APPLYING TANGIBLE AUGMENTED REALITY IN USABILITY EVALUATION

A Thesis
Presented to
The Academic Faculty

by

Xiaotian Zhang

In Partial Fulfillment
of the Requirements for the Degree
Masters in the
School of Industrial Design

Georgia Institute of Technology
May 2016

APPLYING TANGIBLE AUGMENTED REALITY IN USABILITY EVALUATION

Approved by:

Dr. Young Mi Choi, Advisor
School of Industrial Design
Georgia Institute of Technology

Tim Purdy
School of Industrial Design
Georgia Institute of Technology

Wendell Wilson
School of Industrial Design
Georgia Institute of Technology

Date Approved: 1/14/2016

ACKNOWLEDGEMENTS

First and Foremost, I would like to express my sincere gratitude to my advisor Dr. Young Mi Choi for the continuous support of my study and research, for her patience, motivation, enthusiasm, and immense knowledge.

Also, the completion of this research could not have been possible without the participation and assistance of so many people whose names cannot all be enumerated. Their contributions are sincerely appreciated and gratefully acknowledged. However, I would like to express my deep gratitude to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR	Augmented Reality
AF	Augmented Foam
TAR	Tangible Augmented Reality
3D	Three Dimensional
SDK	Software Development Kit
UEI	User Experience Index
TUE	Total User Experience
NPD	New Product Development
BAH	Booz, Allen and Hamilton

SUMMARY

Feedback from users is an invaluable part of the product design process. Prototypes of varying levels of detail are frequently used to solicit feedback for attributes related to the physical and user experience aspects of a product. Using Tangible Augmented Reality Technologies to achieve natural interactions have significance in product design. TAR is one of Augmented Reality related technologies which enable the superposition of a physical model in a real environment and a digital model onscreen to permit evaluation of a product.

This study is to investigate whether usability input from AR/TAR technologies' representations of a product with physical interface elements is similar to input based on a physical prototype. If AR/TAR technology inputs are found to be similar to the real product, it can be an evidence to indicate that AR/TAR can be a useful tool for collecting highly accurate inputs on a product concept focusing on physical controls. User testing was conducted with 20 participants.

And conclusion shows the two AR methods are statistically significantly different from real products, but Tangible AR does not have significant difference from real products. The results of the data analysis strongly support that Tangible AR is an effective method of prototyping for user study.

CHAPTER 1

INTRODUCTION

Augmented Reality refers to a view of the real or physical world in which certain elements of the environment are computer generated. These virtual elements could be a modification of a current element in the real world or could be an entirely new element. AR is one way that some of these intangibility problems can be solved in many fields, such as healthcare, maintenance, urban planning, etc. Product design is also one of these related fields, but involves more natural interactions. Using AR related technologies to achieve natural interactions has significance in product design.

Tangible Augmented Reality is one of the AR related technologies that enables the superposition of a physical model on the real environment and a digital model on the screen.

Problem Statement

Usability evaluation has always been the most important part of the design process. A good method for performing a usability test would be to produce accurate and significant data, which could help improve the design. As technology has developed, varying kinds of prototypes are frequently used to test the concept and gain feedback with respect to the physical and user experience aspects of a product. Consequently, accurate feedback from the early stages of the design process plays the most important role, as it will guide the direction of the project. Designers cannot make final decisions until the end of the design process, which means that the highly detailed prototypes are not available this late point. Accordingly high-fidelity prototypes are the key point to solve this problem.

Goals of the Study

The goal of this study is to investigate whether usability input from AR/TAR technologies representative of products with physical control elements is similar to the usability input based on a real product. If AR/TAR technology inputs are found to be similar to the real product, it can be evidence to indicate that AR/TAR is a useful tool for collecting highly accurate inputs for product concepts focusing on physical controls. This method might allow collection of inputs earlier in the product design process or allow for more thorough testing of potential concepts. In the study, usability input will be collected from a product with physical components along with three different representations using AR/TAR technologies.

CHAPTER 2

BACKGROUND

Introduction

Feedback from users is an invaluable part of the product design process. Usability evaluation has always been one of the most important parts of a design process. Proper usability testing will produce accurate and significant data, which could help improve the overall design. As technology has developed, a variety of different prototypes are frequently used to test an idea and get feedback with respect to physical and user experience aspects of a product. Significantly accurate feedback from early stages of design process plays the most important role, as it will guide the direction of the project. Designers cannot make the final decisions until the end of the design process, which means the highly detailed prototypes are not available this late point.

In this situation, new solutions are urgently needed. New technologies have been applied and have greatly improved the design process. New technologies such as Augmented Reality and correlate technologies.

Computer Aided Design applications are popular tools, which designers use to graphically represent and simulate ideas and concepts. There are numerous applications and technologies available such as Solidworks for 3D modeling or Keyshot for rendering and animations. However, all of these methods have a common problem: The rendered models by these methods are intangible: users cannot interact with the prototype or conceptual model through the software. Augmented Reality is one solution by which some of these intangibility problems may be resolved.

Augmented Reality in Different Field

Augmented reality---first mentioned by Frank Baum at 1901--- 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, or graphics. In an AR system, a cue in an environment (such as picture, photograph, QR code, etc.) is replaced by a new element. This could be a modified element based on the original one, a fictitious element, a video, etc. A current example of this is used by IKEA to provide a virtual preview of furniture from their catalog within a room[1].This ability to dynamically replace certain visual elements with new/different one can be a potentially very helpful aid for product designers.

AR has now been widely used in different field, for example, Héctor Martínez, Seppo Laukkanen and Jouni Mattila presented a new hybrid approach that enables the creation of augmented reality maintenance applications for large and hazardous scientific facilities[2]. J. Zhu & S. K. Ong & A. Y. C. Nee proposed an authorable context-aware AR System (ACARS) to assist maintenance technicians to interact with the AR contents actively[3].

In these fields, the characteristics of AR, which combines real, environmental, and fictious objects together, offers advantages in helping users better understand the use environament and achieve their goals.

Implementations of Tangible Augmented Reality

There are already research studies aimed at extending AR into more tangible forms. Mark Billinghurst, Hirokazu Kato, and Seiko Myojin described several interaction methods that can be used to provide a better user experience, including tangible user interaction, multimodal input and mobile interaction[4]. An example of this is Augmented Foam (AF) introduced by Woohun Lee and Jun Park in 2005[5]. It is based on original AR technology. Regular AR is based on a graphic input, such as a defined picture. When camera sees the defined picture, a digital model replaces it so that the new model (and not the actual picture captured by the camera) is displayed. AF works similarly to AR. A picture cue is used to place a new digital model onto a surface when viewed through a camera. In this case, the picture is placed on a 3D printed model which is printed to resemble the form of the digital replacement. The digital model and the 3D printed model are superposed on each other, and since the physical portion of the model is similar to the digital replacement viewed through the camera, this enables a greater level of physical interaction and variation. (Fig. 1)

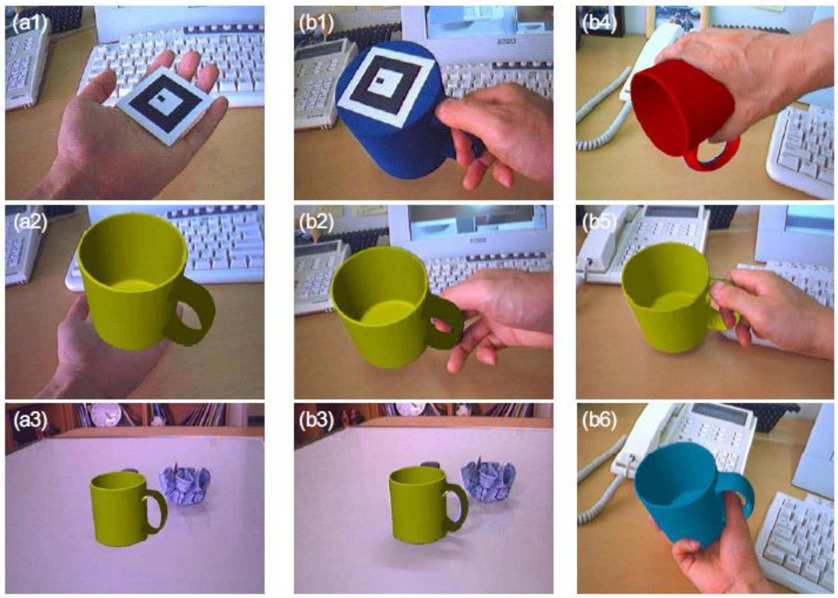


Figure 1. General-Purpose AR (A1~A3) And Augmented Foam (B1~B6): (A1) Plane Marker (A2) Virtual Overlay On A Plane Marker (A3) Virtual Overlay On A Table, (B1) Augmented Foam Without Virtual Overlay, (B2) Augmented Foam With Virtual Overlay (Visibility Problem) [5]

Woohun Lee and Jun Park also implemented AF combined with electrical components, which achieved interactions via electronic signals[5].

There was another research study, which focused on molecular biology conducted by Alexandre Gillet, Michel Sanner and their colleagues[6]. They used a similar technology to create an application that demonstrates the use of auto-fabricated tangible models and augmented reality for research and teaching in molecular biology, and for enhancing the scientific environment for collaboration and exploration. The user manipulates a model, and the model is tracked by a video camera and displayed on the computer screen. A virtual representation (e.g., another 3D rendering of the same molecule, textual labels, or a 3D animation) is superimposed over the video display, and spatially registered with the model as the user explores the structure.

Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev's research focused on Tangible Augmented Reality Interface, which is now popular among many AR-related fields[7]. Their research defined what Tangible AR interfaces are, presented some design guidelines, and prototyped interfaces based on these guidelines. Their experiences with these interfaces showed that the Tangible AR metaphor supports a seamless interaction between the real and virtual worlds, and provides a range of natural interactions that are difficult to find in other AR interface

Significance in Applying Tangible AR in Usability Evaluation

Tangible AR approach can be a suitable method for usability evaluation, the advantages can be huge if it turns out to be feasible. Traditional methods of usability testing have many limitations, such as the final detailed product will not be available until the end of the design process as we have previously mentioned, and performing usability test in the traditional ways, which ask the end-users to use the product and then give feedback can be costly, inconvenient and unsatisfying in many ways. The idea of using Tangible AR to perform usability evaluation meets this design need as it can be useful in simulating the detailed product such as physical interactions even though it is in still in the early conceptual stages. Tangible AR can be easily handled by the user and thus may not show a big difference with the physical product and most importantly it can save a lot of time and money as well as human resources by building computer-generated augmented reality prototypes rather than building actual physical models. However, there may be some factors such as a reduction in the physical experience that could influence the outcome of usability testing.

In this paper, a new method will be introduced, which combined rapid prototyping and AR technologies to detect the button's usage. There was no electrical parts included.

The goal of this study will be to investigate whether usability input from three different representations (AR, AF, and TAR) of a product with physical interface elements is similar to input based on the physical product itself. If one or more of those inputs is found to be similar, it can indicate that it/they can be a useful tool for collecting highly accurate input on a product concept. This might allow collection of input earlier in the product design process or more thorough testing of potential concepts.

CHAPTER 3

METHODOLOGY

Hypothesis

This research compares traditional usability results and results from three representations of a real product, which applies AR and Tangible AR technologies to find out whether there are differences between each of them. If one of those representations is found to have significant differences from traditional usability results, then the method applied to that representation wold be deemed not an effective tool to be used in usability testing during early design process. If there is no significant difference in one of them, then it may be possible to use the method applied to that representation as a lower cost substitute.

To test the hypothesis, this study asked users to finish four experiments in which they interacted with four different objects. One of them was a real product, the other three objects were representations of this real product, which were built using different technologies: pure AR technology, AR with 3D printed model (named as "Augmented Foam"), and Tangible AR technology. (Short for AR, AF and TAR) After the users performed tasks with each object, they provided feedback through a Use Questionnaire, then the results compared for usability between the traditional method and the AR methods.

Product for Usability Testing

The Sunbeam heater SFH5264MW (Fig. 2) was used for usability testing. The reason why this product is chosen was that it is not overly complex and it has clear physical controls, which can highlight the key point of the experiment. Another reason for choosing this product was that its size and weight permit it to be easily carried to different test sites. A product with a knob, which has a ridge-shape design will not be chosen because the marker should not be attached on the front side.

Next, this product will be precisely measured and simulated for the prototypes.

Figure 2. Sunbeam Heater SFH5264MW

Modeling the Product for AR and AF Groups

For AR Group

3D modeling was the first step building all the prototypes. The product was precisely measured and then built in 3D modeling software, which was Solidworks in this case. (Fig. 3)

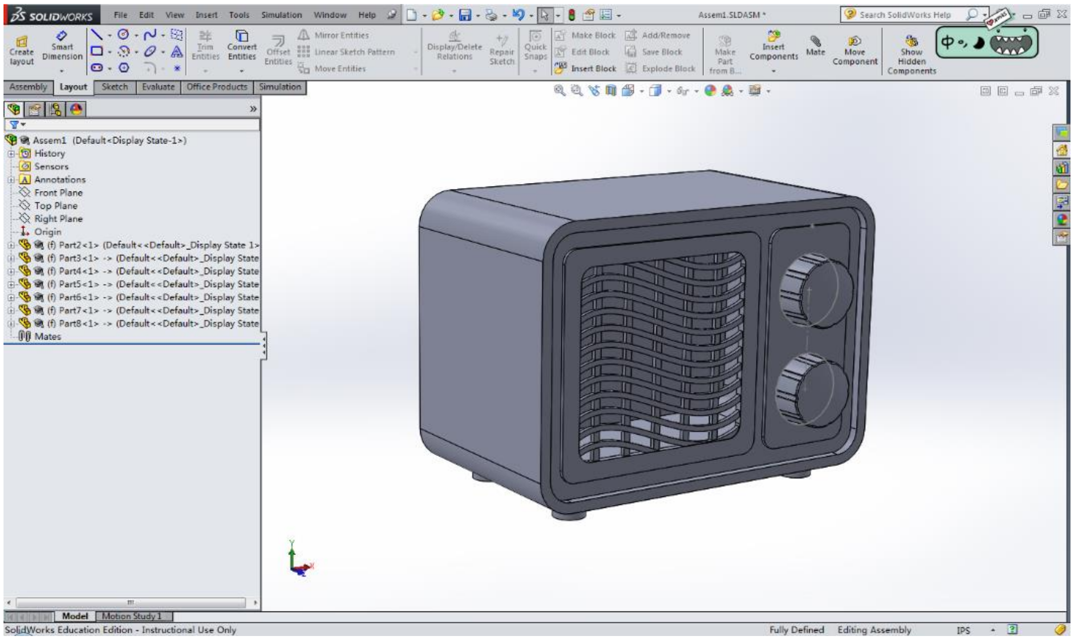


Figure 3. Building 3D Model in Solidworks

After modeling the product in 3D, another web prototype that just shows the front of this fan heater also was finished in the Javascript programming language.

The Unity3D and Vuforia SDK software was used to place the virtual model, which combined 3D digital modeling and a web prototype (Fig. 4) together in the real environment.

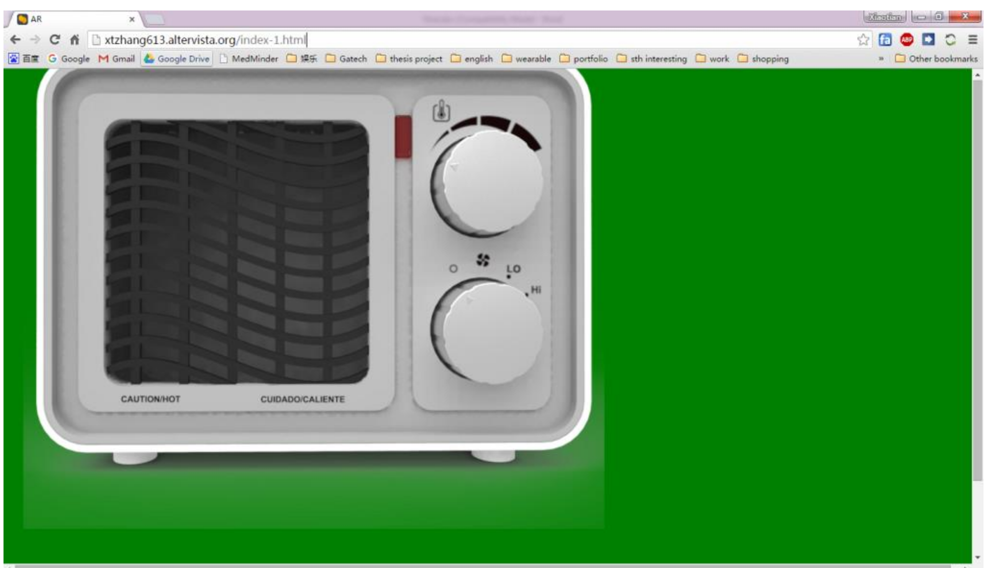


Figure 4. Web Prototype Showing the Front of the Product

Marker detection technology was used to detect the marker (Fig. 5) in the real environment and the 3D model that was created was superinposed on top of the marker when users viewed it through a tablet.

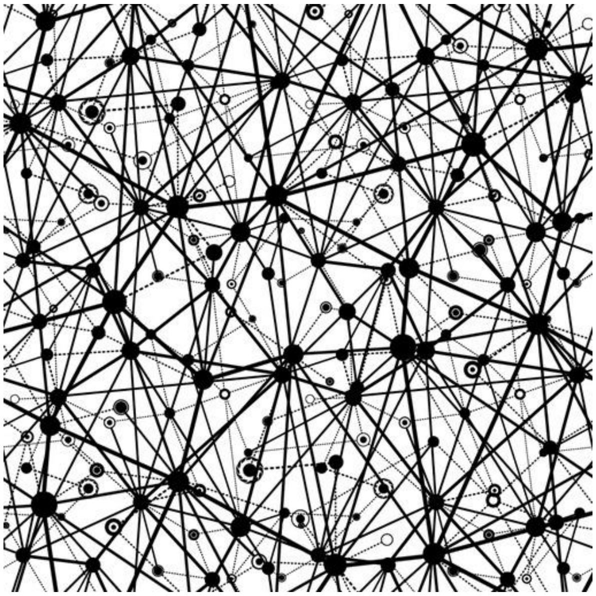


Figure 5. Marker Card Used in AR Group

While the users viewed the virtual model through a tablet, they had the ability to interact with the object through an augmented reality app on the tablet. Figure 6 shows how it looks like when everything is set up for AR group.

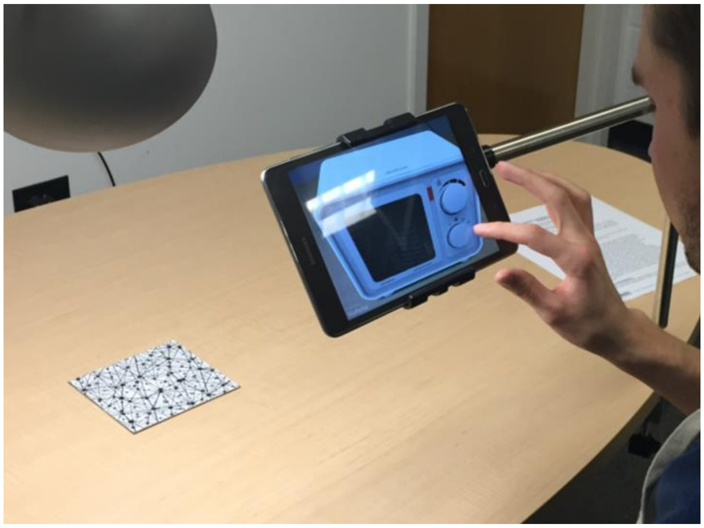


Figure 6. Installed Experiment for AR Group

For AF Group

Compared to the AR group, all the materials and software used in AR group are also being used in the AF group. The difference was the 3D printed model and marker position. The 3D printed model, which was printed using the same digital model as the 3D virtual model was used to let the user have more cognition by touching. The marker was placed on top of the 3D model both in the virtual world and the real world.

Figure 7. 3D Printed Model

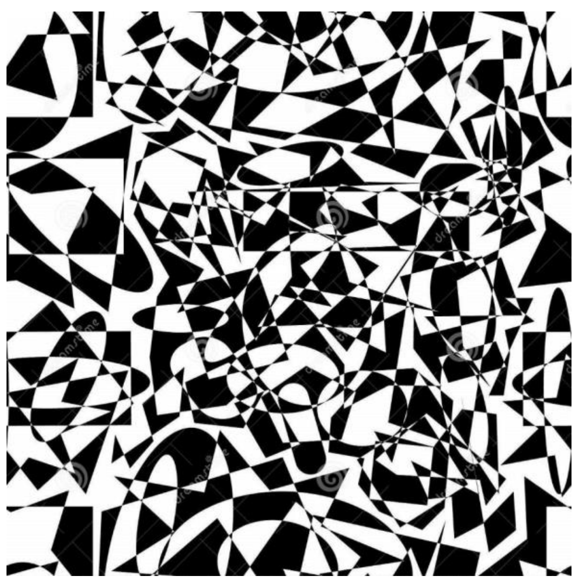


Figure 8. Marker Card for AF Group

Figure 9 shows how it looks like when everything is set up for AF group.

Figure 9. Installed Experiment for AF Group

Prototyping for TAR Group

Unity3D and Vuforia SDK also used to create the tangible interactive model for the TAR group. The first step was linking the three models to each marker, and testing them on the screen. The second step was to define different angle ranges for next step use. The last step was to find proper sounds and then add those sounds to different angle ranges to simulate the real knob experience. For example, a "click" sound appears when user turn the knob from a mode to another mode. After all three steps, the general testing was implemented to determine the performance.

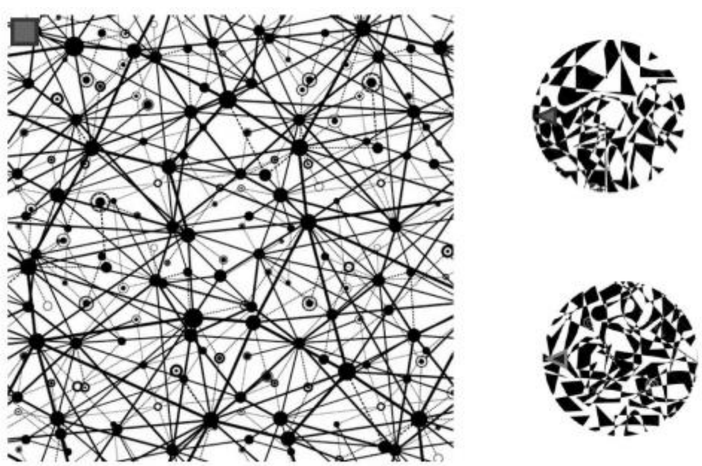


Figure 10. Three Marker Cards for TAR Group

Figure 11 shows how it looks like when everything is set up for TAR group.

Figure 11. Installed Experiment for TAR Group

Design of Experiment

Introduction

Usability testing was completed by 20 participants. Each of them was asked to interact with all the four design representations. Each user did the testing individually. The real product group named as NonAR, the Augmented Reality group as AR, the group applying Augmented Reality with 3D printed model as AF and the Tangible AR group as TAR for short. IRB approval was obtained from the Georgia Institute of Technology to conduct the study and the approved consent form can be found in the appendix.

Experiment Order

Since there are the four groups of experiments, each of which is dealing with different representations, four different experimental orders were designed, to which “A” through “D” were assigned as indicated below.

Table 1. Experiment Order

Order	A	B	C	D
1	Real	AR	AF	TAR
2	AR	AF	TAR	Real
3	AF	TAR	Real	AR
4	TAR	Real	AR	AF

Task Order

Since this Sunbeam heater has two knobs, and four modes for each knob, that means that there are 8 tasks in total as shown below:

1. Turn the Mode Control to “High-heat”
2. Turn the Mode Control to “Low-heat”
3. Turn the Mode Control to “Fan-only”
4. Turn the Mode Control to “Off”
5. Turn the Thermostat Control to Level 1
6. Turn the Thermostat Control to Level 2
7. Turn the Thermostat Control to Level 3
8. Turn the Thermostat Control to Level 4

Since there are four experiments, four different orders of tasks have been created, which can be found in the appendix.

And also, to minimize the influence of task order in our analysis, five different task order were designed, to which “a” through “e” were assigned as indicated below.

Table 2. Task Order

Order	a	b	c	d	e
1	i	ii	iv	iii	ii
2	ii	iii	ii	ii	i
3	iii	i	iii	i	iv
4	iv	iv	i	iv	iii

Design of Experiment Order

Before the experiments, each of 20 participants were assigned a label 1 to 20. In order to totally randomize the experiment order and task order, the experiments were further designed as illustrated below.

Table 3. Designed Experiments Order

User	Experiment Order	Task Order
1	A	a
2	A	b
3	A	c
4	A	d
5	A	e
6	B	a
7	B	b
8	B	c
9	B	d
10	B	e
11	C	a
12	C	b
13	C	c
14	C	d
15	C	e
16	D	a
17	D	b
18	D	c
19	D	d
20	D	e

This design is called a fractional factorial design, by which it minimized the noise influence of experiment order and task order [8].

Questionnaire

To gather feedback for the usability of the product, a Use Questionnaire was used. This questionnaire was used because it is already widely accepted and validated for gathering usability feedback. The same questionnaire was provided to all the users to facilitate a comparison of the feedback received using the two methods. The Use Questionnaire is in Appendix E.

Pilot Study

Pilot study was a preliminary study conducted in order to evaluate feasibility, time, and adverse events in an attempt to improve the study design prior to the performance of a full-scale research project. Two users were recruited to do the pilot test, and each user was asked to finish all the four experiments with 8 simple tasks in each experiment. After each experiment, subjects were asked to fill out a user questionnaire asking about their satisfaction with each prototype. After the experiment, a short interview was implemented with the participants to get their feedback and suggestions on the testing procedure.

After the pilot test, some changes were made based on their feedback. One of the most significant changes was adding a description and a demo to explain what AR is before the experiment since there are many people who are not familiar with AR technology. Another change was implemented a detailed explanation on how to use all the prototypes. All the changes were added to the Study Script.

Setup for Users Testing

The user testing was conducted in College of Architecture building on the Georgia Institute of Technology campus. All the test facilities were located in an empty space. A large desk providing enough space for all the experiment's materials. A table lamp providing additional light, which enabled the marker can be recognized more efficiently. A tablet holder, which enabled the participants to be able to use both hands to interact with the prototypes was provided.

Data Analysis

The next step was to analyze the feedback received from users. Using the exact same questionnaire for all the four groups helped to compare the feedback that it was easy to identify what part of the usability testing works better using what method.

The analysis started by grouping responses on a spreadsheet. Four sets of groupings were made. One for all the users who tested the actual product (NonAR group). Second for all the users who used augmented reality for usability testing just using the marker card (AR group). Third for all the users using augmented reality for usability testing, which used a marker card pasted on a 3D printed model (AF group) and fourth for all the users using Tangible AR for usability testing, which used the physical model for interaction (TAR group). Since there are four sections in the User Questionnaire (Usefulness, Ease of use, Ease of learning, and Satisfaction), responses for each section were grouped separately for analysis. After separate analysis, an overall analysis was implemented.

To test whether there was a significant difference in the responses for the four groups, Friedman's test [9-11] was used. Friedman's Test is only capable of detecting whether there is a statistically significant difference between all the groups, thus if there was a difference between at least two of the groups, it is necessary to find out where the difference is. In this research, Wilcoxon's Signed Rank Test is used to do a paired comparison to see if groups are different from each other. Wilcoxon's Signed Rank Test is a non-parametric statistical hypothesis test used when comparing repeated measurements on a single sample to assess whether their population mean rank differ, it was proposed by Frank Wilcoxon [12] based on the assumption that the data comes from the same population and is measured on ordinal scale.

CHAPTER 4

CHALLENGES DURING THE PROCESS

Product for Usability Testing

The Sunbeam heater SFH5264MW used for usability testing turned out to be a very successful product for this study. The size of this product was not too large, which enabled time efficiency when setting up different experiments for participants, it was also not too small, which enabled the marker card to be located on top of the knobs so it still could be captured easily.

Knob Resistance of 3D Printed Model

Since this study focused on the early stages of the design process, designers usually have not developed a design concept with a great amount of detail, such as a mechanism design, but for our study, the internal structure was designed to simulate the actual product. It didn’t mean that the product technically functions this way, but it was able to simulate the different tactile elements, because the closer the tactile elements were to reality, the more accurate the experiment would be.

It took lots of time and effort to minimize differences in knob resistance. Different mechanisms were tried and finally, a fingerlike structure was selected. Based on the fingerlike structure, different sizes of the “figure” were printed out to see which one was the best. Figure 12 showed all the 3D printed knobs used in resistance simulation. Finally, the most elongated “finger” was determined to be the best to simulate the similar resistance.

Figure 12. Knob Models for Iteration

Programing for TAR

There are some existing online platforms which provide simple AR functions, such as Layar [13, 14] and Augment [13, 15]. The best they could achieve was recognizing just one marker in one screen at same time, but for the TAR group, cooperation between more than one marker was needed, which meant that the existing applications could not provide a shortcut to easily build a TAR prototype.

For TAR group, Unity3D [16] was chosen to build the prototype. Unity3D is a cross-platform game engine developed by Unity Technologies and used to develop video games for PC, consoles, mobile devices, and websites. They already have a toolkit called Vuforia SDK that be used in Unity3D to program the TAR prototype. Vuforia SDK is an Augmented Reality Software Development Kit for mobile devices that enables the creation of Augmented Reality applications. It uses Computer Vision technology to recognize and track planar images (Image Targets) and simple 3D objects. This image registration capability enabled developers to position and orient virtual objects, such as 3D models and other media in relation to real world images when these are viewed through the camera of a mobile device. [17]

For this case, 3 different models needed to export as obj file: main body, upper knob, and lower knob. There were 3 phases for programming. Firstly, the main part of this heater need to be linked to the square marker shown below, and the two knobs also had corresponding round markers. The second step was defining all the rotational ranges for the two knobs, such as: the lower knob had four different modes, requiring 30 degree rotation between each mode. The last group was adding effects to the different rotational ranges based on the second step, such as “click” sound.

After all the programming was finished, the apk file was exported by Unity 3D. Then, the apk file had to be installed on the tablet. After installing the app, we tested it by printing out the trackers and opening the app to interact with the prototype. Any problems would require making modifications in the programming.

The whole programming process was really difficult for designers. It could be better to have an AR developer on the design team to solve such problems. If there is evidence that TAR is really useful, it would be valuable to develop a toolkit or an application that solves the technical problems and enables everyone to use it without programing.

CHAPTER 5

DATA ANALYSIS AND RESULTS

Overall Use Questionnaire Result

In this research study, Use Questionnaire [18] was used to collect, measure, and analyze the user experience. Table 4 shows all the basic demographic information collected by pre-test questionnaire from each participant.

Table 4. Demographics

Pre-test questionnaire & Times			
	Born year	Gender	Race
#1	1990	M	Asian
#2	1987	M	Asian
#3	1990	M	Asian
#4	1995	M	White
#5	1992	F	Asian
#6	1994	M	Black or African American
#7	1993	M	White
#8	1991	M	White
#9	1994	F	White
#10	1993	M	White
#11	1969	F	Black or African American
#12	1990	M	White
#13	1997	M	White
#14	1994	F	White
#15	1993	M	White
#16	1991	F	Asian
#17	1991	M	White
#18	1991	F	Asian
#19	1993	M	White
#20	1992	M	Asian

The USE questionnaire consists of 4 sub-sections: Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Figure 13 shows all the questions in different sections.

			1	2	3	4	5	6	7	
USEFULNESS			1	2	3	4	5	6	7	
1	It helps me be more effective.	strongly disagree	○	○	○	○	○	○	○	strongly agree
2	It helps me be more productive.	strongly disagree	○	○	○	○	○	○	○	strongly agree
3	It is useful.	strongly disagree	○	○	○	○	○	○	○	strongly agree
4	It gives me more control over the activities in my life.	strongly disagree	○	○	○	○	○	○	○	strongly agree
5	It makes the things I want to accomplish easier to get done.	strongly disagree	○	○	○	○	○	○	○	strongly agree
6	It saves me time when I use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
7	It meets my needs.	strongly disagree	○	○	○	○	○	○	○	strongly agree
8	It does everything I would expect it to do.	strongly disagree	○	○	○	○	○	○	○	strongly agree
EASE OF USE			1	2	3	4	5	6	7	
9	It is easy to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree
10	It is simple to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree
11	It is user friendly.	strongly disagree	○	○	○	○	○	○	○	strongly agree
12	It requires the fewest steps possible to accomplish what I want to do with it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
13	It is flexible.	strongly disagree	○	○	○	○	○	○	○	strongly agree
14	Using it is effortless.	strongly disagree	○	○	○	○	○	○	○	strongly agree
15	I can use it without written instructions.	strongly disagree	○	○	○	○	○	○	○	strongly agree
16	I don't notice any inconsistencies as I use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
17	Both occasional and regular users would like it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
18	I can recover from mistakes quickly and easily.	strongly disagree	○	○	○	○	○	○	○	strongly agree
19	I can use it successfully every time.	strongly disagree	○	○	○	○	○	○	○	strongly agree
EASE OF LEARNING			1	2	3	4	5	6	7	
20	I learned to use it quickly.	strongly disagree	○	○	○	○	○	○	○	strongly agree
21	I easily remember how to use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
22	It is easy to learn to use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
23	I quickly became skillful with it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
SATISFACTION			1	2	3	4	5	6	7	
24	I am satisfied with it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
25	I would recommend it to a friend.	strongly disagree	○	○	○	○	○	○	○	strongly agree
26	It is fun to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree
27	It works the way I want it to work.	strongly disagree	○	○	○	○	○	○	○	strongly agree
28	It is wonderful.	strongly disagree	○	○	○	○	○	○	○	strongly agree
29	I feel I need to have it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
30	It is pleasant to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree

Figure 13. Use Questionnaire

For each of the questions, users were asked to rate their feelings towards one specific question on a 7-point scale from “strongly agree” to “strongly disagree”. The questionnaire was chosen because it was more comprehensive and accurate [19] and it was used so that the data could be directly compared with earlier work [20], which also had used the same questionnaire. It measured user experience in four aspects and the 7-point scale allows users more choices. This questionnaire had already been widely used in user experience evaluations.

After collection data, 20 feedbacks for four different groups (Non-AR, AR, AF, and TAR) for a total of 80 feedbacks for each of the 30 questions were collected. After checking the validity, all of the 2400 data was valid and eligible to be included in our analysis. All the raw data can be found in Appendix F.

Before applying statistical testing on data, descriptive statistics for each of the groups is shown in Tables 5-8. Each of the following tables shows mean, standard error, median, mode, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum and 95% confidence level of each question for each group.

Table 5. Descriptive Statistics of NonAR Group

NonAR												
	Mean	Median	Mode	Standard Deviation	Skewness	Standard Error	Variance	Kurtosis	Range	Min	Max	95% Confidence Level
Q1	5.7	6	7	1.26	-0.42	0.28	1.59	-1.53	3	4	7	0.02
Q2	5.5	6	4	1.28	-0.08	0.29	1.63	-1.74	3	4	7	0.02
Q3	5.85	6	7	1.23	-0.45	0.27	1.50	-1.47	3	4	7	0.02
Q4	5.3	5	4	1.26	0.24	0.28	1.59	-1.66	3	4	7	0.02
Q5	5.45	5	7	1.32	0.12	0.29	1.73	-1.82	3	4	7	0.02
Q6	5.1	5	4	1.41	-0.07	0.32	1.99	-0.45	5	2	7	0.02
Q7	6	6	7	0.97	-0.38	0.22	0.95	-1.12	3	4	7	0.01
Q8	6.2	7	7	1.20	-1.45	0.27	1.43	1.38	4	3	7	0.02
Q9	6.4	7	7	0.82	-1.55	0.18	0.67	2.61	3	4	7	0.01
Q10	6.45	7	7	0.83	-1.69	0.18	0.68	2.96	3	4	7	0.01
Q11	6.5	7	7	0.83	-1.86	0.18	0.68	3.44	3	4	7	0.01
Q12	6.05	6.5	7	1.23	-1.22	0.28	1.52	0.58	4	3	7	0.02
Q13	5.55	6	6	1.19	-0.44	0.27	1.42	-0.58	4	3	7	0.02
Q14	6.3	6	7	0.73	-0.55	0.16	0.54	-0.83	2	5	7	0.01
Q15	6.45	7	7	1.00	-2.49	0.22	1.00	7.17	4	3	7	0.01
Q16	6.4	6.5	7	0.68	-0.71	0.15	0.46	-0.45	2	5	7	0.01
Q17	6.1	6	7	0.91	-0.68	0.20	0.83	-0.35	3	4	7	0.01
Q18	6.3	7	7	0.92	-1.12	0.21	0.85	0.36	3	4	7	0.01
Q19	6.5	7	7	0.69	-1.08	0.15	0.47	0.08	2	5	7	0.01
Q20	6.65	7	7	0.49	-0.68	0.11	0.24	-1.72	1	6	7	0.01
Q21	6.6	7	7	0.60	-1.25	0.13	0.36	0.78	2	5	7	0.01
Q22	6.7	7	7	0.66	-2.08	0.15	0.43	3.18	2	5	7	0.01
Q23	6.75	7	7	0.55	-2.24	0.12	0.30	4.66	2	5	7	0.01
Q24	6.25	7	7	1.02	-1.22	0.23	1.04	0.45	3	4	7	0.01
Q25	5.8	6	6	1.24	-1.61	0.28	1.54	3.59	5	2	7	0.02
Q26	4.7	4.5	4	1.59	0.03	0.36	2.54	-0.82	5	2	7	0.02
Q27	6.2	7	7	1.28	-1.75	0.29	1.64	2.32	4	3	7	0.02
Q28	4.85	5	5	1.50	-0.35	0.33	2.24	-0.41	5	2	7	0.02
Q29	4.7	5	5	1.56	-0.55	0.35	2.43	0.61	6	1	7	0.02
Q30	5.3	5	6	1.22	-0.06	0.27	1.48	-1.02	4	3	7	0.02

Table 6. Descriptive Statistics of AR Group

AR												
	Mean	Median	Mode	Standard Deviation	Skewness	Standard Error	Variance	Kurtosis	Range	Min	Max	95% Confidence Level
Q1	3.95	4	4	0.94	0.11	0.21	0.89	0.19	4	2	6	0.01
Q2	3.8	4	4	1.06	0.44	0.24	1.12	0.64	4	2	6	0.01
Q3	4.6	4.5	4	1.10	0.38	0.24	1.20	-0.26	4	3	7	0.02
Q4	4.05	4	4	1.05	0.19	0.23	1.10	-0.22	4	2	6	0.01
Q5	4.1	4	4	1.33	0.98	0.30	1.78	0.66	5	2	7	0.02
Q6	3.65	4	4	1.42	0.45	0.32	2.03	0.60	6	1	7	0.02
Q7	4.75	4.5	4	1.33	0.21	0.30	1.78	-0.08	5	2	7	0.02
Q8	4.8	5	5	1.32	-0.05	0.30	1.75	0.16	5	2	7	0.02
Q9	4.3	4.5	5	1.26	-0.11	0.28	1.59	0.12	5	2	7	0.02
Q10	4.9	5	5	1.02	-0.77	0.23	1.04	2.93	5	2	7	0.01
Q11	4.55	4.5	3	1.54	0.09	0.34	2.37	-1.33	5	2	7	0.02
Q12	4.1	4	5	1.48	-0.08	0.33	2.20	-0.24	6	1	7	0.02
Q13	4.15	4	4	0.88	0.73	0.20	0.77	0.40	3	3	6	0.01
Q14	4.4	4.5	3	1.39	0.10	0.31	1.94	-1.06	5	2	7	0.02
Q15	4.5	4	3	1.61	0.42	0.36	2.58	-1.55	4	3	7	0.02
Q16	3.85	3.5	3	1.42	0.29	0.32	2.03	0.15	6	1	7	0.02
Q17	4.35	4	5	1.31	0.21	0.29	1.71	-0.59	5	2	7	0.02
Q18	4.55	5	5	1.15	0.10	0.26	1.31	-0.42	4	3	7	0.02
Q19	4.5	5	5	1.47	-0.22	0.33	2.16	-1.01	5	2	7	0.02
Q20	5.45	6	7	1.43	-0.55	0.32	2.05	-0.92	4	3	7	0.02
Q21	5.75	6	7	1.25	-0.55	0.28	1.57	-0.67	4	3	7	0.02
Q22	5.5	5.5	7	1.32	-0.46	0.29	1.74	-0.68	4	3	7	0.02
Q23	5	5	5	1.17	-0.22	0.26	1.37	-0.22	4	3	7	0.02
Q24	4.45	4	3	1.39	0.50	0.31	1.94	-0.99	4	3	7	0.02
Q25	4.3	4	3	1.42	0.51	0.32	2.01	-0.52	5	2	7	0.02
Q26	4.7	5	5	1.22	-0.33	0.27	1.48	0.08	5	2	7	0.02
Q27	4.3	4	5	1.03	-0.36	0.23	1.06	-0.02	4	2	6	0.01
Q28	4.2	4	4	1.61	0.14	0.36	2.59	-0.37	6	1	7	0.02
Q29	3.65	3.5	3	1.27	0.40	0.28	1.61	-0.70	4	2	6	0.02
Q30	4.7	4.5	4	1.34	0.47	0.30	1.80	-0.76	4	3	7	0.02

Table 7. Descriptive Statistics of AF Group

AF												
	Mean	Median	Mode	Standard Deviation	Skewness	Standard Error	Variance	Kurtosis	Range	Min	Max	95% Confidence Level
Q1	4.6	4	4	1.23	0.68	0.28	1.52	-0.43	4	3	7	0.02
Q2	4.35	4	4	1.04	0.76	0.23	1.08	0.86	4	3	7	0.01
Q3	4.75	5	5	1.41	0.24	0.32	1.99	-1.12	4	3	7	0.02
Q4	4.25	4	4	0.91	1.31	0.20	0.83	3.41	4	3	7	0.01
Q5	4.05	4	3	1.15	0.83	0.26	1.31	-0.67	3	3	6	0.02
Q6	4.05	4	4	1.36	0.32	0.30	1.84	-0.04	5	2	7	0.02
Q7	5	5	5	1.26	0.35	0.28	1.58	-0.58	4	3	7	0.02
Q8	5.2	5	7	1.47	-0.05	0.33	2.17	-1.35	4	3	7	0.02
Q9	4.9	5	5	1.12	-0.54	0.25	1.25	1.13	5	2	7	0.02
Q10	4.85	5	5	1.27	-0.38	0.28	1.61	0.34	5	2	7	0.02
Q11	4.8	5	5	1.44	-0.32	0.32	2.06	-0.80	5	2	7	0.02
Q12	4.75	4.5	4	1.68	-0.08	0.38	2.83	-1.22	5	2	7	0.02
Q13	4.55	4.5	5	1.15	0.33	0.26	1.31	-0.47	4	3	7	0.02
Q14	4.2	4	3	1.32	-0.10	0.30	1.75	-1.14	4	2	6	0.02
Q15	4.65	5	3	1.60	0.04	0.36	2.56	-1.34	5	2	7	0.02
Q16	4.3	4	3	1.56	0.46	0.35	2.43	-1.21	5	2	7	0.02
Q17	4.6	5	5	1.23	-0.26	0.28	1.52	-0.06	5	2	7	0.02
Q18	4.75	5	5	1.16	-0.57	0.26	1.36	0.78	5	2	7	0.02
Q19	4.7	5	5	1.45	-0.33	0.33	2.12	-0.45	5	2	7	0.02
Q20	5.35	5	5	1.27	-0.74	0.28	1.61	1.11	5	2	7	0.02
Q21	5.4	5.5	5	1.31	-1.00	0.29	1.73	1.26	5	2	7	0.02
Q22	5.55	5.5	5	1.32	-0.89	0.29	1.73	1.21	5	2	7	0.02
Q23	5	5	5	1.34	0.15	0.30	1.79	-0.88	4	3	7	0.02
Q24	4.75	4	4	1.25	0.53	0.28	1.57	-1.02	4	3	7	0.02
Q25	4.55	4.5	5	1.43	0.31	0.32	2.05	-0.47	5	2	7	0.02
Q26	5.05	5	5	1.36	-0.52	0.30	1.84	0.11	5	2	7	0.02
Q27	4.45	4	4	1.28	0.38	0.29	1.63	0.28	5	2	7	0.02
Q28	4.45	4.5	5	1.50	-0.36	0.34	2.26	0.47	6	1	7	0.02
Q29	3.8	4	4	1.15	1.35	0.26	1.33	2.38	5	2	7	0.02
Q30	4.85	5	5	1.31	-0.32	0.29	1.71	-0.06	5	2	7	0.02

Table 8. Descriptive Statistics of TAR Group

	TAR											
	Mean	Median	Mode	Standard Deviation	Skewness	Standard Error	Variance	Kurtosis	Range	Min	Max	95% Confidence Level
Q1	5.3	5.5	6	1.17	-0.22	0.26	1.38	-0.92	4	3	7	0.02
Q2	5.2	5	6	1.11	-0.18	0.25	1.22	-0.75	4	3	7	0.02
Q3	5.7	6	7	1.17	-0.21	0.26	1.38	-1.46	3	4	7	0.02
Q4	5.4	5.5	4	1.27	-0.18	0.28	1.62	-1.18	4	3	7	0.02
Q5	5.45	6	7	1.39	-0.40	0.31	1.94	-1.15	4	3	7	0.02
Q6	5.1	5	6	1.07	-0.22	0.24	1.15	-0.99	4	3	7	0.02
Q7	5.35	5.5	4	1.18	0.08	0.26	1.40	-1.54	3	4	7	0.02
Q8	5.9	6	7	1.12	-0.54	0.25	1.25	-1.08	3	4	7	0.02
Q9	6.25	6.5	7	0.91	-1.02	0.20	0.83	0.26	3	4	7	0.01
Q10	6.05	6	6	0.94	-0.94	0.21	0.89	0.41	3	4	7	0.01
Q11	6.15	7	7	1.23	-1.46	0.27	1.50	1.22	4	3	7	0.02
Q12	5.9	6	7	1.21	-0.98	0.27	1.46	0.19	4	3	7	0.02
Q13	5.5	5.5	5	1.10	-0.40	0.25	1.21	-0.08	4	3	7	0.02
Q14	5.65	6	6	1.39	-1.00	0.31	1.92	-0.11	4	3	7	0.02
Q15	6	6	6	1.08	-1.41	0.24	1.16	2.17	4	3	7	0.02
Q16	5.85	6	5	0.93	-0.11	0.21	0.87	-1.08	3	4	7	0.01
Q17	5.9	6	6	0.79	-0.53	0.18	0.62	0.49	3	4	7	0.01
Q18	5.75	6	7	1.33	-0.97	0.30	1.78	-0.08	4	3	7	0.02
Q19	6.2	7	7	1.06	-1.04	0.24	1.12	-0.20	3	4	7	0.01
Q20	6.6	7	7	0.68	-1.51	0.15	0.46	1.17	2	5	7	0.01
Q21	6.55	7	7	0.69	-1.28	0.15	0.47	0.54	2	5	7	0.01
Q22	6.6	7	7	0.60	-1.25	0.13	0.36	0.78	2	5	7	0.01
Q23	6.55	7	7	0.51	-0.22	0.11	0.26	-2.18	1	6	7	0.01
Q24	6.2	6	7	0.89	-0.92	0.20	0.80	0.22	3	4	7	0.01
Q25	5.7	6	6	1.03	-0.28	0.23	1.06	-0.94	3	4	7	0.01
Q26	6.2	7	7	1.11	-1.22	0.25	1.22	0.18	3	4	7	0.02
Q27	6.2	6	6	0.95	-2.07	0.21	0.91	6.18	4	3	7	0.01
Q28	5.65	6	7	1.31	-0.52	0.29	1.71	-1.01	4	3	7	0.02
Q29	4.55	4	4	1.47	0.33	0.33	2.16	-0.69	5	2	7	0.02
Q30	5.75	6	7	1.07	-0.30	0.24	1.14	-1.09	3	4	7	0.02

To get a clear comparison of the answers for the four groups, the mean of the 20 questions of four groups are compared in Table 9 and the line graph is plotted in Figure 14.

Table 9. Average Rate of Each Question for Four Groups

	NonAR	AR	AF	TAR
Q1	5.7	3.95	4.6	5.3
Q2	5.5	3.8	4.35	5.2
Q3	5.85	4.6	4.75	5.7
Q4	5.3	4.05	4.25	5.4
Q5	5.45	4.1	4.05	5.45
Q6	5.1	3.65	4.05	5.1
Q7	6	4.75	5	5.35
Q8	6.2	4.8	5.2	5.9
Q9	6.4	4.3	4.9	6.25
Q10	6.45	4.9	4.85	6.05
Q11	6.5	4.55	4.8	6.15
Q12	6.05	4.1	4.75	5.9
Q13	5.55	4.15	4.55	5.5
Q14	6.3	4.4	4.2	5.65
Q15	6.45	4.5	4.65	6
Q16	6.4	3.85	4.3	5.85
Q17	6.1	4.35	4.6	5.9
Q18	6.3	4.55	4.75	5.75
Q19	6.5	4.5	4.7	6.2
Q20	6.65	5.45	5.35	6.6
Q21	6.6	5.75	5.4	6.55
Q22	6.7	5.5	5.55	6.6
Q23	6.75	5	5	6.55
Q24	6.25	4.45	4.75	6.2
Q25	5.8	4.3	4.55	5.7
Q26	4.7	4.7	5.05	6.2
Q27	6.2	4.3	4.45	6.2
Q28	4.85	4.2	4.45	5.65
Q29	4.7	3.65	3.8	4.55
Q30	5.3	4.7	4.85	5.75

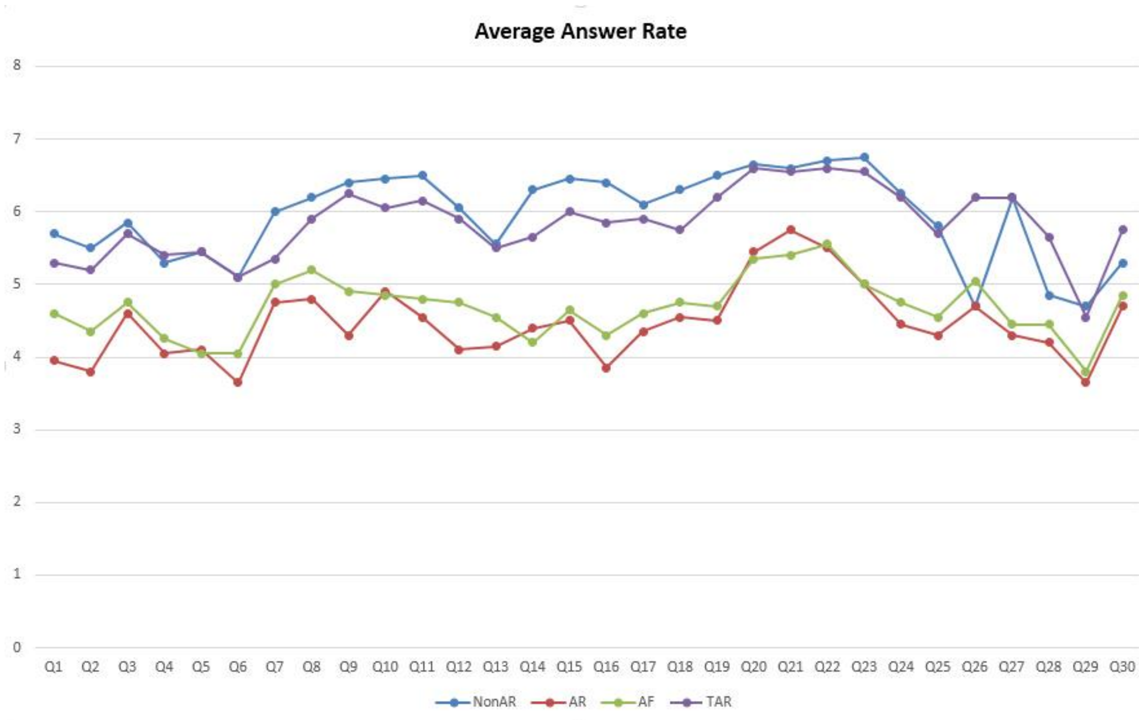


Figure 14. Line Graph of Average Answer Rate

It can be observed that among the four series of answers, the two with higher scores were similar, which belonged to NonAR and TAR, while the lower two were also pretty similar, which belonged to AR and AF. This simply showed that comparably, the user experience for TAR and Non-AR is similar, and the reason might have been that among the three AR experiments, TAR was the one which best simulated the actual products, it not only allowed the interactions between 3D printed model and humans, but also allowed humans to see the interactions from a screen.

Next, based on users' answers to each question in one sub section, the User Experience Index (UEI) for Usefulness, Ease of Use, Ease of Learning, and Satisfaction were calculated. Then simply choosing the average of answers in each subsection as each question stood for one aspect of the subsection, for example, the section of Ease of Learning contains four questions, and the answers of user#1 was 5,6,7,6 for each of them, thus the method for calculating the UEI of Ease of Learning for user#1 is:

$$\text{UEI} - \text{Ease of Learning} = (5+6+7+6)/4 = 6$$

In this way, Table 10, 11, 12, 13 shows the UEI of each section for each user.

Table 10. User Experience Index of NonAR Group

NonAR				
	Usefulness	Ease of use	Ease of learning	Satisfaction
#1	5.63	6.73	7.00	6.00
#2	6.38	5.82	6.25	4.71
#3	4.50	5.91	7.00	5.29
#4	4.50	6.45	7.00	5.57
#5	4.25	4.73	6.50	2.71
#6	6.88	6.64	7.00	5.43
#7	5.50	7.00	7.00	4.86
#8	4.50	6.73	6.50	4.00
#9	5.50	6.18	7.00	5.43
#10	6.50	6.55	7.00	6.43
#11	6.13	6.18	5.75	6.29
#12	5.75	6.00	6.00	5.14
#13	5.00	5.64	7.00	5.57
#14	7.00	6.73	7.00	6.00
#15	7.00	7.00	7.00	7.00
#16	6.75	6.55	6.50	6.71
#17	4.00	6.27	6.50	4.57
#18	6.88	6.82	7.00	7.00
#19	4.88	4.91	5.50	3.57
#20	5.25	6.64	7.00	5.71

Table 11. User Experience Index of AR Group

AR				
	Usefulness	Ease of use	Ease of learning	Satisfaction
#1	4.38	5.36	6.50	5.71
#2	5.50	5.18	4.75	4.43
#3	2.38	2.91	3.25	2.14
#4	3.00	2.82	4.75	3.43
#5	3.00	3.73	6.00	3.57
#6	5.50	5.64	7.00	5.00
#7	5.38	6.45	7.00	6.29
#8	4.50	5.00	6.00	5.14
#9	3.63	5.45	6.50	4.57
#10	6.00	4.91	6.75	5.71
#11	5.25	4.45	3.75	4.57
#12	4.63	3.91	4.50	4.43
#13	4.13	3.91	4.50	3.71
#14	3.63	3.55	5.75	3.57
#15	3.63	4.00	6.00	3.14
#16	4.13	4.64	4.00	4.00
#17	4.00	3.64	6.00	3.57
#18	4.25	4.09	6.00	6.00
#19	3.50	4.09	4.50	3.29
#20	3.88	3.82	5.00	4.29

Table 12. User Experience Index of AF Group

AF				
	Usefulness	Ease of use	Ease of learning	Satisfaction
#1	4.50	5.73	6.50	5.71
#2	5.38	5.00	4.50	4.43
#3	2.88	2.45	2.25	2.29
#4	3.50	3.09	4.75	3.57
#5	3.88	4.36	6.00	3.43
#6	5.50	5.64	7.00	5.00
#7	6.38	6.73	7.00	5.00
#8	4.63	5.00	6.25	5.00
#9	5.00	5.73	5.50	5.00
#10	5.13	5.73	6.50	6.00
#11	4.75	4.55	4.00	4.14
#12	4.88	4.18	5.00	4.57
#13	3.63	4.09	4.50	4.29
#14	4.13	4.45	5.75	4.57
#15	6.50	6.18	6.75	6.86
#16	4.25	4.09	4.50	4.14
#17	4.00	3.82	4.50	3.57
#18	4.25	4.09	6.00	6.00
#19	3.50	3.91	4.50	3.14
#20	4.00	4.00	4.75	4.43

Table 13. User Experience Index of TAR Group

TAR				
	Usefulness	Ease of use	Ease of learning	Satisfaction
#1	5.13	6.73	6.75	6.57
#2	6.25	5.91	6.50	4.86
#3	4.75	5.82	7.00	6.14
#4	4.63	6.18	6.75	5.43
#5	5.50	5.82	6.00	5.57
#6	5.88	6.45	7.00	5.14
#7	5.75	6.82	7.00	6.86
#8	5.38	5.73	6.50	5.86
#9	4.50	5.18	6.75	4.86
#10	6.00	5.91	7.00	5.71
#11	6.00	6.36	5.75	6.14
#12	6.00	5.18	6.00	5.86
#13	4.13	5.55	6.00	5.14
#14	3.75	4.36	7.00	4.00
#15	7.00	7.00	7.00	7.00
#16	6.63	6.55	6.50	6.57
#17	4.00	5.36	6.75	5.57
#18	6.63	6.55	7.00	7.00
#19	4.38	4.64	5.25	4.29
#20	6.25	6.45	7.00	6.43

Figure 15, 16, 17, 18 show four boxplots based on the data listed above in each table for each of the sections.

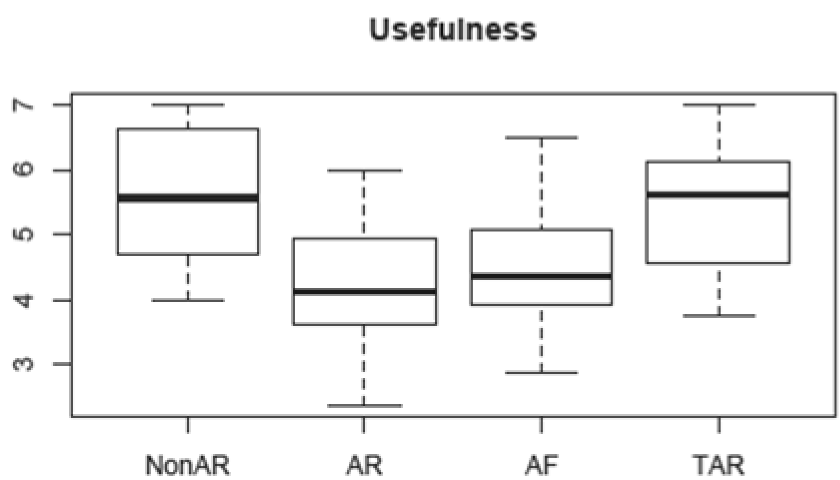


Figure 15. Boxplot of Usefulness

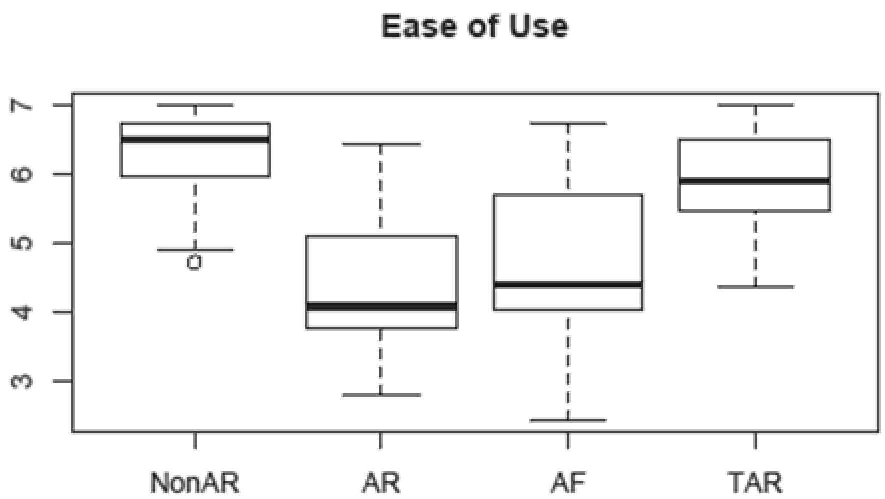


Figure 16. Boxplot of Ease of Use

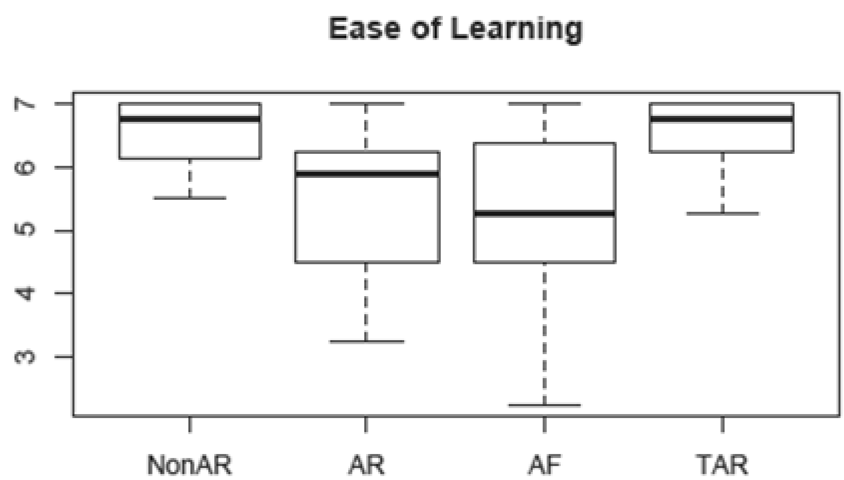


Figure 17. Boxplot of Ease of Learning

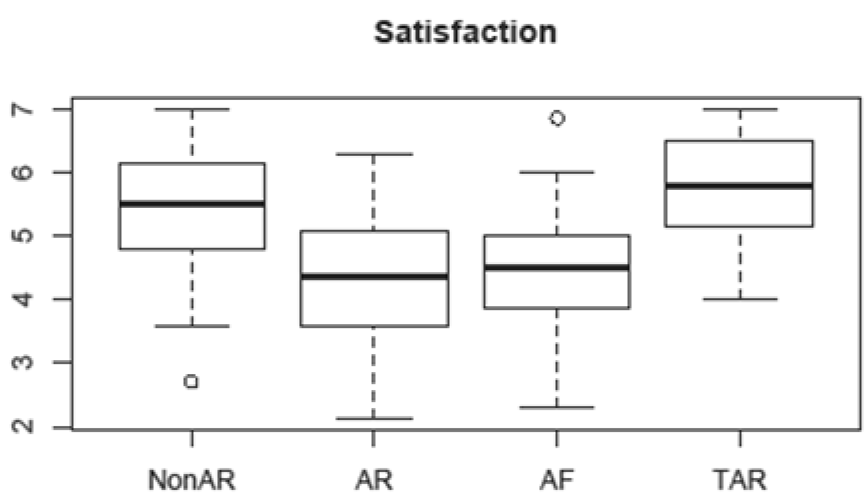


Figure 18. Boxplot of Satisfaction

An observation can be made from the boxplot above. Firstly, for all the four sections, NonAR and TAR were similar in their median, interquartile range, which might indicate that the user experience for NonAR and TAR was similar. Secondly, AR and AF was much lower compared to NonAR, which meant the user experience for these two groups was lower in Usefulness, Ease of Use, Ease of Learning, and Satisfaction. The reasons were obviously that these two groups were not effective in product simulation, AR do not have a 3D printed model for users to touch and feel, while AF only allowed users to interact through the screen, which greatly reduced the quality of the user experience. Next, Friedman's test was used to test whether there were statistically significant differences between the user experiences of the four groups.

Friedman's Test

In this research, Friedman's Test was used to analyze and determine whether there were statistically significant difference between the user experiences of four experiments. Friedman's test was a non-parametric statistical test developed by Milton Friedman [9-11].

It was used to detect differences in treatments across multiple test attempts. The reason of choosing Friedman's test as our main hypothesis testing methodology was that the population of four different experiments was the same group of people, and each of them was asked to provide feedback in four different occasions, and most importantly, the dependent variable which the user's answer in our analysis is ordinal and does not based on normality assumption.

The null hypothesis H_0 and alternative hypothesis H_1 of this test was stated as below:

H_0: There is no difference between four of our groups.

H_1: There are at least two groups that show a difference.

The statistical significant level Alpha was also chosen to be:

$$a = 0.05$$

By conducting Friedman's Test on our data, a p-value measured how much the null hypothesis was supported. So, if the p-value of the Friedman's test was larger than 0.05, it is fail to reject the null hypothesis and would conclude that there was no difference between user experience in four of the experiments, on the other hand, if the p-value of the Friedman's test was lower than 0.05, the null hypothesis is rejected and it is concluded that at least there were two groups that were different in their user experience.

Wilcoxon's Signed Rank Test

The Friedman's Test was only capable of detecting whether there were statistically significant differences between all the groups, thus if it is detected that there were differences between two groups, where the difference must be find out. In this research, the Wilcoxon's Signed Rank Test was used to do paired comparison to see if each groups were different from one another. Wilcoxon's Signed Rank Test was a non-parametric statistical hypothesis test used when comparing repeated measurements on a single sample to assess whether their population mean rank differ, it was proposed by Frank Wilcoxon [12] based on the assumption that the data comes from the same population and measured on ordinal scale. The Wilcoxon's test was performed on two groups of data, and the null hypothesis and alternative hypothesis was stated as following:

H_0: There is no difference between the two groups.

H_1: There is a difference between the two groups.

The statistical significant level Alpha was also chosen to be:

$$a = 0.05$$

So that if the p-value derived from the Wilcoxon's test is larger than a, then the null hypothesis is not rejected and it is concluded that there is no difference between the two groups; if the p-value we get from the Wilcoxon's test is lower than a, the null hypothesis is rejected and it is concluded that there is a difference between the two groups. In our case, Wilcoxon's signed rank test is used as an ad-hoc analysis to assess whether the user experience between two groups is statistically different.

Data Analysis for Each Section

In this section, Friedman's Test was applied on four groups of data for Usefulness, Ease of Use, Ease of Learning and Satisfaction. When it is found that there is a statistically significant difference between four groups, further research is conducted to detect where the difference is, i.e., which of the two groups showed a big difference and which two of them did not have a large difference. Finally, a conclusion about whether the user experience of the three AR groups is statistically significantly different to NonAR or not was got. All the data manipulation and analysis was done in Excel and R.

Usefulness

Table 14 shows the UEI of Usefulness from 20 users for four groups:

Table 14. User Experience Index of Usefulness

Usefulness				
	NonAR	**AR**	**AF**	**TAR**
#1	5.63	4.38	4.50	5.13
#2	6.38	5.50	5.38	6.25
#3	4.50	2.38	2.88	4.75
#4	4.50	3.00	3.50	4.63
#5	4.25	3.00	3.88	5.50
#6	6.88	5.50	5.50	5.88
#7	5.50	5.38	6.38	5.75
#8	4.50	4.50	4.63	5.38
#9	5.50	3.63	5.00	4.50
#10	6.50	6.00	5.13	6.00
#11	6.13	5.25	4.75	6.00
#12	5.75	4.63	4.88	6.00
#13	5.00	4.13	3.63	4.13
#14	7.00	3.63	4.13	3.75
#15	7.00	3.63	6.50	7.00
#16	6.75	4.13	4.25	6.63
#17	4.00	4.00	4.00	4.00
#18	6.88	4.25	4.25	6.63
#19	4.88	3.50	3.50	4.38
#20	5.25	3.88	4.00	6.25

Friedman's Test results is:

Friedman chi-squared = 35.1967, df = 3, p-value = 1.107e-07

As the p-value is extremely low, the null hypothesis is rejected and it is concluded that there is significant difference between four groups.

Since the differences between the four groups were detected, ad-hoc analysis is necessary to determine where the difference is and how much it is. Figure 19 shows a parallel coordinates plot of usefulness.

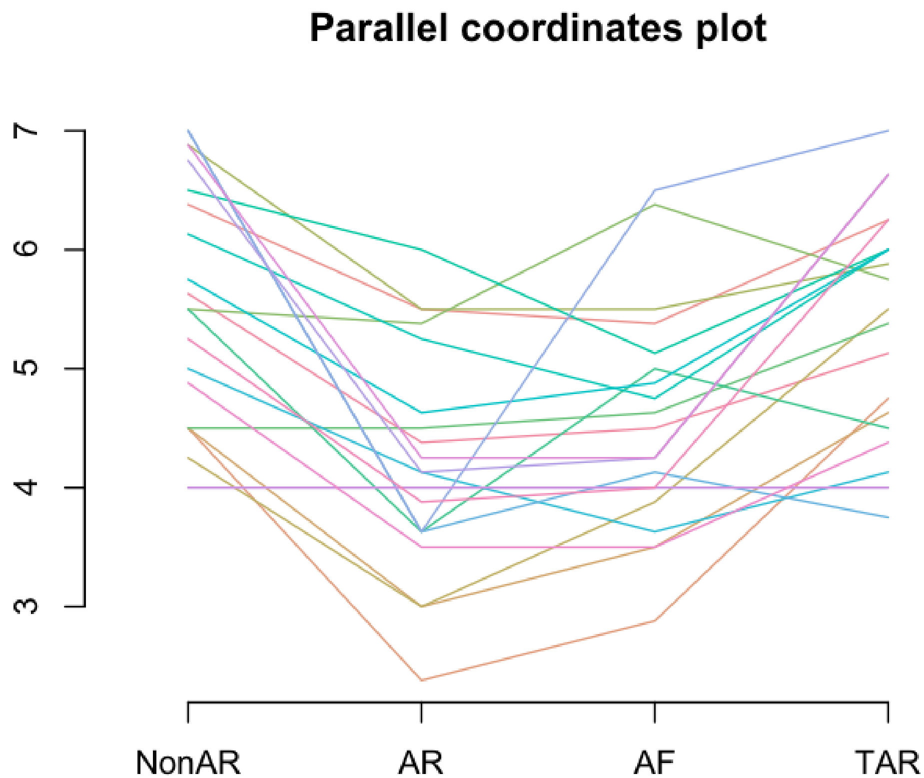


Figure 19. Parallel Coordinates Plot of Usefulness

This plot shows how the answers varies which came from same person in different groups. It is obvious that each user is having higher value when it comes to NonAR and TAR, the data of AF is similar to that of AR, while the data of NonAR is similar to that of TAR. Next, Wilcoxon's test was applied to see whether groups have any difference with each another, the results are shown below:

P-value for paired groups:

AR - AF	4.542611e-01
NonAR - AF	1.343496e-03
TAR - AF	1.703710e-02
NonAR - AR	1.415781e-06
TAR - AR	4.047383e-05
TAR - NonAR	8.954669e-01

Figure 20 shows the boxplot for differences of paired comparison.

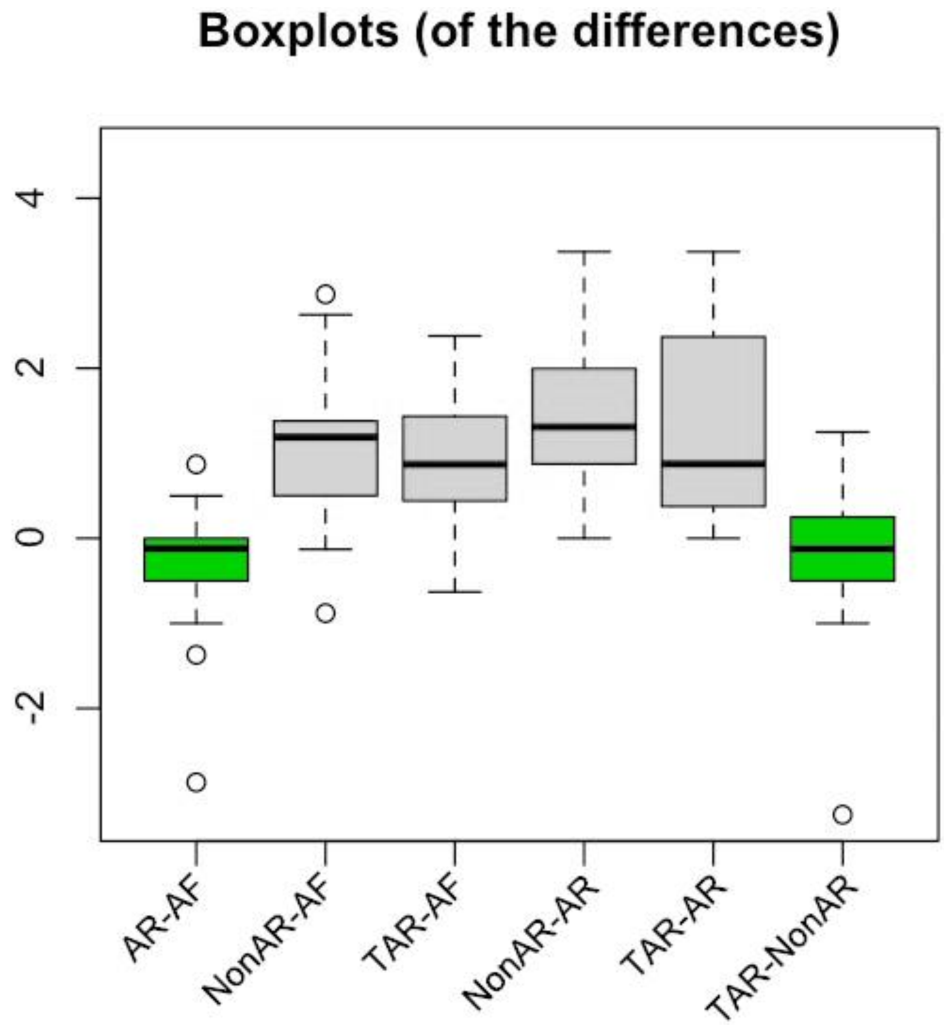


Figure 20. Boxplots of Differences for Usefulness

In the boxplot above, the grey ones are those with small p-value and thus show siginificant differences, while the green ones are those with large p-value and thus show no difference. The results shows that there are two pairs of groups that have no differences, which are:

TAR with NonAR

AR with AF

The result shows that the user experience of Userfulness in AR and AF differs with that in NonAR, and the user experience in Usefulness for users in NonAR is the similar with that in TAR.

Ease Of Use

Table 15 shows the UEI of Ease of Use from 20 users for four groups.

Table 15. User Experience Index of Ease of Use

Ease of use				
	NonAR	AR	AF	TAR
#1	6.73	5.36	5.73	6.73
#2	5.82	5.18	5.00	5.91
#3	5.91	2.91	2.45	5.82
#4	6.45	2.82	3.09	6.18
#5	4.73	3.73	4.36	5.82
#6	6.64	5.64	5.64	6.45
#7	7.00	6.45	6.73	6.82
#8	6.73	5.00	5.00	5.73
#9	6.18	5.45	5.73	5.18
#10	6.55	4.91	5.73	5.91
#11	6.18	4.45	4.55	6.36
#12	6.00	3.91	4.18	5.18
#13	5.64	3.91	4.09	5.55
#14	6.73	3.55	4.45	4.36
#15	7.00	4.00	6.18	7.00
#16	6.55	4.64	4.09	6.55
#17	6.27	3.64	3.82	5.36
#18	6.82	4.09	4.09	6.55
#19	4.91	4.09	3.91	4.64
#20	6.64	3.82	4.00	6.45

Friedman's Test results is:

Friedman chi-squared = 46.9175, df = 3, p-value = 3.619e-10

As the p-value is extremely low, the null hypothesis is rejected and it is concluded that there are significant differences between four groups.

Since the differences between the four groups were detected, ad-hoc analysis is necessary to determine where the difference is and how much it is. Figure 21 shows a parallel coordinates plot of Ease of Use.

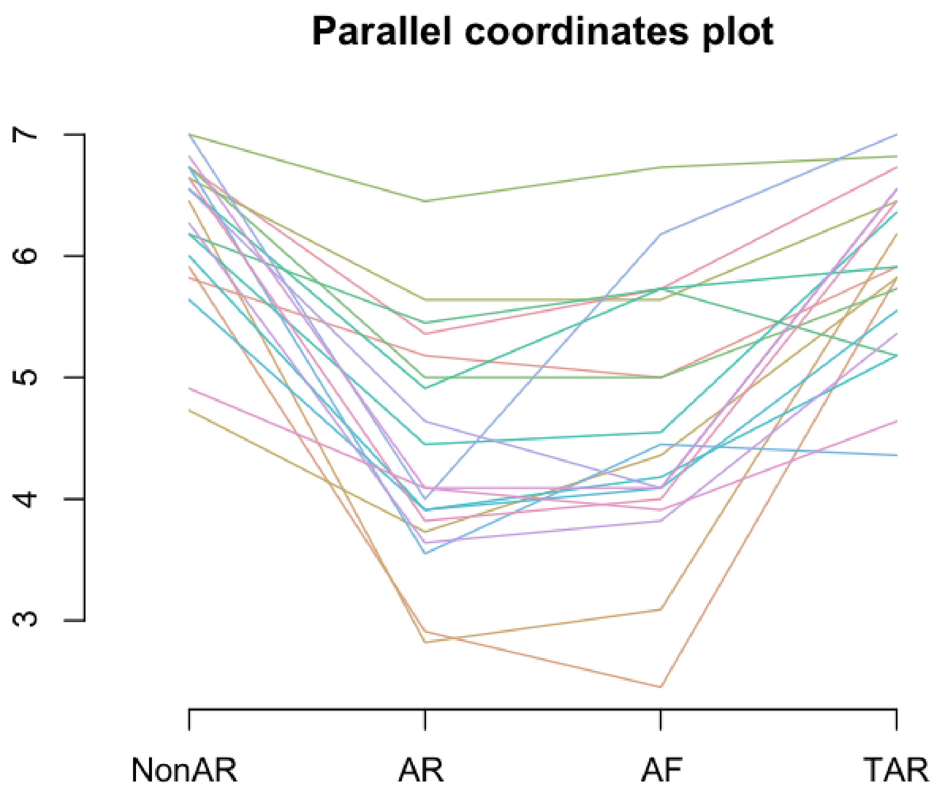


Figure 21. Parallel Coordinates Plot of Ease of Use

This plot shows how the answers varies which came from same person in different groups. One can see that all the lines are having higher value when they come to NonAR and TAR, the data of AF is similar to that of AR, while the data of NonAR is similar to that of TAR. Next, Wilcoxon's test was applied to see whether groups have any difference with each another, the results are shown below:

P-value for paired groups:

AR - AF	5.990523e-01
NonAR - AF	5.198619e-06
TAR - AF	1.050742e-02
NonAR – AR	3.977727e-09

TAR - AR 6.092923e-05

TAR – NonAR 3.023716e-01

Figure 22 shows the boxplots for differences of paired comparison.

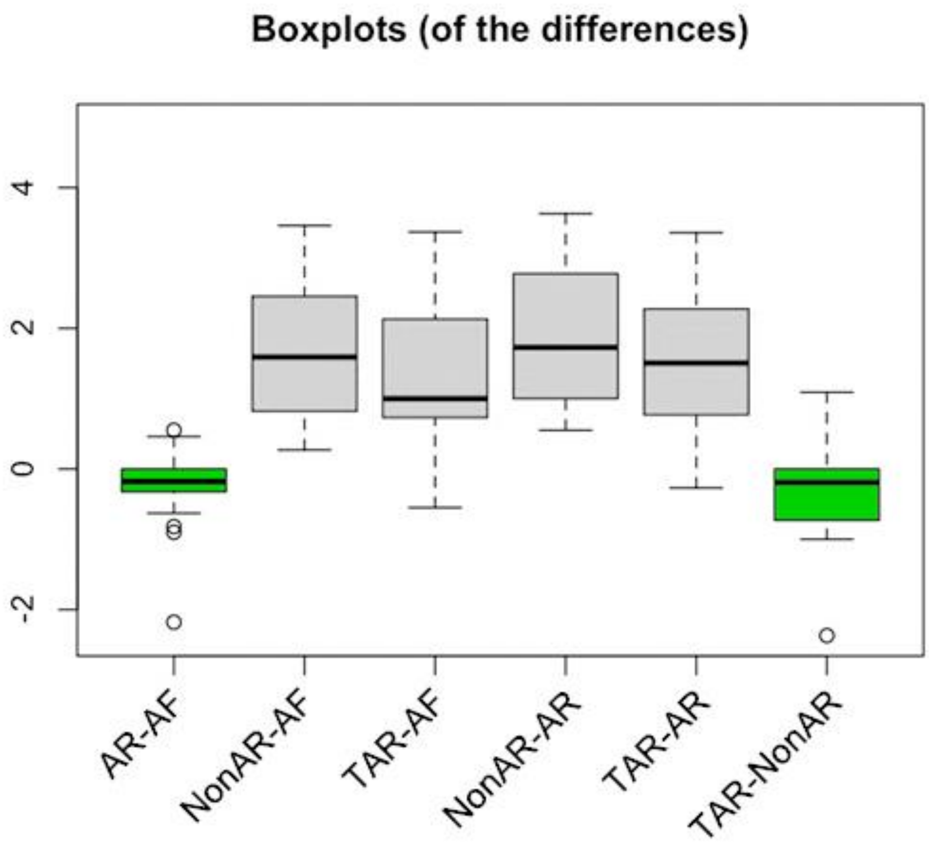


Figure 22. Boxplots of Differences for Ease of Use

In the boxplot above, the gray ones are those with small p-value and thus show big differences, while the green ones are those with large p-value and thus show no difference. The results show that there are two pairs of groups have no differences, which are:

TAR with NonAR

AR with AF

The result shows that the user experience of Ease of Use in AR and AF differs with that in NonAR, and the user experience in Ease of Use for users in NonAR is the similar to that in TAR.

Ease Of Learning

Table 16 shows the UEI of Ease of Learning from 20 users for four groups.

Table 16. User Experience Index of Ease of Learning

Ease of learning				
	NonAR	AR	AF	TAR
#1	7.00	6.50	6.50	6.75
#2	6.25	4.75	4.50	6.50
#3	7.00	3.25	2.25	7.00
#4	7.00	4.75	4.75	6.75
#5	6.50	6.00	6.00	6.00
#6	7.00	7.00	7.00	7.00
#7	7.00	7.00	7.00	7.00
#8	6.50	6.00	6.25	6.50
#9	7.00	6.50	5.50	6.75
#10	7.00	6.75	6.50	7.00
#11	5.75	3.75	4.00	5.75
#12	6.00	4.50	5.00	6.00
#13	7.00	4.50	4.50	6.00
#14	7.00	5.75	5.75	7.00
#15	7.00	6.00	6.75	7.00
#16	6.50	4.00	4.50	6.50
#17	6.50	6.00	4.50	6.75
#18	7.00	6.00	6.00	7.00
#19	5.50	4.50	4.50	5.25
#20	7.00	5.00	4.75	7.00

Friedman's Test results is:

Friedman chi-squared = 36.681, df = 3, p-value = 5.375e-08

As the p-value is extremely low, the null hypothesis is rejected and it is concluded that there is significant difference between the four groups.

Since the differences between the four groups are detected, ad-hoc analysis is necessary to determine where the difference is and how much it is. Figure 23 shows a parallel coordinates plot of Ease of Learning.

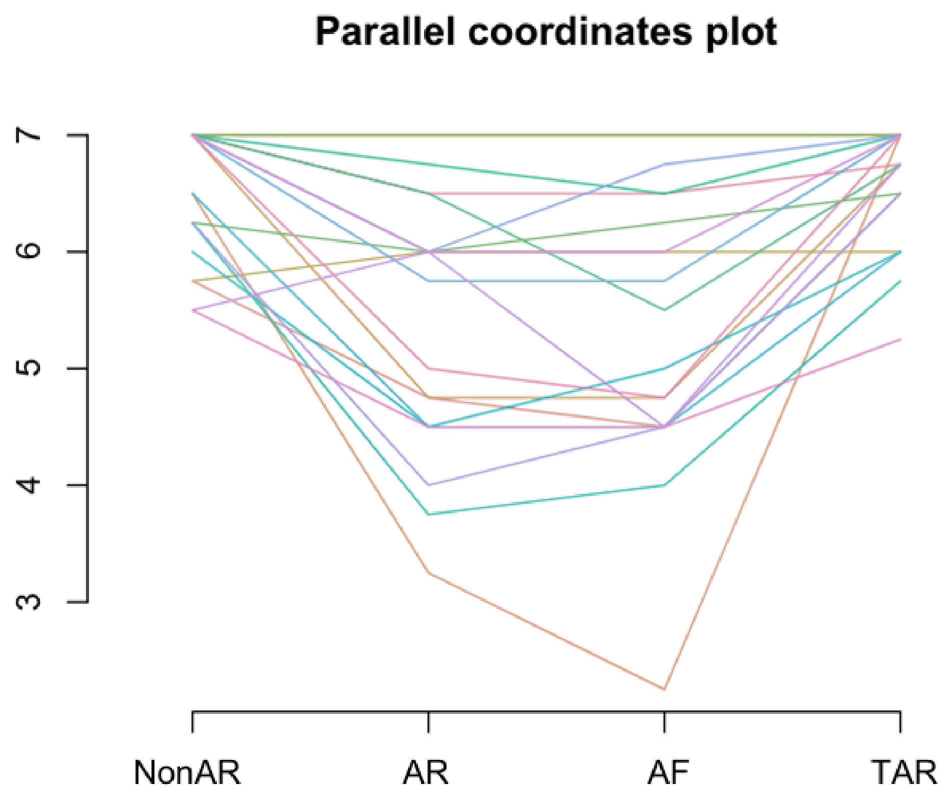


Figure 23. Parallel Coordinates Plot of Ease of Learning

This plot shows how the answers varies which came from same person in different groups. One can see that each user rates higher value when it comes to NonAR and TAR, the data of AF is similar to that of AR, while the data of NonAR is similar to that of TAR. Next, Wilcoxon's test was applied to see whether groups have any difference with each another, the results are shown below:

P-value for paired groups:

AR - AF	9.970198e-01
NonAR - AF	2.137911e-04
TAR - AF	2.421746e-05
NonAR - AR	4.160423e-04

TAR - AR 8.527298e-05

TAR - NonAR 9.865759e-01

Figure 24 shows the boxplots for differences of paired comparison.

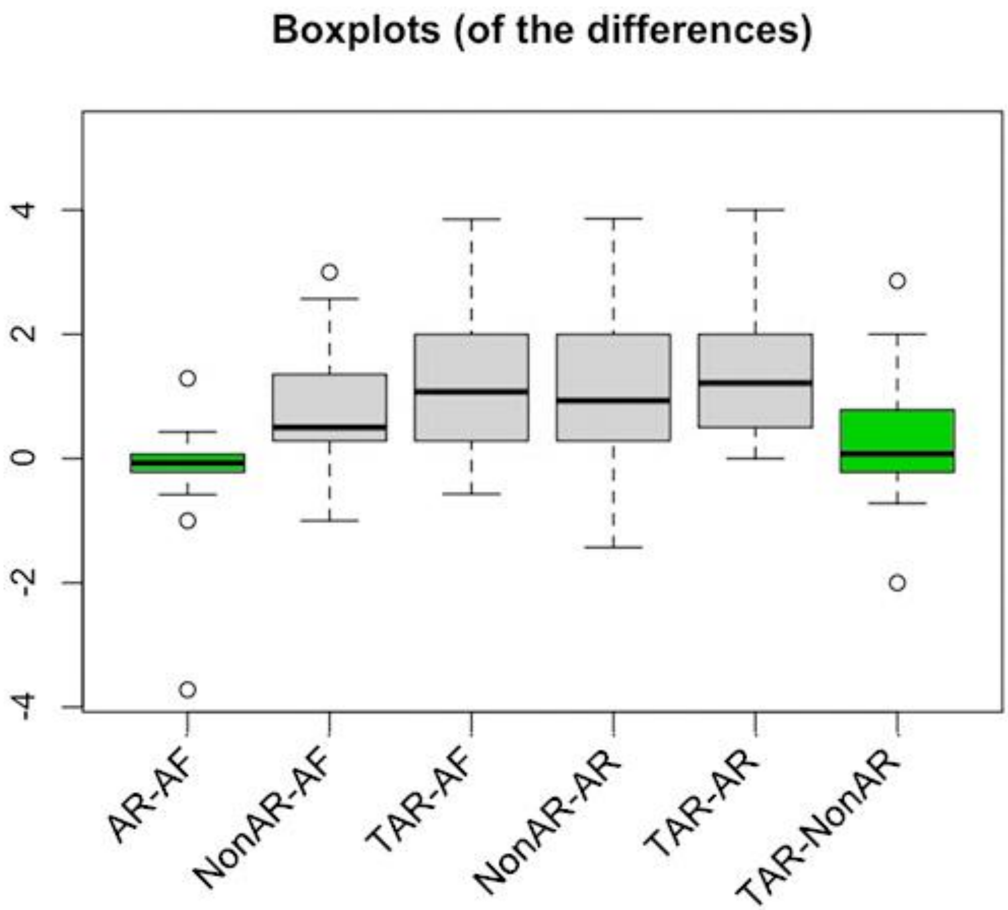


Figure 24. Boxplots of Differences for Ease of Learning

In the boxplot above, the gray ones are those with a small p-value and thus show sizeable differences, while the green ones are those with large p-value and thus show no difference. The results show that there are two pairs of groups that have no differences, which are:

TAR with NonAR

AR with AF

The results show that the user experience of Ease of Learning in AR and AF differs with that in NonAR, and the user experience in Ease of Learning for users in NonAR is the similar with that in TAR.

Satisfaction

Table 17 shows the UEI of Satisfaction from 20 users for four groups.

Table 17. User Experience Index of Satisfaction

Satisfaction				
	NonAR	AR	AF	TAR
#1	6.00	5.71	5.71	6.57
#2	4.71	4.43	4.43	4.86
#3	5.29	2.14	2.29	6.14
#4	5.57	3.43	3.57	5.43
#5	2.71	3.57	3.43	5.57
#6	5.43	5.00	5.00	5.14
#7	4.86	6.29	5.00	6.86
#8	4.00	5.14	5.00	5.86
#9	5.43	4.57	5.00	4.86
#10	6.43	5.71	6.00	5.71
#11	6.29	4.57	4.14	6.14
#12	5.14	4.43	4.57	5.86
#13	5.57	3.71	4.29	5.14
#14	6.00	3.57	4.57	4.00
#15	7.00	3.14	6.86	7.00
#16	6.71	4.00	4.14	6.57
#17	4.57	3.57	3.57	5.57
#18	7.00	6.00	6.00	7.00
#19	3.57	3.29	3.14	4.29
#20	5.71	4.29	4.43	6.43

Friedman's Test results is:

Friedman chi-squared = 30.2656, df = 3, p-value = 1.213e-06

As the p-value is extremely low, the null hypothesis is rejected and it is concluded that there is a significant difference between the four groups.

Since the differences between the four groups are detected, ad-hoc analysis is necessary to determine where the difference is and how much it is. Figure 25 shows a parallel coordinates plot of Satisfaction.

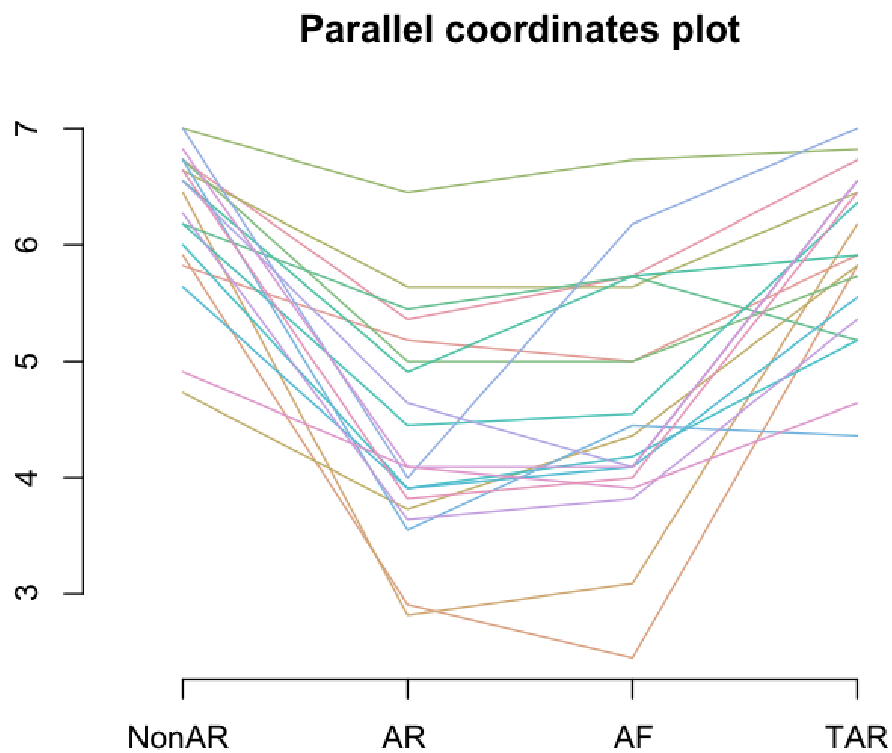


Figure 25. Parallel Coordinates Plot of Satisfaction

This plot shows how the answers varies which came from same person in different groups. One can see that each user rates higher value when it comes to NonAR and TAR, the data of AF is similar to that of the AR, while the data of the NonAR is similar to that of the TAR. Next, Wilcoxon's test was applied to see whether groups have any difference with each another, the results are shown below:

P-value for paired groups:

AR - AF	7.846474e-01
NonAR - AF	1.170734e-02
TAR - AF	1.598410e-03
NonAR - AR	3.683039e-04
TAR - AR	2.349967e-05

TAR - NonAR 9.431413e-01

Figure 26 shows the boxplots for differences of paired comparison.

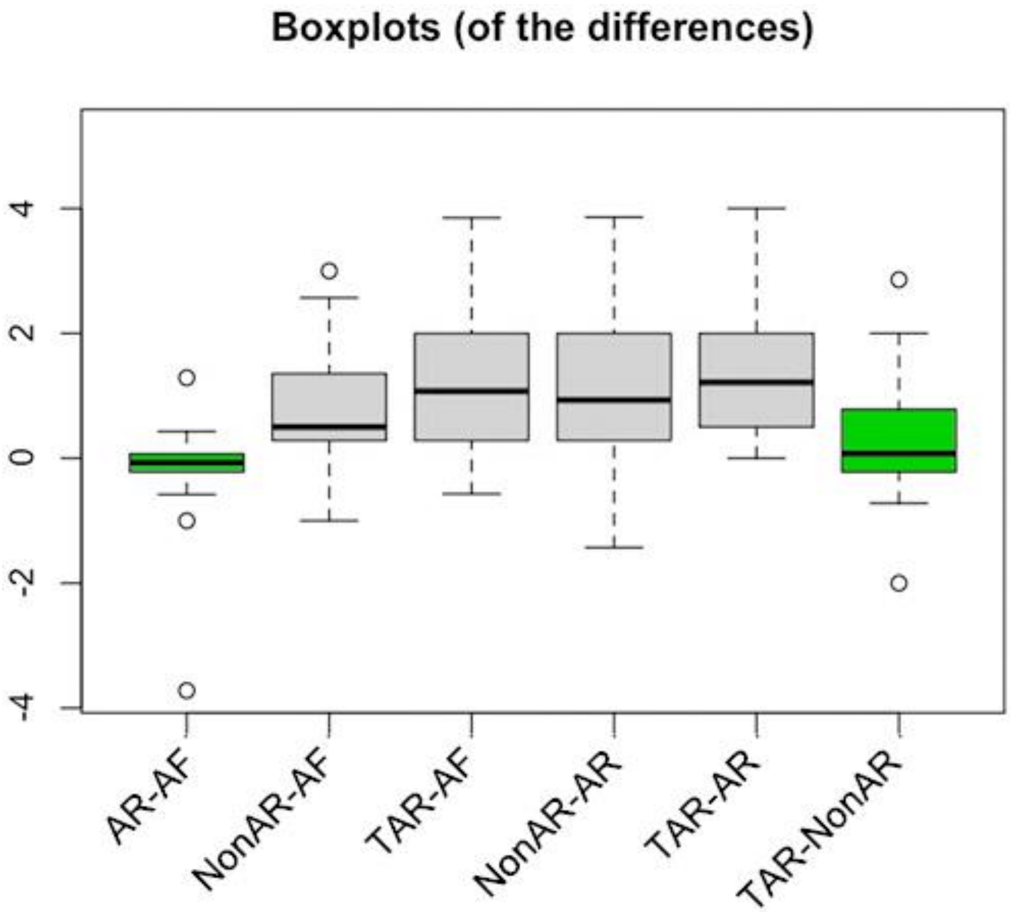


Figure 26. Boxplots of Differences of Satisfaction

In the boxplot above, the gray ones are those with small p-value and thus show significant differences, while the green ones are those with large p-value and thus show no difference. The results show that there are two pairs of groups that have no difference, which are:

TAR with NonAR

AR with AF

The result shows that the user experience of Satisfaction in AR and AF differs with that in NonAR, and the user experience in Satisfaction for users in NonAR is the similar with that in TAR.

Total User Experience

The data was analyzed for each section including Usefulness, Ease of Use, Ease of Learning, and Satisfaction, and all four sections have same conclusions which was user experience in AR and AF is different with NonAR, while user experience in TAR is similar to NonAR.

However, since each of the four sections served as an aspect of the total user experience, it may be advisable to obtain one index for total user experience of each user in each experiment. An index was already calculated for each section as the average of answer rates of all questions in that section, so that now for each user in each group, there are four indexes, which measure the user experience in that section.

The total user experience should be a weighted average of the user experience index for all the four sections, and since in the USE questionnaire, the Usefulness section contains 8 questions, the Ease of Use section contains 11 questions, the Ease of Learning section contains 4 questions, and the Satisfaction section contains 7 questions. Thus, the weight of four sections are found to be 8/30, 11/30, 4/30 and 7/30, and the total user experience should be the weighted sum of user experience index of all four sections. For example, user-1 has his user experience index to be 5, 6, 7, 6 in four sections, so that his total user experience should be:

$$\text{Total User Experience (TUE)} = 5*(8/30) + 6*(11/30) + 7*(4/30) + 6*(7/30) = 5.87$$

In this way, the total user experience of each user in each experiment may be calculated and obtained through the following table:

Table 18. Total User Experience Rate

Total User Experience rate				
	NonAR	AR	AF	TAR
#1	6.302333	5.332333	5.5	6.268667
#2	5.70	5.03	4.90	5.83
#3	5.47	2.63	2.50	5.77
#4	5.80	3.27	3.53	5.67
#5	4.27	3.80	4.23	5.70
#6	6.47	5.63	5.63	6.07
#7	6.10	6.20	6.27	6.57
#8	5.43	5.03	5.07	5.77
#9	5.93	4.90	5.33	5.13
#10	6.57	5.63	5.74	6.03
#11	6.20	4.60	4.43	6.13
#12	5.73	4.30	4.57	5.67
#13	5.57	4.00	4.07	5.14
#14	6.67	3.87	4.57	4.47
#15	7.00	3.97	6.50	7.00
#16	6.60	4.27	4.20	6.57
#17	5.17	4.03	3.90	5.23
#18	6.90	4.83	4.83	6.74
#19	4.67	3.80	3.70	4.57
#20	6.10	4.10	4.20	6.47

Friedman's test was applied on the data and the results are generated as below:

Friedman chi-squared = 42.0761, df = 3, p-value = 3.866e-09

Since the p-value is extremely low, the null hypothesis is rejected and it is concluded that there are at least two groups that show difference.

Next, it is necessary to detect how much the difference between each groups is. Figure 27 shows the parallel coordinates plot.

P-Value for paired comparison:

AR - AF 9.266927e-01

NonAR - AF	1.283184e-04
TAR - AF	1.088669e-04
NonAR - AR	6.266038e-06
TAR - AR	9.782220e-06
TAR - NonAR	1.000000e+00

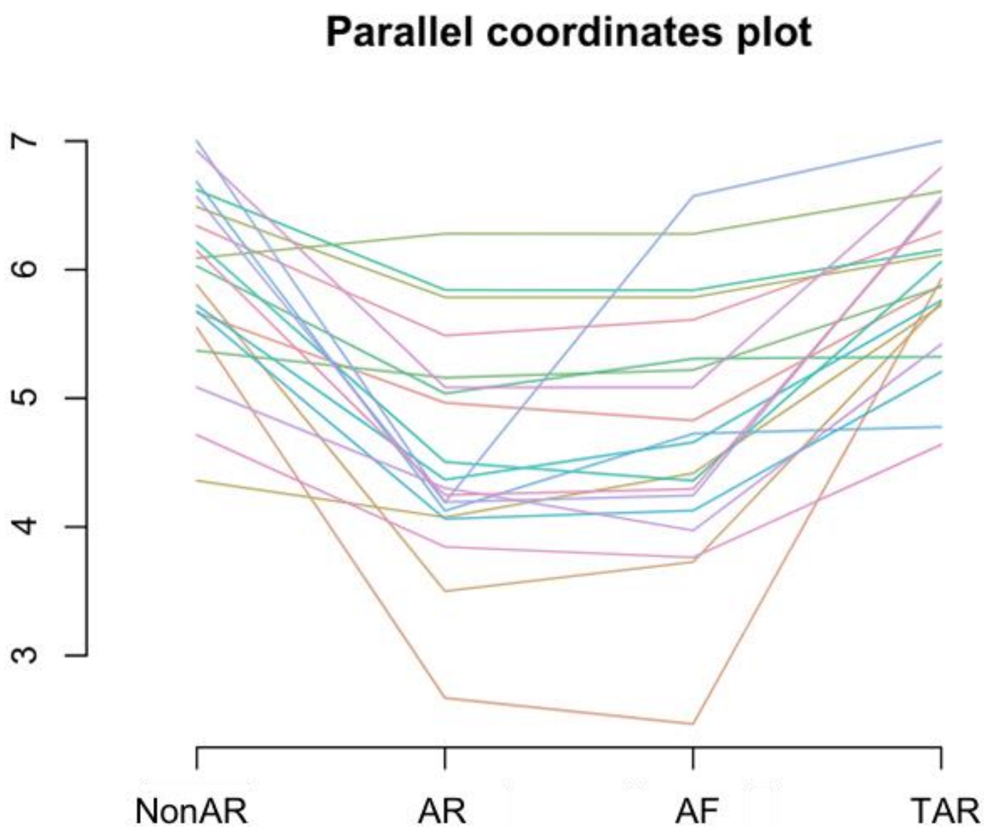


Figure 27. Parallel Coordinates Plot of Total User Experience Rate

This plot shows how the answers varies which came from same person in different groups. One can see that each user rates higher value when it comes to NonAR and TAR, the data of AF is similar to that of AR, while the data of NonAR is similar to that of TAR. Next, Wilcoxon's test was applied to see whether groups have any difference with each another, the results are shown below:

P-Value for paired comparison:

AR - AF	9.266927e-01

NonAR - AF	1.283184e-04
TAR - AF	1.088669e-04
NonAR - AR	6.266038e-06
TAR - AR	9.782220e-06
TAR - NonAR	1.000000e+00

The boxplot for the differences of paired comparison is:

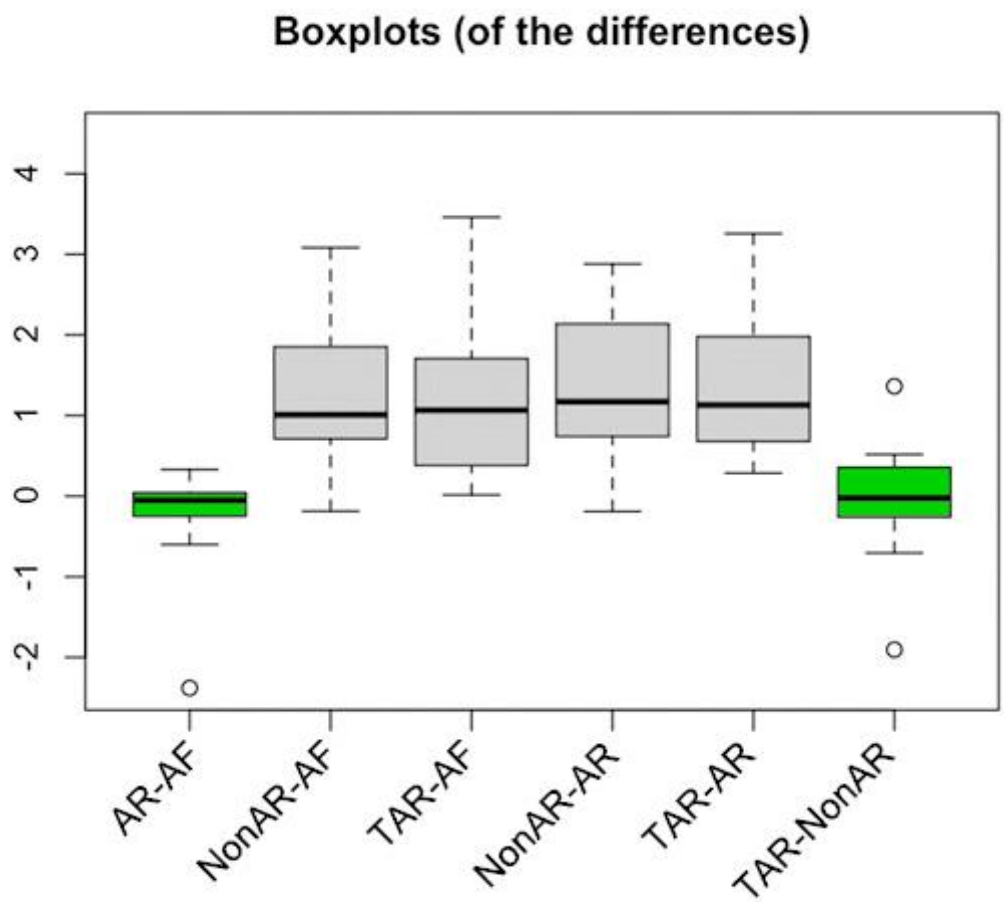


Figure 28. Boxplots of Differences for Total User Experience Rate

In this boxplot, the green ones are the groups that shows significant difference, while the grey are those who do not show significant difference. There are only AR – AF and NonAR – TAR in gray, which means the total user experience in AR and AF is similar and the total user experience in TAR and NonAR is similar.

Result

In the data analysis, statistical tests on each of the four sections were conducted: Usefulness, Ease of Use, Ease of Learning, and User Satisfaction. By conducting Friedman's test on the data of four sections, the p-values of Friedman's test were found to be extremely low on all the four sections, which meant that at least two of the four groups: NonAR, AR, AF and TAR showed a difference in each of the four sections. In order to detect which of pair of the four groups showed a difference, paired Wilcoxon's test was further conducted on the data. Since there are four groups, there are 6 pairs of data. By conducting the Wilcoxon's test of each of the 6 pairs for each section, results found for all the four sections are the same. The p-value of NonAR with TAR, and the p-value of AR and AF are below 0.05, while the p-values for the other paired groups are larger than 0.05. The results showed that for each of the four sections, paired groups of NonAR and TAR, AR and AF do not have statistically significant differences, while all the other paired groups are showing significant difference. Statistical tests were also conducted on the total user experience index as a whole to see whether the user experience between groups was different. The same results as that for each of the sections was found, the p-value of the Friedman's test was extremely low, which showed at least two groups were different in user experience, and by conducting the Wilcoxon's test, one also found that NonAR and TAR, AR and AF did not show a statistically significant difference.

CHAPTER 6

CONCLUSION

Reflection on the Results

The usability testing on Sunbeam heater SFH5264MW was done using four different methods. 20 users tested the product using all the four representations. First was the traditional usability testing methodology, using the actual product and completing certain tasks and then providing their feedback using a questionnaire. The second method involved using augmented reality for usability testing, all the users used this method for testing using a marker card. The third method involved not only a marker card, but also a 3D printed model. The forth method used three marker cards combined with a 3D printed model to achieve Tangible AR.

Friedman's Test was used to detect difference in user experience between the experimental groups and Wilcoxon's Test was used to indicate differences between pairs of groups. For Usefulness, Ease of Use, Ease of Learning, Satisfaction and also the total user experience, AR and AF were statistically significantly different with NonAR, but TAR did not have a statistically significant difference with NonAR. The results of the data analysis strongly supported the assertion that TAR was effective in simulating the real product in the user study.

As was shown from the results of data analysis, TAR did a better job than AR and AF in taking the place of real product in usability testing, it made sense for the following reasons. Firstly, among the three prototyping methods, TAR was the only one that kept the natural interactions between participants and products, the simulation of physical interactions was an important part of user experience in this study. Secondly, the less workload also made TAR a better approach in simulating the real product, the only big difference between real product and TAR representation was the augmented view shown on screen. TAR enabled participants to interact with the prototype without taking effort to learn how to use, and they only need to get used to the augmented view shown on screen while all the other interactions are the same as the real product.

The application of TAR in usability testing was quite promising. As will be shown in later section, using TAR methods instead of traditional prototyping methods came with significant save of time and cost and thus potentially made great contributions to new product development. It also provides great flexibility in design process which will improve the efficiency of research, enable designers and engineers to concentrate more on developing and get free from tedious prototyping and testing process.

Comparison with Previous Study

This study was extended from regular AR method to Tangible AR from a previous study. The previous study focused on another product which was the Sony Walkman NWZ-E463 (Figure 29
) and implemented AR and AF technology to set up two experimental groups. Then the previous study compared the usability evaluation with AR and AF groups. The results between the actual product evaluations and augmented reality evaluation appear no different in the study which means AR and AF method can be a valid method of collecting usability data in the process. [20]

Figure 29. Sony Walkman NWZ-E463 Used in Sanchit's Study

In this study, a different product was chosen which was the Sunbeam Heater SFH5264MW and there was one more experimental group using TAR technology. The results between the actual product evaluations and augmented reality evaluation appear different in this study which shows differences in AR and AF groups and no difference in the TAR group.

The differences between the previous study and this study may be caused by the different features of two selected products. The Sony Walkman NWZ-E463 used in previous study had several physical clickable buttons on it, but the Sunbeam heater SFH5264MW used in this study only had two rotatable knobs. So, when designing the conversion of the user experience from natural interactions of each product to touch screen interactions using the same methods in AR and AF groups, the previous study using Sony Mp3 has only explored the use of touchscreen for the main interaction and the buttons did not have any physical feedback. The buttons were simple hotspots on a webpage that was laid on top of the physical model in the AR environment so that there were only minute differences between the natural interactions and the touch screen interactions. For the Sunbeam heater used in this study, the prototype used in the AR and

AF groups enabled users to rotate the knob by “Tap and drag” following a curve, which was different from how people actually turn a knob.

Another achievement that has been added into this experiment is simulating the real knobs in the TAR experiment. Instead of translating natural interactions from clickable buttons and rotatable knobs into touchable screen, TAR group simulates the natural interaction which allowed users have almost the some experience with the real product.

However the difference between the results in the previous study and this study provide supplements to this study, which showed that AR/AF could be a substitute for a real product having clickable buttons, but was not suitable for other types of buttons.

Significance of Time and Cost Efficiency

Companies have been spending a lot of time and money in developing new products. Big companies such as Microsoft spend nearly 10 billion US dollars every year on research and development [21]. Product development and testing are the most significant part of new product development cycle (NPD). They consume 54% of the resources [22]. Applying AR and TAR in Product development and testing could save a lot of time and cost, thus making significant improvement on new product development.

Using the traditional method, it takes about 18,000 dollars to make just one main body shell of the heater, according to a quote from an online Cost Estimator.[23] Figure 30 shows all the input details. The expense is quite high because of the tooling fee for injection modeling. In addition, circuits design and structure design inside the knobs also cost a lot. As a result, the estimated total cost of prototyping of one generation could be as high as 30,000 dollars using no less than one month.

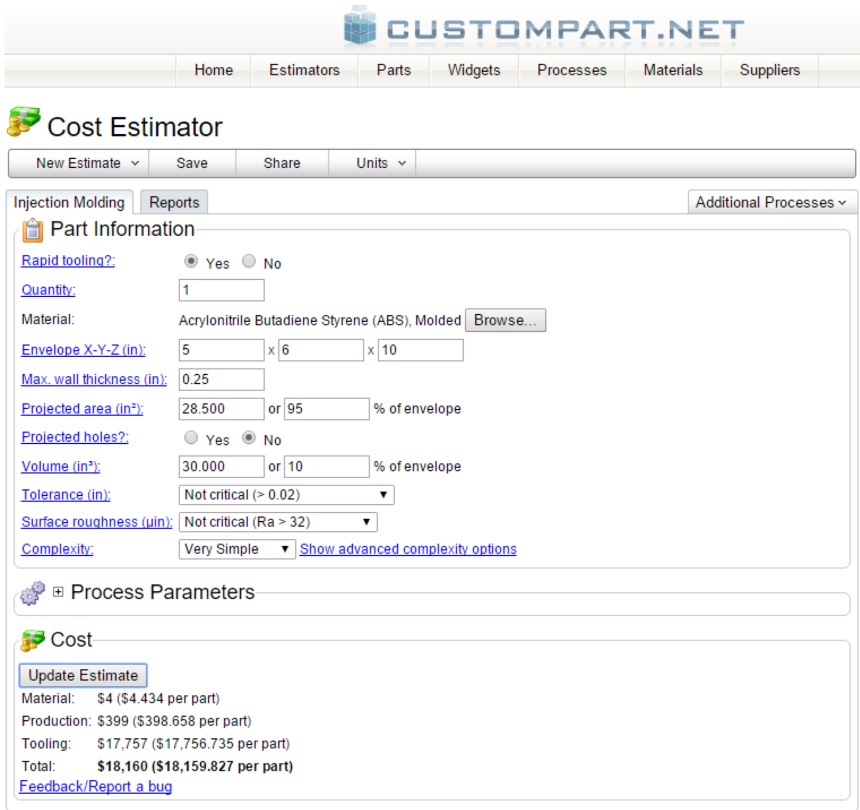


Figure 30. Estimated Cost of Modeling the Shell By Injection Modeling[23]

In contrast with the traditional method, the AR method does not require real materials. This method only costs 300 dollars to hire a project-based engineer for 2 days to program the prototype.

The advantage of the AF method over the AR method is the 3D printed model. This method costs additionally 400 dollars and 2 days for the 3D printed model. Thus, the estimated total cost of this method is 700 dollars and 4 days.

The TAR method used in this study also includes a 3D printed model. Consequently, this method costs 400 dollars for the 3D printed model and another 2000 dollars for hiring an engineer to program the prototype. The prototype and the model could be completed within 5 days.

Table 19. Time and Cost Comparison of Different Method

	Traditional Method	AR Method	AF Method	TAR Method
Costs	≈30,000	300	700	2,400
Time	30 days	2 days	4 days	5 days

By simply comparing the time and cost of the four groups in this experiments, it is very clear that the AR, AF and TAR methods have significant advantage over the traditional method. The AR method takes only 1% of the total expense and 6.7% of the time of traditional method. The TAR method costs 8% of dollars and 16.7% of time of the traditional method. The new methods significantly improve the efficiency and reduce the cost of the prototyping process.

Furthermore, there are other factors that need to be considered. First of all, the product development process is iterative. The prototypes are being revised all the time after receiving analyzing user experience and receiving feedbacks. With the traditional method, it is impractical to revise the prototype whenever it has to be rebuilt due to little change of product design. The cost and time would grow linearly as the iterative design process goes on. In contrast, using the AR, AF and TAR methods, arbitrary update of the design could be achieved through code update without changing the real prototypes. Thus, the cost and time grow slowly as the design process goes on.

Secondly, prototypes need to be highly available for further analysis and improvement. In the traditional method, the prototypes are built in physical form with all functions implemented. The prototypes can be accessed by only few designers at a time for test purposes. However in the AR, AF and TAR groups, once the programming is finished, the tools such as apps and markers are always easily accessible online.

There are also some challenges in applying AR, AF and TAR is usability testing. For example, in order to let participants get more familiar with those technologies, training is conducted before experiment in this study. However, such challenges may not exist in the future. AR related technologies have become increasingly popular recently. People will be more and more familiar with such technologies and will not need to get training before usability testing. The cost for hiring specialized engineers for programming may also be eliminated, if in the future some easy-for-use TAR development tools or software are available for designers.

In summary, the TAR method presented in this study introduces a new form of development process. It makes it possible to achieve significant time and cost efficiency in development process.

Limitation of the Study

While the Tangible AR method produced successful results, it had some limitations. The results of TAR method potentially only could apply to products with large rotatable knobs with a flat surface. The described AR method might not be a good approach for products which do not have physical knobs. The screen size of the display is a limitation. In this study, using a tablet to view augmented reality interactions, somewhat limits the size of the product that can be tested, but this problem could be solved by using a touchscreen monitor and desktop system.

For AR and AF groups, the technology limitation cannot be ignored. While the prototypes for AR and AF groups were programmed in Unity3D, the web prototype which shows the front view of the product with interaction functions cannot perform as smooth as in augmented view even every possible improvement of the web program has been tried. However, it is believed that the current limitation will be solved with the continuous progress of technology.

The users chosen for testing in this study were all college students, which was not a diverse demographic. This study might have produced totally different results with older adults assuming that they might not be as familiar with using a tablet as college students.

Future Work

In this study, representations used in AR, AF, TAR groups were taken as prototypes with different fidelities which utilized augmented reality related technology. Within the TAR technology, there are different prototyping methods which can influence the fidelity of prototypes and thus provide different user experience. A more immersing TAR technology was already developed which used the 3D printed model as marker by simply scanning it with 3D scanner, this technology enables users to watch the 3D printed model from all directions with augmented effects. And the hand visibility correction techniques used in Woohun Lee and Jun Park's study[5] is also a good method which provides a better user experience more similar to real product's. The two technologies mentioned above both could be used in future study in order to achieve a higher fidelity of prototype compared to this study. For the next step, it would be great to explore how usability results change along with TAR prototyping methods with different fidelities.

Secondly, it would be a good idea to apply TAR prototyping methods to other kinds of products with different features such as clickable button, switchable button, spring switch, etc. These three product features are suggested because they are all commonly used in product design and it is helpful to investigate whether TAR is a good prototyping method which fits other design features in usability testing.

APPENDIX A

CONSENT DOCUMENT FOR ENROLLING ADULT PARTICIPANTS IN A RESEARCH STUDY

Georgia Institute of Technology
Project Title: Applying Tangible Augmented Reality in Usability Evaluation

Investigators: Dr. Young-Mi Choi, Xiaotian Zhang
Protocol and Consent Title: Main 11/26/14

You are being asked to be a volunteer in a research study.

Purpose:
The purpose of this study is to provide usability evaluation with a way, which allows users to come up with useful feedbacks more efficiently by applying Augmented Reality related technologies. Designers can leverage our gain from this research to improve their design in the Usability Evaluation Process. We expect to enroll 20 people in this study.

Exclusion/Inclusion Criteria:
All participants' age must range from 18 to 69 and they must not be upper limb disabled and must have normal or corrected to normal vision.

Procedures:
If you agree to participate in this study, your part will involve one visit. The total amount of time you will be in the laboratory is no more than 1 hour. Before all the procedures, you will be randomly assigned to one group. The procedures to be followed include: (1) a briefing on the experiment; (2) completing practice and test trials successively with the help of each of four different design representatives: a. AR group: participants will be able to interact with the prototype via a display screen. b. AF group-I: participants will be able to interact with the prototype by touching a physical model but watching the augmented view via the display. c. AF group-II: main performance will be the same with previous group, but with hand visibility correction. d. Control group: participants will evaluate the actual product. (During practice and test trials, you will be given specific steps to follow); (3) ranking of the relative importance of workload and system usability factors for each of the four systems; (4) debriefing. With your permission, we will videotape your actions as you perform the task steps. Only the researcher and principal investigator will have access to the footage, and it will not be presented or published. You may stop the experiment at any time and for any reason.

Risks or Discomforts:
The risks associated with participation in this study are minimal. They include potential inadaptation on AR representatives that may cause

APPROVED *Consent Form Approved by Georgia Tech IRB: October 19, 2015 - Indefinite* slight dizziness from viewing the interactive displays. To minimize the risks, there will be a short break between each phase. Users can also stop this experiment or ask for a break at any time, for any reason.

Benefits:
You are not likely to benefit in any way from joining this study. We hope that what we learn will someday help designers and developers make decisions on future product development and thus improve the efficiency in field service and related industries.

Compensation to You:

There is no compensation for participation.

Confidentiality:
The following procedures will be followed to keep your personal information confidential in this study: The data collected about you will be kept private to the extent allowed by law. To protect your privacy, your written and video records will be kept in locked in a file cabinet in a private office. Electronic records will be kept in a pass-coded file on a computer in a private office. Only the researcher and principal investigator will have access and knowledge of the information contained within the data. We will use a code rather than your name to identify study records. The code will be kept on a separate password protected computer from the data in a private office. Your name and any other fact that might point to you will not appear when results of this study are presented or published. The video footage will be erased after all the necessary information is collected and analyzed. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records.

Costs to You:
There are no costs to you, other than your time, for being in this study.

In Case of Injury/Harm:
If you are injured as a result of being in this study, please contact Principal Investigator, Dr. Young-Mi Choi, at telephone (404) 277-2748. Neither the Principal Investigator nor Georgia Institute of Technology has made provision for payment of costs associated with any injury resulting from participation in this study.

Participant Rights:
- Your participation in this study is voluntary. You do not have to be in this study if you don't want to be.
- You have the right to change your mind and leave the study at any time without giving any reason and without penalty.

APPROVED *Consent Form Approved by Georgia Tech IRB: October 19, 2015 - Indefinite*

- Any new information that may make you change your mind about being in this study will be given to you.
- You will be given a copy of this consent form to keep.
- You do not waive any of your legal rights by signing this consent form.

Questions about the Study:
If you have any questions about the study, you may contact Dr. Young-Mi Choi, at telephone (404) 277-2748 or christina.choi@gatech.edu.

Questions about Your Rights as a Research Participant:
"If you have any questions about your rights as a research participant, you may contact

Ms. Melanie Clark, Georgia Institute of Technology
Office of Research Integrity Assurance, at (404) 894-6942."
[or]
"Ms. Kelly Winn, Georgia Institute of Technology
Office of Research Integrity Assurance, at (404) 385- 2175."

If you sign below, it means that you have read (or have had read to you) the information given in this consent form, and you would like to be a volunteer in this study.

____
Participant Name (printed)

____ ____________
Participant Signature Date

____ ____________
Signature of Person Obtaining Consent Date

APPROVED Consent Form Approved by Georgia Tech IRB: October 19, 2015 - Indefinite

APPENDIX B

STUDY SCRIPT

Double Check All the Materials

(All the materials: pre-test questionnaire, 2 copy of consent form; order of 4 experiment, order of 8 tasks, 4 questionnaires.)

Explaining the Study

“The purpose of this study is to explore if augmented reality can be beneficial in the usability testing process. Augmented Reality---you can also call it as “AR” for short---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. Now I will set up a demo as a warming up to let you experience AR a little bit more in advance.”

(Demo by using Augment App: open the Augment App, and choose the beverage can model, and then create a marker for it, then let participant see the screen)

“Now you can see the real environment is showing on the screen. When you put this magazine in front of the camera, you will see a beverage can on top of the magazine. The magazine is called ‘Marker’ which is an element of real-world environment and the beverage can is the augmented view of this magazine, which means you can see it from the screen but you cannot touch it in real world.”

(Let participant change the position of the tablet or the marker to experience AR technology.)

“Do you have a sense of AR after this short demo?”

“This study will specifically explore if AR can be used to get better usability input which specifically focused on physical knobs. Pay attention that the knob I just

mentioned is not like clickable buttons such as a home button on an iPhone or a touch screen. The physical knob in this experiment must be rotatable."
(Show the knobs of the product to participants.)

"The results of this research might provide evidence that AR can be used for usability testing to improve the time efficiency in prototyping."

Overview of To Do List

"If you decide to participate this study, your part will involve four experiments up to one hour. In first part of each experiment, you will be asked to finish 8 simple tasks on four different prototypes that I will provide you. Those tasks will be to interact with some basic features of the product, for example, Turn the mode control knob to "fan-only". This process will take up to 10 minutes. The second part of each experiment would be answering a user questionnaire about usability satisfaction of each prototype, which will take about 10 minutes to fill out. Remember, you may stop at any time."

"There is no compensation for participation."

Handing Out Consent Form

"There are two copies of the consent form, you need to carefully read it and sign one of them for me in order to place it on file. You can keep another one for yourself as record."

Fill Out Demographic Questionnaire

"This is a pre-test questionnaire. The purpose of this questionnaire is to gather your basic information. You need to fill out all the 3 questions. It will take less than 2 minutes to complete."

For Experiment Using Actual Product (Non-AR)

Introduction:

"I will set up the Sunbeam heater for you and you will be asked to finish 8 tasks with it."

(Set up product for experiment)

"This is the heater you will be asked to interact later, the two main parts of this heater are the two knobs. The upper knob is called Thermostat Control, which controls the level of wind. The scale over this knob means four different wind levels from 1 to 4. You can hold this knob and then turn it like this. The lower knob is called Mode Control, which controls four modes of this heater. The four symbols here mean four different modes: Off, Fan-Only, Low-Heat, High-Heat. You can hold this knob and then turn it the same way as the upper knob, but you will feel the different knob resistance."

(Show how to interact with the product to participant: for the upper knob, hold the knob and then turn to the right slowly in order to let the participant see clearly; for the lower knob, do the same way)

"Now you can play with this product to get familiar with it and make sure you clearly know all the functions of this heater. You can ask me if you have any questions."

(Participant try the product)

After Completing the Tasks:

(Make sure all the photos are done properly and double check the questionnaire marked in the right way)

"I am handing you a user questionnaire which contains several questions. These questions are about the usability satisfaction of the product you just used. For each of the question, you are asked to rate your feelings towards one specific user experience related questions on a 7-point scale from strongly agree through strongly disagree. Please try to answer them based on your experience of using the product. Take as much time you need to answer the questions."

For Experiment Using Marker Card (AR)

Introduction:

"I will set up a marker card with some graphic printed on it."

(Set up prototype for experiment)

"Now you will look at the model through the screen. You can change the position of the tablet or the marker to see different angles of this model. You have as much time you like to observe the AR model on your screen. Then, you will use touch interface to finish 8 tasks on two knobs. The upper knob is called Thermostat Control, which controls the level of wind. The scale over this knob means four different wind levels from 1 to 4. You can keep touching and switch the angle to rotate the knob, but keep in mind: your finger must keep in this circle. (Point the "circle" to show the active area to the participant.) The lower knob is called Mode Control, which controls four mode of this heater. The four symbols here mean four different modes: Off, Fan-Only, Low-Heat, High-Heat. You can control it by a swipe on the screen. Keep in mind: your finger must keep in this circle. (Point the "circle" to show the active area to the participant.)"

(Show how to interact with the prototype to participant: for the upper knob, put one finger on the screen and then touch the little pointer on the upper knob and drag it to switch the knob; for the lower knob, just swipe a curve on the "knob" to switch mode)

"Now you can play with this prototype to get familiar with it and make sure you clearly know all the functions of this prototype, especially for the interface. You can ask me if you have any questions"

(Participant try the prototypes)

After Completing the Tasks:

(Make sure all the photos are done properly and double check the questionnaire marked in right way)

“I am handing you a user questionnaire which contains several questions. These questions are about the usability satisfaction of the product you just used. For each of the questions, you are asked to rate your feelings towards one specific user experience related questions on a 7-point scale from strongly agree through strongly disagree. Please try to answer them based on your experience of using the product. Take as much time you need to answer the questions.”

For Experiment Using 3D Printed Model (AF)

Introduction:

“I will set up a marker card with some graphic printed on it and a 3D printed model.”

(Set up prototype for experiment)

“You will look at the 3D printed model through the screen and you will see an augmented view of it on the tablet. You can change the position of the tablet or the 3D printed model to see a different angle of this model. You have as much time you like to observe the model on your screen. Then, you will use touch interface to finish 8 tasks on two knobs. The upper knob is called Thermostat Control, which controls the level of wind, you can keep touching and switch the angle to rotate the knob. The scale over this knob means four different wind levels from 1 to 4. You can keep touching and switch the angle to rotate the knob, but keep in mind: your finger must stay in this circle. (Point the “circle” to show the active area to the participant.) The lower knob is called Mode Control, which controls four mode of this heater the four symbols here mean four different modes: Off, Fan-Only, Low-Heat, High-Heat. You can control it by a swipe on the screen. Keep in mind: your finger must stay in this circle. (Point the “circle” to show the active area to the participant.)”

(Show how to interact with the prototype to participant: for the upper knob, put one finger on the screen and then touch the little pointer on the upper knob and drag it to switch the knob; for the lower knob, just swipe a curve on the "knob" to switch mode)

"Now you can play with this prototype to get familiar with it and make sure you clearly know all the functions of this prototype, especially for the interface. You can ask me if you have any questions"

(Participant try the prototype)

After Completing the Tasks:

(Make sure all the photos are done properly and double check the questionnaire marked in right way)

"I am handing you a user questionnaire which contains several questions. These questions are about the usability satisfaction of the product you just used. For each of the question, you are asked to rate your feelings towards one specific user experience related questions on a 7-point scale from strongly agree through strongly disagree. Please try to answer them based on your experience of using the product. Take as much time you need to answer the questions."

For Experiment Using 3D Printed Model (TAR)

Introduction:

"I will set up a 3D printed model with three marker cards with some graphics printed on it."

(Set up prototype for experiment)

"You will look at the 3D printed model through screen. You can change the position of the tablet or the 3D printed model to see a different angle of this model. You have as much time you like to observe the model on your screen. Then, you will be asked to finish 8 tasks with two knobs. The upper knob is called Thermostat Control, which controls the level of wind. The scale over this knob means four different wind levels from

1 to 4. You need to rotate the knob on 3D printed model to trigger it. The lower knob is called Mode Control, which controls four modes of this heater. The four symbols here mean four different modes: Off, Fan-Only, Low-Heat, High-Heat. You also can control it by rotating the 3D printed knob. The two check images on the right side of each knob mean that the knob is active, if the check image disappears, you need to change the 3D printed model or the tablet to active it again. Keep in mind: you just see the augmented view from the screen, but cannot touch the screen at this experiment, and keep all three markers visible to the camera."

(Show how to interact with the prototype to participant: as the same with Non-AR group, just keep the markers on the knobs always visible to the camera.)

"Now you can play with this prototype to get familiar with it and make sure you clearly know all the functions of this prototype, especially for the interface. You can ask me if you have any questions"

(Participant try the prototype)

After Completing the Tasks:

(Make sure all the photos are done properly and double check the questionnaire marked in right way)

"I am handing you a user questionnaire, which contains several questions. These questions are about the usability satisfaction of the product you just used. For each of the question, you are asked to rate your feelings towards one specific user experience related questions on a 7-point scale from strongly agree through strongly disagree. Please try to answer them based on your experience of using the product. Take as much time you need to answer the questions."

All the Tasks

For different participants, there will be 20 different sequences of the tasks, but they only have little difference by order. Here is one example of a series of tasks:

Task 1:

“Imagine you are feeling cold and you want to use your Sunbeam heater to get warm. First, turn the Mode Control to ‘Low-heat’.”

Task 2:

“Then, turn the upper knob to Level 4.”

Task 3:

“Then, turn the upper knob to Level 1.”

Task 4:

“Then, turn the lower knob to ‘Fan-only’.”

Task 5:

“Then, turn the lower knob to ‘High-heat’.”

Task 6:

“Then, turn the upper knob to Level 3.”

Task 7:

“Then, turn the upper knob to Level 2.”

Task 8:

“Then, turn the lower knob to ‘Off’.”

End of All the Experiments

“Thank you so much for your participation and valuable feedback. If you have any questions you can contact me. Have a good day!”

APPENDIX C

USER TASKS OF PROJECT EXPERIMENT

i.

1. Turn the Mode Control (Lower knob) to “Low-heat”
2. Turn the Thermostat Control (Upper knob) to Level 4
3. Turn the Thermostat Control (Upper knob) to Level 1
4. Turn the Mode Control (Lower knob) to “Fan-only”
5. Turn the Mode Control (Lower knob) to “High-heat”
6. Turn the Thermostat Control (Upper knob) to Level 3
7. Turn the Thermostat Control (Upper knob) to Level 2
8. Turn the Mode Control (Lower knob) to “Off”

ii.

1. Turn the Mode Control (Lower knob) to “Fan-only”
2. Turn the Thermostat Control (Upper knob) to Level 4
3. Turn the Thermostat Control (Upper knob) to Level 3
4. Turn the Mode Control (Lower knob) to “High-heat”
5. Turn the Thermostat Control (Upper knob) to Level 1
6. Turn the Mode Control (Lower knob) to “Low-heat”
7. Turn the Thermostat Control (Upper knob) to Level 2
8. Turn the Mode Control (Lower knob) to “Off”

iii.

1. Turn the Mode Control (Lower knob) to “High-heat”
2. Turn the Thermostat Control (Upper knob) to Level 2

3. Turn the Mode Control (Lower knob) to “Fan-only”
4. Turn the Mode Control (Lower knob) to “Low-heat”
5. Turn the Thermostat Control (Upper knob) to Level 4
6. Turn the Thermostat Control (Upper knob) to Level 1
7. Turn the Thermostat Control (Upper knob) to Level 3
8. Turn the Mode Control (Lower knob) to “Off”

iv.

1. Turn the Mode Control (Lower knob) to “High-heat”
2. Turn the Thermostat Control (Upper knob) to Level 4
3. Turn the Mode Control (Lower knob) to “Fan-only”
4. Turn the Thermostat Control (Upper knob) to Level 1
5. Turn the Mode Control (Lower knob) to “Low-heat”
6. Turn the Thermostat Control (Upper knob) to Level 3
7. Turn the Thermostat Control (Upper knob) to Level 2
8. Turn the Mode Control (Lower knob) to “Off”

APPENDIX D

PRE-TEST QUESTIONNAIRE

Participant No: ______________________

Introduction:
※The purpose of this questionnaire is to gather the subjects' basic information.
※This survey has 3 questions and will take less than 2 minutes to complete.

1. In what year were you born?

____

2. What's your gender?

☐ Male
☐ Female

3. What's your race?

☐ American Indian or Alaska Native
☐ Asian
☐ Black or African American
☐ Native Hawaiian or Other Pacific Islander
☐ White

APPENDIX E

USE QUESTIONNAIRE

Questionnaire for User Satisfaction of Sunbeam Heater

Participant No._______________

Group: Non-AR AR AF TAR

※ Try to respond to all the items.
Please rate your agreements with these statements.

USEFULNESS			1	2	3	4	5	6	7	
1	It helps me be more effective.	strongly disagree	○	○	○	○	○	○	○	strongly agree
2	It helps me be more productive.	strongly disagree	○	○	○	○	○	○	○	strongly agree
3	It is useful.	strongly disagree	○	○	○	○	○	○	○	strongly agree
4	It gives me more control over the activities in my life.	strongly disagree	○	○	○	○	○	○	○	strongly agree
5	It makes the things I want to accomplish easier to get done.	strongly disagree	○	○	○	○	○	○	○	strongly agree
6	It saves me time when I use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
7	It meets my needs.	strongly disagree	○	○	○	○	○	○	○	strongly agree
8	It does everything I would expect it to do.	strongly disagree	○	○	○	○	○	○	○	strongly agree

EASE OF USE			1	2	3	4	5	6	7	
9	It is easy to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree
10	It is simple to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree
11	It is user friendly.	strongly disagree	○	○	○	○	○	○	○	strongly agree
12	It requires the fewest steps possible to accomplish what I want to do with it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
13	It is flexible.	strongly disagree	○	○	○	○	○	○	○	strongly agree
14	Using it is effortless.	strongly disagree	○	○	○	○	○	○	○	strongly agree
15	I can use it without written instructions.	strongly disagree	○	○	○	○	○	○	○	strongly agree
16	I don't notice any inconsistencies as I use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
17	Both occasional and regular users would like it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
18	I can recover from mistakes quickly and easily.	strongly disagree	○	○	○	○	○	○	○	strongly agree
19	I can use it successfully every time.	strongly disagree	○	○	○	○	○	○	○	strongly agree

EASE OF LEARNING			1	2	3	4	5	6	7	
20	I learned to use it quickly.	strongly disagree	○	○	○	○	○	○	○	strongly agree
21	I easily remember how to use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
22	It is easy to learn to use it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
23	I quickly became skillful with it.	strongly disagree	○	○	○	○	○	○	○	strongly agree

SATISFACTION			1	2	3	4	5	6	7	
24	I am satisfied with it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
25	I would recommend it to a friend.	strongly disagree	○	○	○	○	○	○	○	strongly agree
26	It is fun to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree
27	It works the way I want it to work.	strongly disagree	○	○	○	○	○	○	○	strongly agree
28	It is wonderful.	strongly disagree	○	○	○	○	○	○	○	strongly agree
29	I feel I need to have it.	strongly disagree	○	○	○	○	○	○	○	strongly agree
30	It is pleasant to use.	strongly disagree	○	○	○	○	○	○	○	strongly agree

APPENDIX F

RAW DATA

NonAR																				
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20
Q1	7	6	4	4	4	7	4	4	6	7	6	6	6	7	7	7	4	7	5	6
Q2	5	6	4	4	4	7	4	4	7	7	6	6	4	7	7	6	4	7	5	6
Q3	7	7	4	4	5	7	4	5	7	7	6	6	5	7	7	7	4	7	5	6
Q4	4	7	4	4	4	6	7	5	5	5	6	6	4	7	7	6	4	7	4	4
Q5	5	7	4	4	4	7	7	4	5	6	7	4	5	7	7	7	4	6	5	4
Q6	4	5	4	2	5	7	4	4	4	6	5	6	5	7	7	7	4	7	5	4
Q7	6	6	6	7	5	7	7	5	5	7	6	6	5	7	7	7	4	7	5	5
Q8	7	7	6	7	3	7	7	5	5	7	7	6	6	7	7	7	4	7	5	7
Q9	7	7	6	7	4	7	7	6	6	7	5	6	6	7	7	6	7	7	6	7
Q10	7	6	6	7	4	7	7	7	5	7	6	6	6	7	7	7	7	7	6	7
Q11	7	6	6	7	4	7	7	7	6	7	7	6	5	7	7	7	7	7	6	7
Q12	7	5	5	4	4	7	7	7	6	7	6	6	6	7	7	7	6	7	3	7
Q13	6	5	5	5	5	3	7	7	6	6	7	6	4	4	7	7	6	6	5	4
Q14	7	6	6	6	5	7	7	7	6	6	6	6	5	7	7	7	7	6	5	7
Q15	7	7	7	7	6	7	7	7	6	7	5	6	6	7	7	6	7	7	3	7
Q16	7	6	6	7	6	7	7	7	6	5	6	6	6	7	7	6	7	7	5	7
Q17	5	5	6	7	4	7	7	6	7	6	6	6	6	7	7	7	5	7	5	6
Q18	7	5	6	7	5	7	7	6	7	7	7	6	6	7	7	6	4	7	5	7
Q19	7	6	6	7	5	7	7	7	7	7	7	6	6	7	7	6	6	7	5	7
Q20	7	6	7	7	6	7	7	6	7	7	6	6	7	7	7	7	6	7	6	7
Q21	7	6	7	7	7	7	7	6	7	7	5	6	7	7	7	6	6	7	6	7
Q22	7	7	7	7	7	7	7	7	7	7	5	6	7	7	7	6	7	7	5	7
Q23	7	6	7	7	6	7	7	7	7	7	7	6	7	7	7	7	7	7	5	7
Q24	6	5	7	7	4	7	7	6	7	7	6	6	7	7	7	7	4	7	5	6
Q25	6	5	6	7	2	6	7	5	6	7	6	6	6	7	7	6	4	7	5	5
Q26	6	4	4	4	2	5	3	2	4	5	7	6	5	5	7	7	4	7	3	4
Q27	7	5	6	7	3	7	7	5	6	7	7	6	6	7	7	7	7	7	3	7
Q28	6	5	4	4	2	5	4	2	5	7	6	4	5	5	7	6	4	7	3	6
Q29	5	4	5	4	2	4	1	4	5	6	5	4	5	5	7	7	5	7	3	6
Q30	6	5	5	6	4	4	5	4	5	6	7	4	5	6	7	7	4	7	3	6

AR																				
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20
Q1	4	5	2	3	3	6	4	4	3	5	5	4	4	3	4	5	4	4	3	4
Q2	4	6	2	2	3	4	4	4	3	5	6	4	4	3	3	4	4	4	3	4
Q3	5	5	3	4	4	7	6	6	5	6	4	5	4	4	3	5	4	5	3	4
Q4	4	6	3	3	3	5	6	5	2	5	5	4	4	4	4	3	4	4	3	4
Q5	4	7	2	3	3	4	6	5	3	7	4	5	4	3	4	4	4	3	3	4
Q6	4	4	2	1	2	4	4	5	4	7	6	5	4	3	3	3	4	2	3	3
Q7	5	6	3	4	2	7	6	4	4	7	7	5	4	4	4	4	4	5	5	5
Q8	5	5	2	4	4	7	7	3	5	6	5	5	5	5	4	5	4	7	5	3
Q9	5	5	2	3	4	6	7	5	5	4	5	4	4	5	3	5	2	4	3	5
Q10	5	6	2	4	4	6	7	5	5	5	4	5	5	5	6	4	5	5	5	5
Q11	6	5	2	3	3	7	7	6	6	6	5	4	4	3	3	6	3	5	3	4
Q12	5	5	4	1	4	7	6	5	6	5	3	3	5	3	3	5	3	2	3	4
Q13	5	4	3	4	4	3	4	4	5	6	4	4	4	4	3	4	6	4	5	3
Q14	6	5	3	3	3	6	5	7	5	6	6	4	5	3	3	4	2	3	5	4
Q15	6	5	3	3	6	7	7	3	6	3	4	4	3	3	7	5	6	3	3	3
Q16	5	5	3	3	4	4	7	4	6	5	3	3	3	3	1	5	2	3	5	3
Q17	5	5	3	2	3	5	7	6	6	5	4	4	3	4	6	4	4	5	3	3
Q18	6	6	4	3	3	5	7	5	5	3	5	5	3	3	5	4	5	5	5	4
Q19	5	6	3	2	3	6	7	5	5	6	6	3	4	3	4	5	2	6	5	4
Q20	7	4	3	5	6	7	7	6	7	7	3	5	3	6	6	4	7	6	5	5
Q21	7	5	4	5	6	7	7	7	7	7	4	5	5	6	7	3	7	6	5	5
Q22	7	5	3	5	6	7	7	6	7	7	5	4	5	6	5	4	7	6	3	5
Q23	5	5	3	4	6	7	7	5	5	6	3	4	5	5	6	5	3	6	5	5
Q24	5	4	3	3	3	7	6	6	5	7	5	5	3	3	4	4	3	6	3	4
Q25	5	5	2	3	3	7	6	5	4	7	4	5	4	4	3	3	4	6	3	3
Q26	6	4	2	4	6	5	7	6	5	6	3	4	4	5	3	4	5	5	5	5
Q27	5	4	2	4	4	4	5	5	5	6	5	4	4	4	3	5	3	6	3	5
Q28	7	4	1	3	4	5	7	6	4	5	6	4	3	3	3	4	3	6	2	4
Q29	5	5	2	3	2	3	6	3	3	3	4	5	4	2	3	4	4	6	2	4
Q30	7	5	3	4	3	4	7	5	6	6	5	4	4	4	3	4	3	7	5	5

AF																				
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20
Q1	4	6	3	5	4	6	7	4	6	5	4	5	4	3	7	4	4	4	3	4
Q2	4	6	3	3	4	4	4	5	5	5	5	5	4	3	7	5	4	4	3	4
Q3	6	5	3	4	4	7	7	5	5	6	3	6	3	3	7	5	4	5	3	4
Q4	4	5	3	3	4	5	7	5	4	5	4	4	4	5	4	4	4	4	3	4
Q5	4	6	3	3	4	4	6	6	5	3	4	4	3	3	6	3	4	3	3	4
Q6	4	4	2	2	3	4	6	5	4	4	6	5	3	5	7	4	4	2	3	4
Q7	5	5	3	4	4	7	7	4	5	6	7	5	3	4	7	5	4	5	5	5
Q8	5	6	3	4	4	7	7	3	6	7	5	5	5	7	7	4	4	7	5	3
Q9	6	5	2	4	4	6	7	5	6	6	5	4	5	4	6	4	5	4	5	5
Q10	6	5	2	4	4	6	7	5	6	5	3	5	5	4	7	5	5	5	3	5
Q11	6	6	2	3	4	7	7	5	6	6	5	5	3	5	6	5	3	5	3	4
Q12	6	4	3	2	4	7	7	6	6	7	3	5	5	6	7	4	4	2	3	4
Q13	5	5	3	4	4	3	6	5	6	5	4	4	3	6	7	3	5	4	5	4
Q14	6	4	2	3	3	6	5	5	6	3	6	3	4	5	5	4	2	3	5	4
Q15	6	5	2	4	6	7	7	3	6	6	5	3	3	5	7	4	5	3	3	3
Q16	5	6	2	3	6	4	7	4	6	6	4	3	3	3	7	3	5	3	3	3
Q17	5	4	3	2	4	5	7	6	5	6	5	5	5	4	6	5	4	5	3	3
Q18	6	5	3	3	5	5	7	6	5	6	5	5	5	4	5	4	2	5	5	4
Q19	6	6	3	2	4	6	7	5	5	7	5	4	4	3	5	4	2	6	5	5
Q20	7	5	2	5	6	7	7	7	5	6	4	5	5	6	6	4	5	6	5	4
Q21	7	4	2	5	6	7	7	6	6	6	3	5	5	6	7	5	5	6	5	5
Q22	7	5	2	5	6	7	7	7	6	7	5	5	4	6	7	4	5	6	5	5
Q23	5	4	3	4	6	7	7	5	5	7	4	5	4	5	7	5	3	6	3	5
Q24	6	4	3	4	4	7	4	6	6	7	4	5	4	4	6	4	4	6	3	4
Q25	5	4	2	3	3	7	5	5	5	7	4	5	4	5	7	4	4	6	3	3
Q26	6	5	2	4	4	5	7	6	6	7	3	5	5	5	7	5	5	5	3	6
Q27	5	4	3	4	4	4	3	5	5	7	5	4	4	5	7	4	2	6	3	5
Q28	7	5	1	3	3	5	6	4	5	5	5	4	4	4	7	5	4	6	2	4
Q29	5	4	3	3	3	3	4	3	3	3	4	4	4	4	7	4	2	6	3	4
Q30	6	5	2	4	3	4	6	6	5	6	4	5	5	5	7	3	4	7	5	5

TAR																				
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20
Q1	5	6	4	5	6	6	4	5	4	6	6	6	4	3	7	7	4	7	5	6
Q2	5	5	4	5	6	6	4	5	4	6	6	6	4	3	7	6	4	7	5	6
Q3	7	6	5	5	5	7	4	6	5	7	7	6	5	4	7	7	4	7	4	6
Q4	4	7	5	4	5	5	7	6	5	7	6	6	4	3	7	6	4	7	4	6
Q5	4	7	5	6	6	4	7	6	5	3	7	6	4	3	7	7	4	6	5	7
Q6	4	6	5	4	6	5	6	5	4	6	5	6	4	4	7	6	4	6	3	6
Q7	6	6	5	4	5	7	7	4	4	6	5	6	4	4	7	7	4	6	4	6
Q8	6	7	5	4	5	7	7	6	5	7	6	6	4	6	7	7	4	7	5	7
Q9	7	7	6	7	6	7	6	7	5	7	6	5	5	4	7	6	7	7	6	7
Q10	7	7	6	7	6	6	6	7	4	5	6	6	5	4	7	6	6	7	6	7
Q11	7	6	6	6	6	7	7	7	4	7	7	6	5	4	7	7	7	7	3	7
Q12	7	5	6	4	6	7	7	6	3	7	7	5	6	5	7	7	6	7	4	6
Q13	6	6	6	5	5	3	7	6	5	6	7	6	5	4	7	7	5	4	5	5
Q14	7	6	6	7	6	6	7	6	5	3	6	4	6	3	7	7	5	6	3	7
Q15	7	7	5	6	6	7	7	3	6	7	6	4	5	6	7	6	6	7	6	6
Q16	6	5	5	6	5	7	7	5	6	7	6	5	6	5	7	7	4	6	5	7
Q17	6	5	6	6	6	7	7	6	6	6	5	6	6	5	7	6	4	7	5	6
Q18	7	5	6	7	6	7	7	5	6	3	7	6	6	4	7	6	4	7	3	6
Q19	7	6	6	7	6	7	7	5	7	7	7	4	6	4	7	7	5	7	5	7
Q20	7	7	7	7	6	7	7	7	7	7	5	6	6	7	7	7	6	7	5	7
Q21	7	6	7	7	6	7	7	7	7	7	5	6	6	7	7	6	7	7	5	7
Q22	7	7	7	7	6	7	7	6	6	7	6	6	6	7	7	7	7	7	5	7
Q23	6	6	7	6	6	7	7	6	7	7	7	6	6	7	7	6	7	7	6	7
Q24	7	5	7	7	6	7	6	6	5	7	6	6	6	4	7	6	7	7	5	7
Q25	6	6	6	5	5	5	7	6	4	6	7	6	5	5	7	7	4	7	4	6
Q26	7	4	6	7	6	4	7	7	6	7	6	7	5	4	7	7	7	7	6	7
Q27	7	6	6	6	6	7	7	7	6	6	7	6	6	5	7	6	6	7	3	7
Q28	7	5	6	4	6	4	7	6	4	5	6	6	5	3	7	7	7	7	4	7
Q29	5	4	6	4	5	5	7	3	4	3	5	4	4	3	7	6	2	7	3	4
Q30	7	4	6	5	5	4	7	6	5	6	6	6	5	4	7	7	6	7	5	7

APPENDIX G

HOW TO SET UP APP

App Programing and Installation

Setting up Demo App

The App used for the demo which I showed to participants at beginning was "Augment". Augment is a mobile app that lets you and your customers visualize your 3D models in Augmented Reality, integrated in real time in their actual size and environment. Users can get 30 day free trial by registering at http://www.augmentedev.com/. The app is free and it doesn't need register if you don't upload your own models. This app is available at Play store and Apple store. (Figure 1)

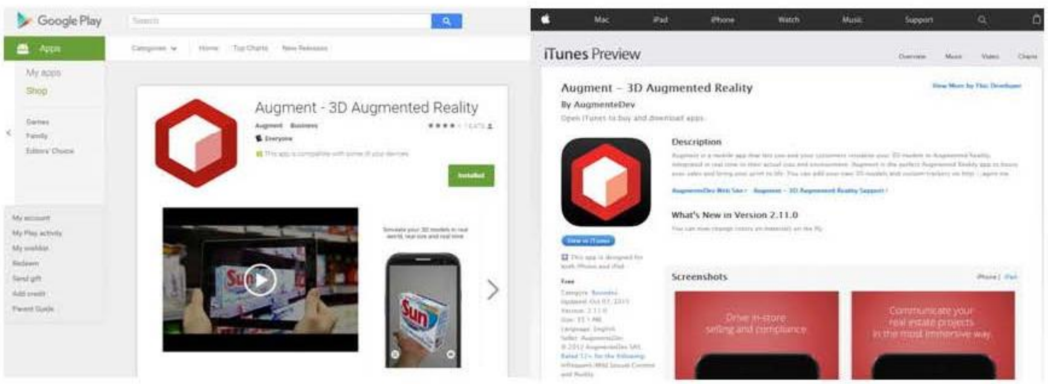
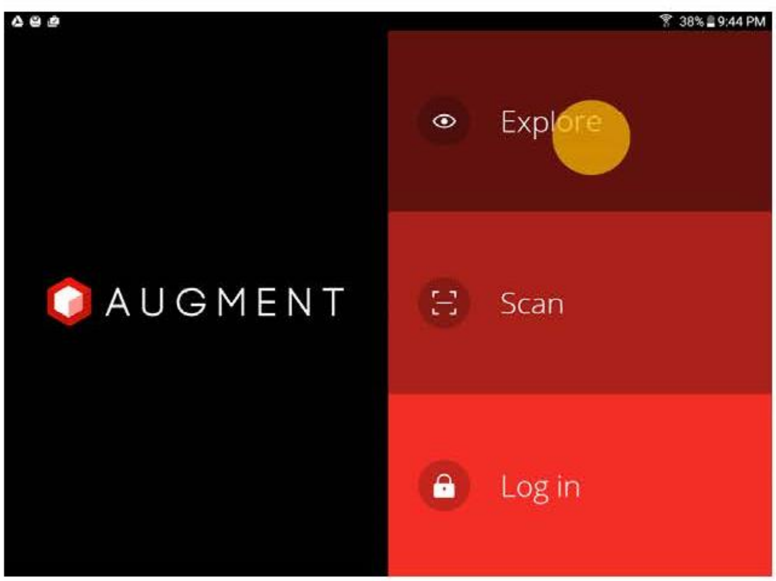


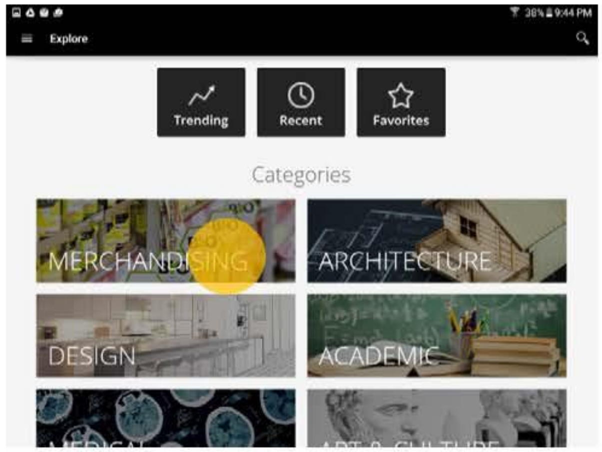
Figure 1. Augment App on Play store and Apple store

After downloading and installing Augment App, we can start to set up the demo for our experiment.

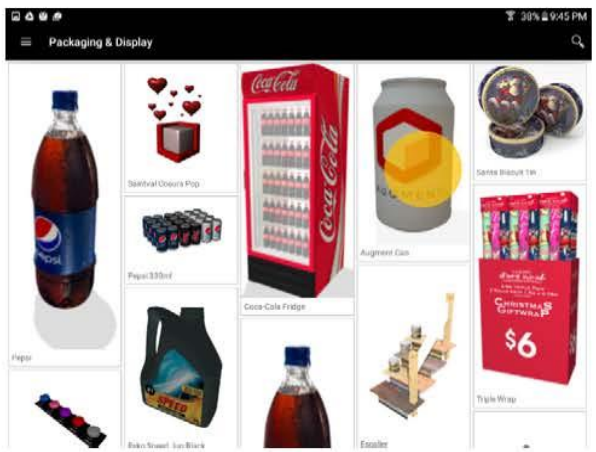
Step 1: Open Augment App, and select "Explore".



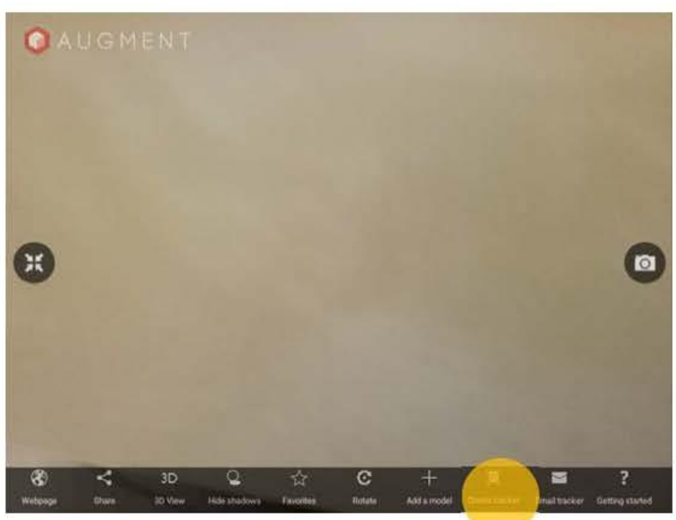
Step 2: Select “Merchandising” in category.



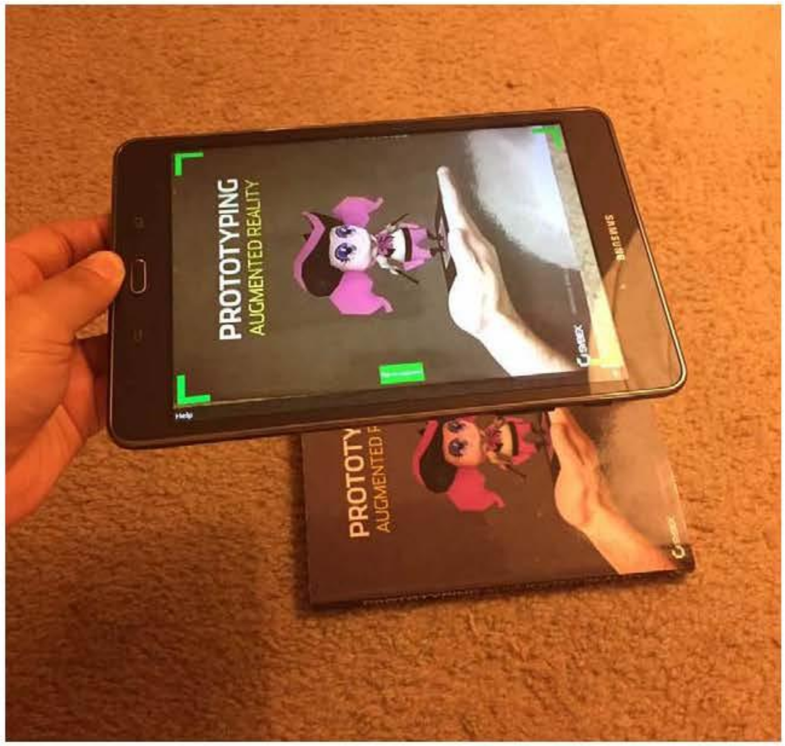
Step 3: Select beverage can from all the models.



Step 4: Select “Create tracker” at the bottom.



Step 5: Find any image which can meets the requirements of tracker, such as a magazine, a book, a picture, etc. Then tap screen when indicators turn to green.



Step 6: Now you can see the model stand on the tracker you just created.

Programing Prototypes for AR and AF groups

The prototypes used in AR and AF groups are programed in Unity 3D. (Figure 2) Unity is a cross-platform game engine developed by Unity Technologies and used to develop video games for PC, consoles, mobile devices and websites.

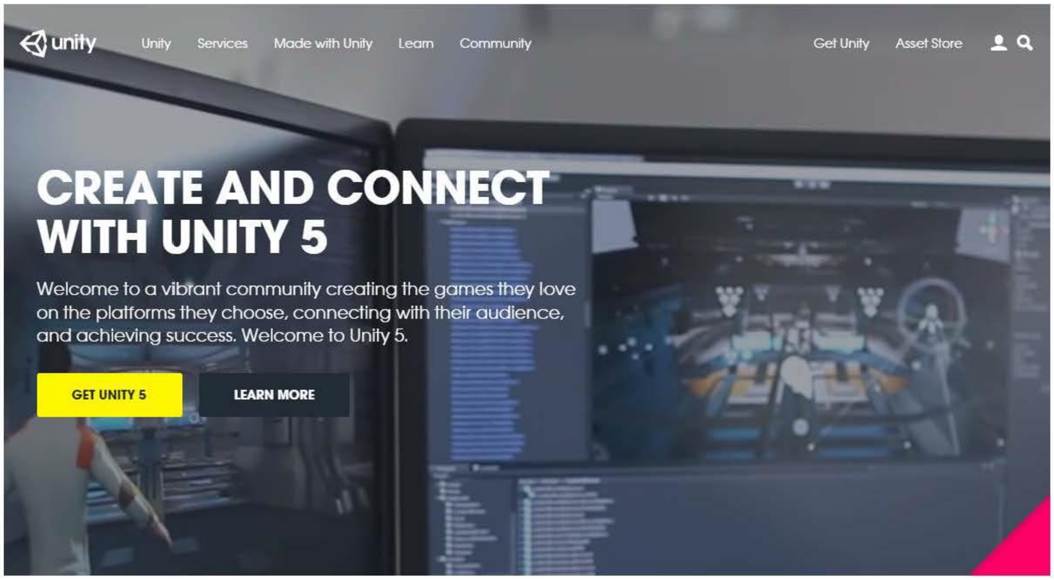


Figure 2. Unity 3D

The first step was to build the 3D models. I finished building models with Solidworks and exported all the models as 3ds files. And then I imported 3ds files into Blender and exported them to obj files for programmer. (Figure 3)

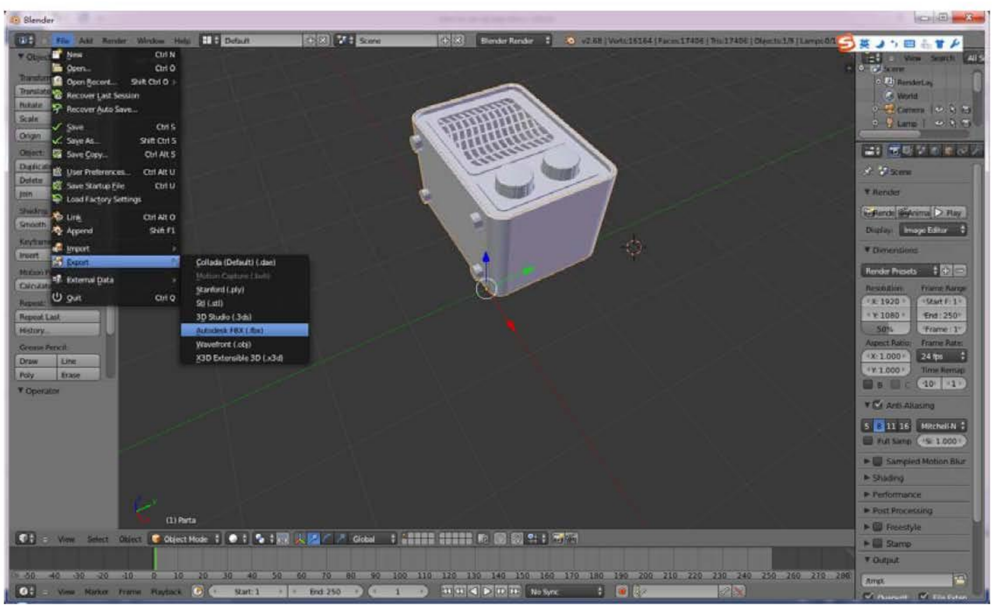


Figure 3. Export obj files

The second step was building the web version prototype by using Javascript. (Figure 4)

Figure 4. Web version prototype

The last step was to put all the 3D models and web version prototype together in Unity 3D and to link them with the image which we used as tracker. After all the programming finished, we got the apk file exported by Unity 3D. Then we need to put the apk file on to tablet. Click the apk file to install the app. (Figure 5.6.7.8)

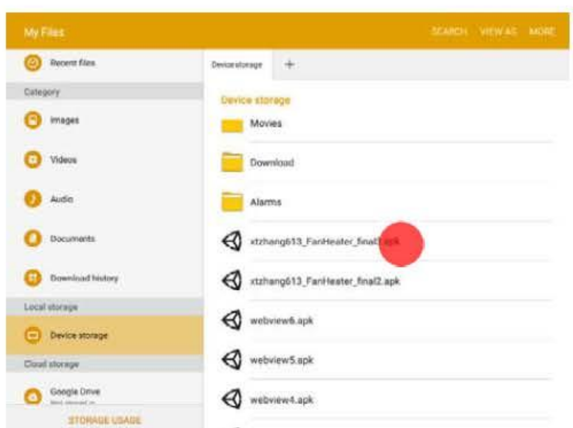


Figure 5. Click the apk file

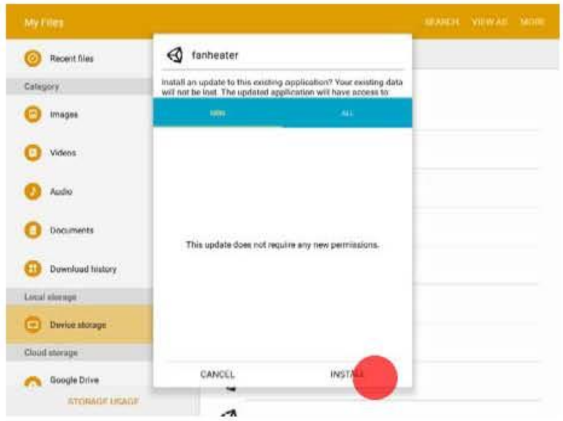


Figure 6. Click “Install”

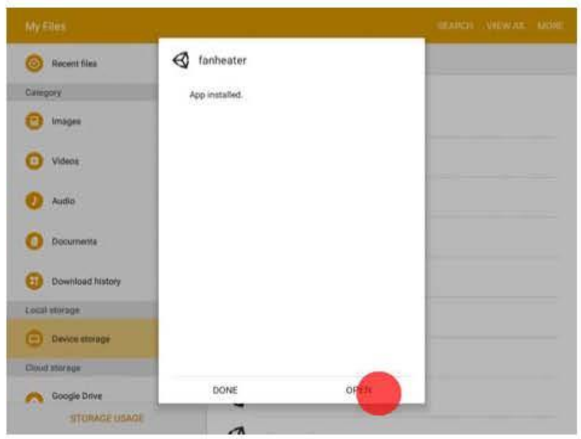


Figure 7. Click “Open”

Figure 8. You can also find the app on desktop

After installing the app, we can test it by printing out the trackers and opening the app to interact with the prototype.

Figure 8. Testing the app

Programing Prototypes for TAR group

The prototype used in TAR group is also programed in Unity 3D. For this case. We need to export 3 different models as obj file: main body, upper knob and lower knob. There were 3 phases for programming.

Firstly, we need to link the main part of this heater to the square marker showing below, and the two knobs also have corresponding round markers.

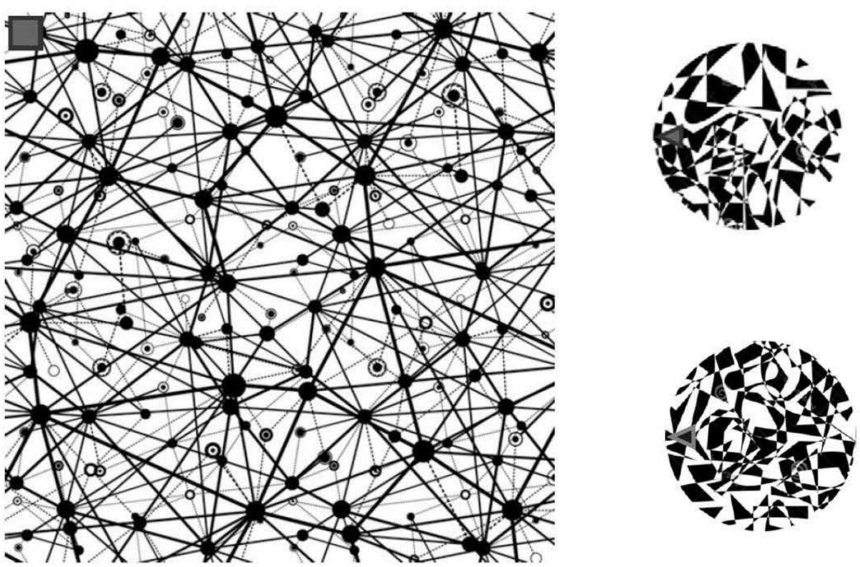


Figure 9. Markers for TAR group

The second step was defining all the angles for the two knobs, such as: the lower knob have four different mode, so that it would be 30° between each mode. The last group was adding effects on to different angles based on second step.

The testing process was the same as AR and AF groups showing above.

REFERENCES

1. IKEA. *Place IKEA furniture in your home with augmented reality*. 2013 [cited 2013 26 July]; Available from: https://www.youtube.com/watch?v=vDNzTasuYEw.
2. Martinez, H., S. Laukkanen, and J. Mattil, *A New Hybrid Approach for Augmented Reality Maintenance in Scientific Facilities.* International Journal of Advanced Robotic Systems, 2013: p. 1.
3. Zhu, J., S.K. Ong, and A.Y.C. Nee, *An authorable context-aware augmented reality system to assist the maintenance technicians.* The International Journal of Advanced Manufacturing Technology, 2012.
4. Billinghurst, M., H. Kato, and S. Myojin, *Advanced interaction techniques for augmented reality applications*, in *Virtual and Mixed Reality*. 2009, Springer. p. 13-22.
5. Lee, W. and J. Park. *Augmented foam: a tangible augmented reality for product design*. in *Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on*. 2005. IEEE.
6. Gillet, A., et al., *Tangible interfaces for structural molecular biology.* Structure, 2005. **13**(3): p. 483-491.
7. Billinghurst, M., H. Kato, and I. Poupyrev. *Collaboration with tangible augmented reality interfaces*. in *HCI International*. 2001.
8. Wu, C.J. and M.S. Hamada, *Experiments: planning, analysis, and optimization*. Vol. 552. 2011: John Wiley & Sons.
9. Friedman, M., *The use of ranks to avoid the assumption of normality implicit in the analysis of variance.* Journal of the American Statistical Association, 1937. **32**(200): p. 675-701.
10. Friedman, M., *A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance.* Journal of the American Statistical Association, 1939. **34**(205): p. 109.
11. Friedman, M., *A comparison of alternative tests of significance for the problem of m rankings.* The Annals of Mathematical Statistics, 1940. **11**(1): p. 86-92.
12. Wilcoxon, F., *Individual comparisons by ranking methods.* Biometrics bulletin, 1945: p. 80-83.
13. 2015; Available from: https://en.wikipedia.org/wiki/Layar.
14. *Layar App*. Available from: https://www.layar.com/.
15. *Augment App*. Available from: http://www.augmentedev.com/.
16. Kim, S.L., et al. *Using Unity 3D to facilitate mobile augmented reality game development*. in *Internet of Things (WF-IoT), 2014 IEEE World Forum on*. 2014. IEEE.
17. *Vuforia Augmented Reality SDK*. 2013; Available from: https://en.wikipedia.org/wiki/Vuforia_Augmented_Reality_SDK.
18. Lund, A.M., *Measuring usability with the USE questionnaire.* Usability interface, 2001. **8**(2): p. 3-6.

19. Lund, A., *Measuring Usability with the USE Questionnaire. STC Usability SIG Newsletter*. 2001, Retrieved 5/3/2009, from http://hcibib. org/perlman/question. cgi.
20. Choi, Y.M. and S. Mittal. *EXPLORING BENEFITS OF USING AUGMENTED REALITY FOR USABILITY TESTING*. in *DS 80-4 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 4: Design for X, Design to X, Milan, Italy, 27-30.07. 15*. 2015.
21. Krantz, M. *Microsoft, Intel, Google outspend Apple on R&D*. 2012; Available from: http://usatoday30.usatoday.com/money/perfi/columnist/krantz/story/2012-03-20/apple-marketing--research-and-development-spending/53673126/1.
22. Cooper, R.G. and E.J. Kleinschmidt, *Resource allocation in the new product process.* Industrial Marketing Management, 1988. **17**(3): p. 249-262.
23. *Custompart*. Available from: http://www.custompartnet.com/estimate/injection-molding/.