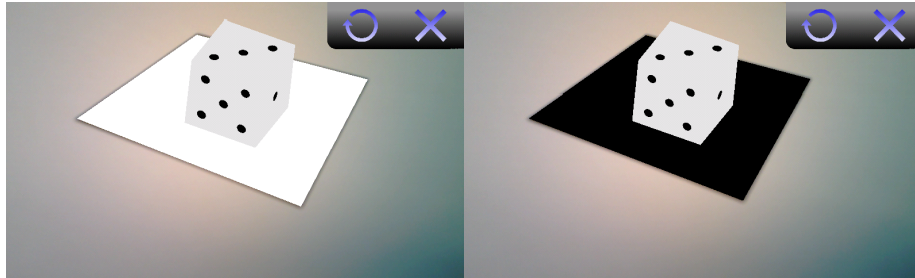


Figure 8: The color of the fiducial marker indicates if the object is locked (left picture) or unlocked (right picture)

gestures are disabled and when the marker is white, the gestures are enabled.

5.4.2 Enlarge and shrink

These two manipulations are performed by pressing on the screen, varying the distance between the mobile phone and the marker and releasing to stop (see table 3). In this case, the tracking data is used to determine how big or small the object is. Thus, the user is forced to keep looking at the object while performing the gesture, giving real time feedback and being possible to perform the gesture from any position as long as the marker is on the camera's viewport.

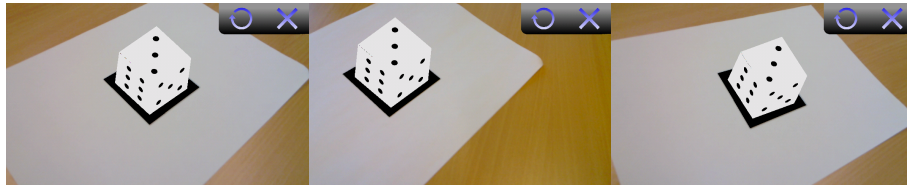


Figure 9: From left to right: no gesture is performed, rotation around the Y axis and rotation around the Z axis

The AR object can be enlarged to the double or shrank to half of it. The current size of the object is calculated by using the area of the marker in the image captured by the camera. The scale factor is the result of dividing the current area of the marker by its previous area but keeping it in the range defined above.

5.4.3 Rotate around the X axis

This rotation is detected by the accelerometers of the mobile. In order to start the rotation, the user 'tilts-tilts back' the mobile phone, as explained in table 5. By performing the same gesture on the opposite direction, it stops the manipulation and resets the position of the mobile. The gesture can be performed clockwise or counter clockwise.

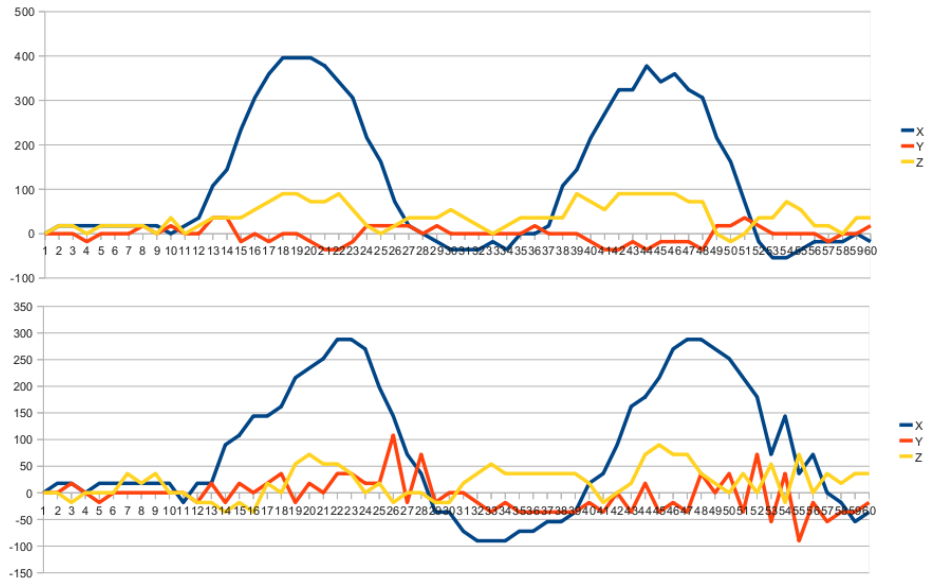


Figure 10: Graphics with the values of the accelerometers while performing the rotation around the Y axis (top picture) and around the Z axis (bottom picture).

5.4.4 Rotate around the Y and the Z axis

As explained in the table 5, these two rotations are invoked by 'tilting-tilting back' the mobile around each axis (Y or Z). Even though these gestures are visibly different, for the accelerometers in the mobile, the gestures are very similar. As shown in the figure 10, both rotations produce the same graphics. The differences are insufficient to distinguish between the two gestures.

The solution is to recognize with the accelerometers these values and use the data from the marker to distinguish between the rotations around the Y and Z axis. The figure 9 shows how the marker is moved on the screen while performing both gestures.

The position of the center of the marker in the camera's viewport allows the distinction of both gestures.

6 Evaluative study

6.1 Purpose

The implementation was based on the results of the user study done to understand how people would like to interact through gestures with an AR application. This was done to ensure that the interaction was intuitive and learnable by the vast majority of people.

Once the implementation was finished, the application was evaluated in a new user study to test if the final result achieved the initial goals. The study was divided into three parts.

The first part aimed to know what people would think when someone interacted with the application. One of the objectives of the application was that gestures should be learnable by observing another person performing them. Thus, people should not only be able to understand the gestures by observing someone else, but also to reproduce them.

The second part evaluated technical aspects of the application. Gestures are done with slight differences between people. In the study, it was tested the robustness and the success of the application in recognizing gestures performed by many people. The interface and the visual feedback shown for the different actions of the user were also evaluated through a questionnaire.

The last one consisted of asking the participants which gestures they would like to perform to invoke the manipulations not implemented. The reason to repeat this part of the first study was to analyze if the methodology and the results collected from that study were accurate. If the results were different, it would mean that the simulation of the application was not enough for users to get an idea of the application and that the results were modified by the procedure.

6.2 Design of the study

A qualitative study was carried out at 'Lava', a youth activity center in Stockholm. Visitors to the center were asked to participate in the study. The study aimed to understand why and what did or did not work the AR demo.

At the beginning, participants were told that the application interacted with an invisible object through gestures. The evaluator performed two manipulations: rotate the AR object around the Z axis and enlarge it. Participants should tell what they thought the evaluator was doing with the mobile phone. Then, they should place a real object where they thought the invisible object was located.

In the next step, users should use the mobile themselves and figure out what the purpose of the application was. They should imitate the gestures done by the evaluator and see the effect.

Having a clear idea of the application, the evaluator did the rest of the gestures. For each one, they should represent with a real object what they

thought it was happening to the AR object. Then, participants had to imitate again the gestures and see the effects.

Finally, the evaluator switched to the alternative rotations, enlarge and shrink, explained in section 5.2.1. Participants were asked to perform the gestures again and see how the AR object was manipulated.

At the end of the study, participants should answer some questions about their experiences with the application and the alternatives manipulations for each gesture. As in the first study, they were asked which gestures they would like to perform to invoke the manipulations not implemented.

The whole study lasted around 20 minutes and each session was recorded with a videocamera. The structure of the study as well as the questionnaire are available in the appendix B.

6.3 Results

Nine people participated in the study, four men and five women aged between 15 and 54. Next subsections presents a deep analysis of the results of the study.

6.3.1 Understanding and learning to use the AR application

In order to verify the application's learnability, the first part of the study explores the application from a performative perspective. Thus, it aimed to know whether third parties would understand how a person was interacting with the application. The results are summarized in table 6.

As explained above, the evaluator first performed the rotation around the Z axis and enlarged the virtual object. Seven out of nine thought he was using the camera or taking a picture. Three of them also suggested as a second option that he was playing some game.

More precisely, for the rotation around the Z axis, seven people said that the evaluator was rotating, turning, switching or navigating through different options.

When the evaluator enlarged the AR object, five participants suggested that he was zooming. Another three pointed that he was taking a picture.

Participants were asked where the invisible object was located. All of them placed the real object around the fiducial marker. Only one put it on the marker. Some of them were looking carefully at the camera position to determine where the object should be.

Once they had seen the application, they should think about the manipulations invoked by the rest of the gestures. Eight out of nine guessed correctly that the object was being shrank while performing its gesture. Six participants knew that the object was being rotated around the Y axis, while eight guessed it for the X axis.

#	Manipulation	Impression	No.
1.1	General impression	Taking a picture	7
1.2		Playing a game	3
2.1	Tilt around the Z axis	Rotating, turning, switching, tilting	7
3.1	Enlarge	Zooming	5
3.2		Taking a picture	3
4.1	Tilt around the X axis	Rotate the AR object around the X axis	8
4.2		Rotate the AR object in another way	1
5.1	Tilt around the Y axis	Rotate the AR object around the Y axis	6
5.2		Rotate the AR object in another way	2
6.1	Shrink	Shrink the AR object	8

Table 6: Summary of the third person’s impressions while looking someone using the application

6.3.2 Usage experience

The usability, robustness and learnability of the application was tested when users performed the gestures themselves. Enlarge and shrink got the best results, with only one person having problems to use them.

The rotation around the X axis was performed also by eight people, but having some difficulties using it. They had to repeat the gestures a few times before their gestures were precise enough to be recognized by the application. All of them surpassed the difficulties and managed to rotate the AR object.

The rotation around the Y and the Z axis got the lowest success ratio. Seven and four out of nine respectively managed to perform gestures.

Some participants were also confused with the locking and unlocking system. The visual information added to know if it was locked or unlocked, was noticed by four out of the nine people. This provoked some difficulties using the applications.

The questionnaire revealed that 8 out of 9 people considered the rotations intuitive and 7 liked the gestures to invoke the rotations. All the participants agreed that the manipulation and the gestures to enlarge and shrink were intuitive and easy to use.

6.3.3 Gestures for non-implemented manipulations

For the manipulations not implemented, participants in the evaluative study were asked, as in the first study, which gestures they would like to perform to invoke them. More precisely, they were asked about the pick up, place, drop off, move to another position and move towards a direction.

Table 7 describes the gestures with the graphical language defined in table 2.

Two participants picked up the virtual object by moving the mobile phone farther from the marker (1.2 in table 7), another two by moving closer and











#	Effect	Textual description	Graphics	No.
1.1	Pick up	Screen-based interaction	-	3
1.2		Move farther from the marker		2
1.3		Move closer and farther from the marker		2
2.1	Drop off	Throw gesture	-	2
2.2		Move closer fast		2
2.3		Screen-based interaction	-	2
2.4		Shake		1
3.1	Place	Move slightly closer to the marker		4
3.2		A slower drop off movement	-	2
3.3		Screen-based interaction	-	2
4.1	Move to another position	Press, mirror mobile's movements, release		3
4.2		Click, mirror the mobile's movements, click		2
4.3		Tilt the mobile phone		1
5.1	Move towards a direction	Move the mobile towards the direction		3
5.2		Tilt the mobile to indicate the direction		2
5.3		Screen-based interaction	-	3

Table 7: Results from the study

then farther from the marker (1.3 in table 7). The rest of the users proposed a screen-based interaction (1.1 in table 7).

Four gestures were proposed for dropping the AR object off. Two people suggest to do a 'throwing gesture' (2.1 in table 7). Two participants moved the mobile phone closer to the marker (2.2 in table 7) to invoke the manipulation and two more proposed a screen-based interaction (2.3 in table 7). Finally, one person suggested to shake the mobile phone (2.4 in table 7).

Participants placed the virtual object on the marker by doing three different gestures. Four of them proposed to move slightly closer to the marker (3.1 in table 7), another two to perform a slow drop off movement (3.2 in table 7) and two more to use the screen (3.3 in table 7) to invoke these actions.

More than a 50% of the participants used the following pattern to provoke the manipulation: use an event to start, mirror the mobile's movement, and use an event to stop. Three of them pressed the screen to start the interaction and released to stop (4.1 in table 7). Another two used a single click to start and stop the manipulation (4.2 in table 7), and one person suggested to tilt the mobile phone (4.3 in table 7).

For moving the AR object towards a direction, three users suggested to move the mobile in the direction they want to move the virtual object (5.1 in table 7). Two participants, tilted the mobile phone to indicate the direction (5.2 in table 7) and three more used a screen-based interaction (5.3 in table 7).

6.4 Analysis

6.4.1 Performative gestures

The results from the evaluative study shows that participants figured out how the evaluator was interacting with the application. Despite the non-experience with AR, they interpreted the gestures with their own experiences. People familiar with modern mobiles suggested that he was taking a picture or zooming. Three participants, aged between 15 and 21 years old, suggested that he was playing some game. On the other hand, a user aged 54 suggested that the evaluator was tuning the radio.

Even before seeing the application, thinking that the evaluator was interacting with an 'invisible object', they related the gestures to previous experiences but having a similar meaning as in the application.

After they used the application and having some experience with augmented reality, a high percentage of the participants could guess which kind of manipulation was done to the AR object.

6.4.2 Robustness and adaptability

Enlarging, shrinking and rotating around the X axis got a high success ratio in terms of usability by the participants in the study. They were able to perform the gestures with non or a few instructions. However, some users had problems performing the rotations around the Y and Z axis.

As explained in section 5.4.4 these two rotations are detected with the accelerometers and distinguished from each other with the marker data. The study showed that the detection of these gestures were not robust enough. In some occasions while participants were doing the gestures, the application was not able to distinguish correctly between both gestures.

6.4.3 Manipulations' preference for each gesture

In section 5.2.1 was explained that rotations would be performed in two different ways: by steps or continuous rotation. The reason was to know which manipulation was better accepted by the participants in the study. The results show that there was not a clear preference between the rotation in small steps or the continuous rotation. Four of the participants said that both are valid depending on the application for which are used. Thus, depending on the application, it should be used one or the other.

The evaluative study aimed to verify the results from the user study where participants were divided on how to enlarge and shrink, as shown in table 3. The enlarging and shrinking could not be simulated in the user study which suggested that the division of opinions could be induced by not seeing the resulting manipulation. However, the same division of opinions appeared in the second study as well. Five participants preferred to shrink the object when moving the mobile closer to the marker. They argued that it was easier to see the AR object. On the other hand, the other four preferred the opposite because it is natural and intuitive. In the real world, an object becomes bigger when you get closer to it. Thus, both options are valid for enlarging the object to the double and shrank it to reduce half of it.

6.4.4 Usability issues

The study revealed some usability problems which should be considered for future applications. Only two of the participants were able to reproduce the gestures without explaining how to use the application. In most of the cases, gestures had to be performed a few times before they could imitate them properly. They were doing the gestures similarly but forgetting, for instance, to press on the screen for changing the size of the AR object or doing a slow movement rather than with a fast 'flick' movement to rotate the object.

Light conditions provoked some tracking problems. The rotations around the Y and Z axis were the most affected, as they were using the marker data to distinguish between both gestures.

The lack of experience of the users in AR applications and the tracking problems generated an unexpected problem. When users were doing the gestures and the tracking system failed, the AR object disappeared or blinked. Some participants thought that was the effect invoked by the gesture they did with the mobile phone. Although the movement of the phone provoked the tracking problem, it was not a desired effect. Thus, the lack of experience in AR applications led to misunderstandings. The application should, then, indicate that

there is a technical error.

6.4.5 Methodology used for designing the gesture repertoire

The results from the evaluative study shows that the method used to collect data to design the gesture repertoire was appropriate. As explained in the previous section, the gestures are intuitive and participants were able to map them to their own experiences with a similar meaning as it had in the AR application.

Thus, the simulation of the application done in the first study, not only was understood by the participants, but also allowed them to give accurate feedback on which gestures could fulfill the requirements of the application.

The gestures proposed in both studies for the not implemented manipulations reinforce the chosen methodology to define the repertoire of gestures. Although there are slight differences, many of the suggestions of the users appeared in both studies. It should be noticed that the different background of the participants in each study. The fact that in the second study most of gestures got a screen based suggestion explains this statement. Participants in the first study were familiar with new technologies and many of them knew what augmented reality was. However, participants in the evaluative study were not familiar with AR and not necessarily in brand new technologies.

7 Conclusions

7.1 Summary

The work presented in this thesis can be summarized in three main blocks:

1. A user study was done in order to build the project upon the user's experience. With this study, we got closer to the users needs and, thus, we could reach our goal to have a learnable and intuitive gesture-based user interface. The data collected in the study was analyzed in order to build a repertoire of gestures. The selection of the gestures for each manipulation was done according to three criteria: a gesture should be recognizable with the hardware resources available in the mobile phone, it should be consistent with the rest of the gesture repertoire and in case of having two or more possible gestures, the most frequently suggested was selected.
2. The design of the demo application took into account the results from the user study. Due to time limitations, the implementation was narrowed down to the following actions: lock and unlock system, enlarge, shrink and the rotations around the X, Y and Z axis.
3. An evaluative study was conducted in order to verify the results of the investigation implemented in the AR application. Its robustness, learnability and usability were tested. The results showed that the implementation was not robust enough for some gestures. Some unexpected technical problems appeared during the test, which led to misunderstandings because of the lack of experience of the user in AR applications. The results also pointed out that the chosen methodology to design the set of gestures was appropriate and gave accurate results.

7.2 Discussion and conclusion

The goal of this thesis was to explore the possibilities of a gesture-based interaction within an AR application and to define a standard repertoire of gestures which could be used in future mobile AR applications. The iterative research methodology guided this thesis to achieve its goals. From the beginning, the user's point of view was considered in any decision related to the development of the application. Technical feasibility was also considered as an important criterion. A well-implemented gesture is easier to recognise and, thus, the interaction with the application is easier. The combination of those criteria helped to develop a natural and learnable gesture-based interaction with a high acceptance ratio by the users.

Due to time limitations, only two iterations could be performed: the first user study followed by the development of the AR application, and the evaluative study to test the demo. As it has been explained, the tracking problems and the unexperience of the participants in the field of AR caused some misunderstandings. It would have been better to add an iteration and divide the evaluative study in two.

After the development of the AR application, a technical study could have been carried out to test the usability of the demo, the robustness of the recognition system with users who had no experience with augmented reality. This study would have allowed to correct many technical problems that appeared in the evaluative study. Once all the issues were corrected, the evaluative study would have been carried out.

Although it would have been desirable to evaluate the AR demo in two iterations instead of one, the results of the thesis are satisfactory since they prove that gestures are an excellent interaction method for AR applications. The combination of both technologies provides a realistic and natural experience to the user to interact with digital information.

From a performance point of view, the use of a gesture-based interaction within an AR application is not only easy to learn but also a natural way of interaction which can be understood by third parties. As shown by the results, the participants involved in the study were able to guess which kind of manipulation was done by the gestures.

This thesis was focused on a very particular case of augmented reality. The application was limited to use only one fiducial marker. We believe that the gesture-based interaction presented in this master thesis is applicable to other AR scenarios. For instance, an AR application which uses more than one AR object could use the same interaction technique as long as there is a way to select which object the user is interacting with. Markerless tracking applications are different from a technical perspective, however, there is no sign which suggests that these gestures cannot be used in markerless tracking applications as interaction method.

The defined set of manipulations was focused on simple interactions with a virtual object. The manipulations modified in different ways the state of the virtual object and allowed the user to observe the AR object in more or less detail as well as from different positions and perspectives.

7.3 Future work

The results of this thesis point the feasibility of combining augmented reality and gesture-based interaction. However, they also show that more research is required in this area.

The robustness of the gestures should be improved and the rest of the gestures of the repertoire should be implemented and tested. A technical study should be carried out to identify the difficulties that may be encountered and improve the usability of the application.

A deeper research on the learnability of the application should be done. It would be interesting, for instance, to give the application to a group of people and to observe if they can learn themselves how to use the application and the amount of instructions they give to each other.

It would be also interesting to analyze other kinds of manipulations or in other scenarios. For instance, how to select an object to be manipulated in an application that uses many markers simultaneously.

The manipulations and its gestures were defined to interact with a 3D model. It would be interesting to implement this repertoire in an application to interact with 2D information and see if they work or they are redefined in a specific way.

Gestures have a wide range of possibilities to become a natural interaction method for augmented reality applications. The combination of both technologies can offer the user a new experience while interacting with digital systems.

References

- [1] Azuma, R. 1997. A Survey of Augmented Reality
- [2] Milgram, P., Kishino, F. 1994. A taxonomy of mixed reality visual displays
- [3] Rohs, M. 2005, Real-world interaction with camera phones.
- [4] Rohs, M. and Gfeller, B. 2004. Using camera-equipped mobile phones for interacting with real-world objects.
- [5] Sturman, D., Zeltzer, D. 1994. A survey of glove-based input.
- [6] Davis, J., Shah, M. 1994. Recognizing hand gestures.
- [7] Sánchez-Nielsen, E., Antón-Canalís, L., Hernández-Tejera, M. 2003. Hand gesture recognition for human-machine interaction.
- [8] Wang, J., Zhai, S., Canny, J., 2006. Camera phone based motion sensing: interaction techniques, applications and performance study.
- [9] Bahar, B., Burcu Barla, I., Boymul, Ö., Dicle, Ç., Erol, B., Saraçlar, M., Metin Sezgin, T., Železný, M., 2007. Mobile-phone based gesture recognition.
- [10] Niezen, G., Hancke, G., 2008. Gesture recognition as ubiquitous input for mobile phones.
- [11] Prekopcsák, Z., 2008. Accelerometer based real-time gesture recognition.
- [12] Chen, L-H., Yu, C-J., Hsu, S-C., 2008. A remote chinese chess game using mobile phone augmented reality.
- [13] Watts, C., Sharlin, E., 2008. Photogeist: An augmented reality photography game.
- [14] Wetzel, R., Waern, A., Jonsson, S., Lindt, I., Ljungstrand, P., Åkesson, K-P., 2009. Boxed pervasive games: An experience with user-created pervasive games.
- [15] Harvainen, T., Korkalo, O., Woodward, C., 2009. Camera-based interactions for Augmented reality.
- [16] Xu, Y., Gandy, M., Deen, S., Schrank, B., Spreen, K., Gorbsky, M., White, T., Barba, E., Radu, J., Bolter, J., Macintyre, B., 2008. Bragfish: Exploring physical and social interaction in co-located handheld augmented reality games.
- [17] Rohs, M. and Zweifel, P. 2005. A conceptual framework for camera phone-based interaction techniques.
- [18] Bury, K. F. 1984. The iterative development of usable computer interfaces.

- [19] Kruchten, P. 2000. From the waterfall to iterative development - A challenging transition for project managers.
- [20] Reeves, S., Benford, S., O'Malley, C., and Frased, M., 2005. Designing the spectator experience.
- [21] Norman, D., 2002. The design of everyday things - Chapter 7.
- [22] Forney, G.D., 1973. The viterbi algorithm.
- [23] Foehrenbach, S., König, W., Gerken, J., Reiterer, H., 2008. Natural Interaction with Hand Gestures and Tactile Feedback for large, high-res Display.

A User study

Before starting, for this study I will record you on video to analyze your gestures and comments afterwards. Do you agree?

The aim of this study is to explore gestures through a mobile to interact with a virtual object. Augmented reality is a technology that allows drawing virtual objects over the real world, using a camera and a screen.

As we still do not have the system implemented, for this test I will be move an object as if it was virtual. You will make a gesture with the mobile that make sense to you to provoke the movement I am doing with the object.

Keep in mind that you are interacting with a virtual object, so you need to **see it through the camera**. Out of the camera you would not see it.

You are free to think of any movement to interact with the virtual object in a specific way. You are allowed to **move the mobile and touch the screen with one finger** as if it was a button. You can not select, scratch or do anything else with the screen. **Just press and release**.

Once I ask you to think of a gesture, you will be asked to **think aloud** and to **try different movements**. Once you choose one movement, you will be asked to **perform it three times** to make sure you feel comfortable doing the movement several times continuously.

Any question?

1. To start interacting with the object, first you need to lock it. Until you don't attach it, no movement will have effect on the virtual object. Which movement would you do to:
 - Lock the object
 - Unlock the object
2. Now I want you to think a gesture to shake the object. Which gesture would you do?
3. Now I want you to think a gesture to change the size of the object.
 - Enlarge
 - Shrink
4. Which gesture would you do to:
 - Pick up an object
 - Drop off an object
 - Place an object
5. Think of a gesture to:
 - Move the object from its current position to another position

- Move the object towards a specific direction

6. Now I want you to think of a gesture to:

- Rotate around the X axis
- Rotate around the Y axis
- Rotate around the Z axis
- Rotate around a specific axis
- Rotate a certain amount of degrees around a specific axis

A.1 Questionnaire

- Age:
 - Gender
1. Can you think of any other interaction with the virtual object? If so, which movement would you do to perform that action?
 2. Which actions do you think have a more natural or obvious gesture interaction?
 3. Which action do you think have a less natural gesture interaction?
 4. Which kind of rotation do you think is more useful, easy to use or intuitive?

One problem of building a gesture-based interaction systems is the difficulty on defining gestures for different interactions without overlapping them. So, for a big set of gestures, the implementation is much more difficult and the usability decreases as the user must do a more precise movement so that the system does not misunderstood user's intentions.

In order to make implementation feasible and interaction easier, we could have two different modes. In each mode, you would have a set of gestures available. Thus, there are less gestures which makes it easier to recognize and program.

5. I would like you to divide the actions performed before in two different sets according to a specific criteria.

For instance, you could have the navigation mode or the interaction mode. In the navigation mode you would perform actions like rotate, move, shrink or enlarge while in the interaction mode you would pick up and drop, shake.

6. How would you change between the modes? By which gesture?
7. Would you like to have some visual information to know in which mode you are?
8. Would you like to have visual information regarding the possible actions available?
9. Do you think you should be able to lock/unlock the object from any mode?

B Evaluative study

Before starting, for this study I will record you on video to analyze your comments how you interacted with the application afterwards. Do you agree?

I am developing an application that uses the gestures you do with the phone to interact with an invisible object in the real world. Now I'll do two different gestures provoking two interactions with this invisible object.

1. Imagine you saw me on the street doing this. What would suggest to you?
2. Put your object where you think the invisible object is located
3. What do you think it happens to the object when I...? Represent it with your real object.
 - Rotate Z clockwise
 - Enlarge
4. Try it out
Now, I will perform some more movements which provoke different interactions to the invisible object. Observe the movements.
5. What do you think it happens to the object when I...? Represent it with your real object.
 - Rotate Z clockwise
 - Rotate Z counter clockwise
 - Enlarge
 - Shrink
 - Rotate Y clockwise
 - Rotate Y counter clockwise
 - Rotate X clockwise
 - Rotate X counter clockwise
6. Try it out
7. [Change to rotation by steps and swap the enlarge/shrink movements]
8. I change some properties of the application. Try it out again and find out the differences
9. According to what you have seen, how would you:
 - Pick up
 - Drop off
 - Place
 - Move to another position
 - Move towards a specific direction

B.1 Questionnaire

- Age:
 - Gender
1. Do you think the way the 3D object rotates is intuitive?
 2. Would you like the rotations to be implemented in another way?
 3. Which of the two rotations would you prefer and why?
 4. Is it intuitive to scale the object?
 5. Would you like the scaling to be implemented in another way?
 6. Which of the two scaling you prefer and why?
 7. Were you able to recognize visually if the gesture interaction was enabled?