Usability Evaluation for Augmented Reality

Satu Elisa Schaeffer, editor

Tekijä — Författare — Author
Satu Elisa Schaeffer, editor

Työn nimi — Arbetets titel — Title

Usability Evaluation for Augmented Reality

Tiivistelmä — Referat — Abstract

In Spring 2014, a small group of students at University of Helsinki took on the task of adapting and applying usability-evaluation techniques for evaluating four different types of augmented-reality applications. This report combines their final reports so that other students, researchers, and IT professionals around the world facing similar situations can draw from their experiences and findings. The course was instructed by the editor of this work.

ACM Computing Classification System (CCS):

H.5 [Human-centered computing]
I.3.2 [Computing methodologies]

Contents

iv

Preface

Satu Elisa Schaeffer

New technology surfaces every once in a while and begins to reshape how people interact with their environment. Devices that allow the creation of an *augmented reality* (AR) have been around for quite a bit now and even before that were envisioned in fiction — both in books and on the big screen as well as television series. With such technology, the reality perceived by a person is modified by computer-generated stimuli — usually visual or auditive information, but also other senses are possible such as perceiving touch (for example vibrations like the already-standard method of cellphones to "ring" silently in our pockets), odors, or even temperature changes.

It is not necessary to *increase* the amount of stimuli, but also the exclusion of undesired perceptions is sometimes necessary; this is called as *diminished reality* and a rather common manifestation of this are the noise-cancelling headphones that make long flights so much more bearable. Also active and adaptive camouflage devices have been proposed to hide visual cues; objects and people can be either "erased" and replaced with "background" or turned into something else entirely (this latter option offering interesting opportunities for future role-playing games). I personally would not mind walking around town and seeing three-dimensional models of how construction sites will eventually look like instead of the hole in the ground that the site presently happens to be.

One of the challenging aspects with building products with AR technology is the interaction it requires — when the system output is presented in the form of stimuli to the senses of the user, it makes little sense for the system input to require the user to press a button on a physical device. Most contemporary AR applications are still based on pointing a smart-phone camera to an object of interest, but the interest that headset devices such as the Google Glass and those manufactured by Vuzix have been generating in the past years, we may soon be past that and simply looking at things and hearing sounds instead of holding a phone. Wearable gadgets that interpret gestures like the MYO bracelet and external input devices that observe our movements like Microsoft's Kinect or the LeapMotion device open very concrete possibilities to methods of interaction far beyond manipulating a touch

1

screen. Also speech-recognition technology and natural-language processing has gotten extremely popular with interfaces like Apple iOS's Siri.

I don't expect us to keep pressing buttons for long — the near-future AR applications will interpret hand gestures, react to voice commands, and gently buzz a wearable device to alert us. No more walking to a lamp post while staring at Google Maps on a tablet when navigating the tourist attractions in Paris, Rome, or Barcelona. Texting while driving will also become a much more complex issue when the car interacts with the user through AR technology and actually does most of the driving autonomously. These are interesting times for an early-adopter tech geek and for a computer-science researcher likewise.

Having been invited to spend my sabbatical year at Helsinki Institute for Information Technology HIIT by Patrik Floréen to collaborate on adaptive ubiquitous computing, the obvious potential of AR technology was the driving force in deciding to hold a seminar on the topic in the Fall semester 2013 at the Kumpula Campus of University of Helsinki. The students innovated and prototyped AR applications during the seminar, including vehicular assistants, tutoring for learning a foreign language or to play an instrument, facial and product recognition (so that we remember to whom we owe money and what we were supposed to remember when picking up a laundry detergent at the store), and even live-action role-playing games. I would like to extend my thanks to those students for the experience.

Also, as it is evident that interacting with AR technology is not directly comparable to the accustomed human-computer interaction with desktop, web, or mobile applications, it was decided that a course of *usability* aspects of augmented reality is also given. The present volume includes the reports of the students who took part in that during the Spring semester of 2014. We discussed the existing methods for usability evaluation and their applicability and adaptability for the purpose of determining the ease of use of AR applications. Each student, individually or in pairs, first selected either an existing AR application or chose to develop one a new one from scratch, and then examined usability-evaluation techniques to experiment with how practical it is to apply such a technique, what modifications are required, and how informative are the results obtained in improving the design of the user interaction for the AR application in question.

After finishing the course, pleased with what we had created, the participants embraced the idea of making their results available for others (students, people in the industry, and researchers). Should the reader wish to contact the authors, we have included short academic biographies along with e-mail addresses in Author information. I have edited the course reports into somewhat consistent chapters and am quite likely to blame for some of the typos, grammar issues, or errors that managed to remain in the manuscript despite the several revisions it has gone through.

My thanks to the students of the course for their hard work and to HIIT and University of Helsinki for having me here, as well as to my home institution Universidad Autónoma de Nuevo León (located in northeastern Mexico, in case the curious reader wonders where that might be) for letting me take this time to explore future technologies and the challenges they bring to our already diverse field of study.

<div style="text-align: right">

Elisa Schaeffer
Kumpula
June 25, 2014

</div>

Chapter 1

Usability evaluation of an AR application for overlaying 3D models

Héctor Martínez & Payel Bandyopadhyay

The way users interact with computers and mobile devices is changing drastically with the new emerging technologies. *Augmented reality* (AR) is one of these technologies that define new ways of user interaction. There has been a large amount of research work done in evaluating user interfaces of traditional systems such as mobile devices and web interfaces. Since AR is one of the new emerging technologies, the number of systematic evaluations done in AR interfaces is relatively low.

In this work, a systematic evaluation of the user interface of an AR application has been done. Out of the existing usability evaluation methods, four methods have been chosen as a guideline to evaluate the targeted AR application. In order to cover all the aspects of usability methods, two methods from usability inspection (namely cognitive walk-through and heuristic evaluation), one from usability testing (laboratory observation) and one from user reports (questionnaire) have been chosen. The AR application that has been evaluated in this project is **Augment — 3D Augmented Reality**. The results obtained from the four usability methods have been described in this document.

Usually, due to limited time and resources, applying all the methods to evaluate an user interface is not feasible. Hence, a comparison of the results of the four usability methods has been carried out. In this comparison, a justification, based on the results obtained, about which usability evaluation method would be more suitable in case of AR interfaces is presented. Finally, a set of design guidelines for AR applications has been proposed.

1.1 Introduction

Nowadays, there are many innovative ways of user interaction with environments, starting from real to virtual environments. Figure 1.1 shows an

overview of the reality-virtuality continuum defined by Milgram and Kishino
[27]. In virtual environment (also known as virtual reality), user sees a com-
pletely synthetic environment which bears no connection with the real en-
vironment. This means that user remains unaware of the surrounding real
environment. Augmented reality is a variation of virtual environment [6].
It combines real and virtual objects in a real environment [40]. This means
that user can see the real environment with some additional objects added
to the real environment. Any application having the following properties
can be classified as an AR application [22]:

- Combination of real and virtual objects in a real environment.

- Interactive and real-time operation.

- Registration (alignment) of real and virtual objects with each other.

Augmented reality is one of the most promising research areas of user
interface. Whenever a user interface is involved with any system, the concern
of usability evaluation and design guidelines of the user interface appears.
Usability evaluation plays an important role in the application development
process. Usually, application developers are experts in their respective fields
and their developed user interfaces might seem very simple to use from their
point of view. Unfortunately, the target users are often novice users of
the applications and find difficult to use the application. Hence, usability
evaluation plays a crucial role in any application having a user interface.

Though AR applications have existed for many years already, the number
of usability evaluations applied to AR interfaces is low [13]. Therefore, in
this project a systematic evaluation of an AR application has been done.
The AR application that has been chosen is **Augment** [5]. The application
is available to download in *Google play* and *App Store*.

The remainder of this chapter has been structured in the following man-
ner. Section 1.2 provides background information of usability methods in
general with focus on the usability methods adapted in this project. Section
1.3 provides the description and analysis of the AR that has been used as
a prototype for evaluation. Section 1.4 describes the detailed adaptation
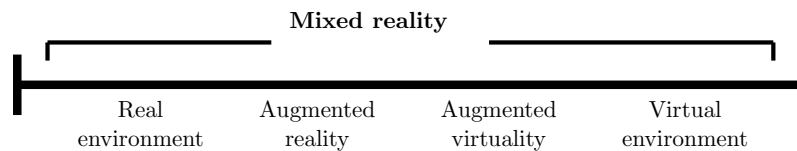
Figure 1.1: Milgram's reality-virtuality continuum [27].

and results of the methods that have been used to evaluate the AR application. Section 1.5 shows a comparison of the results of the adapted usability methods. Section 1.6 finally concludes the whole work done in this project.

1.2 Background

AR being one of the most promising technologies is gaining its importance in various fields of application. Therefore, usability evaluation of AR applications is of prime concern. Usability evaluation of traditional systems like mobile applications or web interfaces is done based on pre-defined usability methods. These methods can be categorized [22, 25] as *inspection* methods, *testing* methods, and *user reports*. Table 1.1 lists some common usability evaluation methods and their corresponding category; note that this is approximate categorization. There may be other methods also which fall in these categories.

1.2.1 Inspection methods

In inspection methods, trained evaluators are involved. These methods are less time consuming than the other categories. Methods like heuristics, cognitive walk-through, feature inspections, guideline checklist, and perspective based inspection fall in this category. In this document, two of these methods have been chosen to evaluate the AR prototype.

Heuristic evaluation   For *heuristic evaluation* [40], one or more evaluators are recruited. The evaluators are often novice to the given system's design. Evaluators examine the user interface of a given prototype and try to find problems in the interface's compliance according to ten standard usability principles. These principles are often called as *heuristics* because they are more in the nature of rules of thumb than specific usability guidelines. The ten heuristics most commonly used are the following [32]:

1. **Visibility of system status:** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

2. **Match between system and the real world:** The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system- oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

3. **User control and freedom:** Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

4. **Consistency and standards:** Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

5. **Error prevention:** Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

6. **Recognition rather than recall:** Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. **Flexibility and efficiency of use:** Accelerators — unseen by the novice user — may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

8. **Aesthetic and minimalist design:** Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

9. **Help users recognize, diagnose, and recover from errors:** Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

10. **Help and documentation:** Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Cognitive walk-through   *Cognitive walk-through* [44] is an evaluation method which is used to inspect the interaction between the user and the interface through some pre-defined tasks. In this method, the main focus is on exploratory learning [35]. Exploratory learning in this context means how well a novice user is able to use the interface without any prior training.

This method can either be applied at the early stage of designing an interface with paper prototypes or during beta testing phase. The method also includes recruiting evaluators who are system designers and designers who are novice to the given user interface [11].

The system designers prepare the first phase of activity which involves preparing the task and the corrective actions. The novice designers of the

Table 1.1: Usability evaluation methods in corresponding categories [25].

Category	Usability-evaluation methods
Inspection methods	Heuristics
	Cognitive walk-throughs
	Pluralistic walk-throughs
	Feature inspections
	Guideline checklist
	Perspective-based inspection
Testing methods and co-discovery	Question asking protocol
	Think aloud protocol
	Performance measurement
	Field observation
	Laboratory observation
User reports	Interview
	Questionnaire

given user interface then try to analyze the task from the perspective of a novice user. This method is based on CE+ theory [43] which defines the four phases of activity, states the following:

1. The user sets a *goal* to be accomplished.

2. The user searches the interface for available *actions*.

3. The user *selects* an action that seems likely to make progress toward the goal.

4. The user *performs* the action and checks to see whether the *feedback* indicates that progress is being made towards the goal.

For every given step that a user would have taken to complete a task, all the above steps are repeated.

The task of system designers in the first phase includes several prerequisites to the cognitive walk-through procedure which are [43]:

- A **general description** of *who* the users will be and what *relevant knowledge* they possess.

- A **specific description** of one or more *representative tasks* to be performed with the system.

- A list of the **correct actions** required to complete each of these tasks with the interface being evaluated.

Then an evaluator — novice to the user interface — attempts to answer the following four questions for each correct action involved in a task [44]:
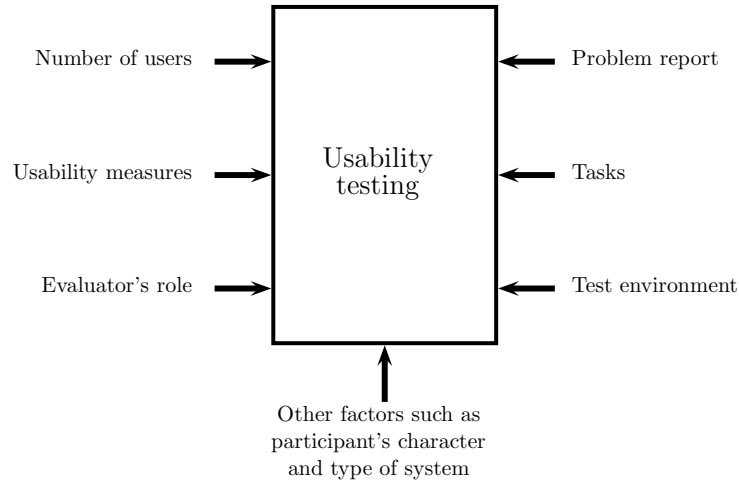
Figure 1.2: Usability evaluation methods (adapted from Alshamari and Mayhew [3]).

1. Will the user *try* to achieve the *right* effect?
2. Will the user *notice* that the correct action is available?
3. Will the user *associate* the correct action with the desired effect?
4. If the correct action is performed, will the user see that *progress is being made* towards the solution of the task?

1.2.2 Testing methods

In *testing methods*, potential users (not designers, not usability professionals) are involved. These methods are more time consuming and costly than the other categories [36]. These methods measure the extent to which the product satisfies its target users. Factors affecting usability testing are shown in Figure 1.2. Methods like co-discovery, question asking protocol, think aloud protocol, performance measurement, laboratory testing, and field observation fall in this category. In the present project, one of these methods has been chosen to evaluate the AR prototype.

Laboratory evaluation   Laboratory evaluation [3] is done in controlled environment where the experimenter has full control of assignments of subjects, treatment variables and manipulation of variables [34]. This "laboratory" does not need to be a dedicated laboratory [21]. Laboratory in this context means controlled environment which mimics real life scenario.

This method is most advantageous in case of evaluating user interfaces

because there is a possibility for the experimenter to do video/audio recordings of the user interface and user interactions. This helps the experimenter to analyze that given a user interface, how the users are going to use it. Since (in laboratory evaluation), the experimenter has relative good control of the assignment of variables and recruitment of participants, the experimenter can recruit participants depending on the target users.

The participants can be either novice users of the application, have used similar type of applications before, have computer science (CS) background (like students, researchers, professors) or non-CS background. Laboratory evaluation gives the experimenter the entire freedom to choose the target users. All depends on the experimenters' need of evaluating the application.

The users perform a set of pre-defined tasks in a usability laboratory. In laboratory testing, the experimenter can provide the users with two types of tasks to find the usability problems: structured tasks and unstructured tasks [3]. The details of the these tasks are explained below:

Structured tasks: Tasks that are structured in a way that the experimenter creates a step by step to-do list which the user performs in order to complete the task. The steps needed to complete the task are clearly defined by the experimenter. This type of tasks can be written down in a very detailed manner, like providing a realistic scenario explaining what the user needs to do. Table 1.2 shows an example of structured task.

Unstructured tasks: Tasks that are written down in an abstract level. The users have the full control of the steps needed to be taken in order to complete the given task. In our project, we have used this task type, hence an example of this type of task can be found in our task description.

Usually, video or at least audio recordings of user task and interaction with user interface are done. From the recordings, the evaluators then seek to analyze the number of errors, the time spent on task, as well as user satisfaction.

1.2.3 User reports

In user-reporting methods, users are naturally involved. These methods are less time consuming than the other categories. Methods like interviews and questionnaires fall in this category. In this project, one of these methods has been chosen to evaluate the AR prototype.

Questionnaire *Questionnaires* [18] are usually performed before or after the testing methods. Often it is difficult to measure certain aspects of users

Table 1.2: A structured task definition by Andreasen et al. [4].

#	Description
1	Create a new e-mail account (data provided).
2	Check the number of new e-mails in the inbox of this account.
3	Create a folder with a name (provided) and make a mail filter that automatically moves e-mails that have the folder name in the subject line into this folder.
4	Run the mail filter just made on the e-mails that were in the inbox and determine the number of e-mails in the created folder.
5	Create a contact (data provided).
6	Create a contact based on an e-mail received from a person (name provided).
7	Activate the spam filter (settings provided).
8	Find suspicious e-mails in the inbox, mark them as spam, and check if they were automatically deleted.
9	Find an e-mail in the inbox (specified by subject-line contents), mark it with a label (provided), and note what happened.

objectively, in those cases this method is used to gather subjective data from users. This involves querying users to gather user opinion and preferences related to *operability, effectiveness, understandability,* and *aesthetics* of the user interface [22].

The questionnaires are *indirect* and cannot be used as a single method to analyze the user interface. The main reason for this is that this technique does not analyze the user interface of the prototype but collects opinions about the user interface from the users. Collected raw data of users' behaviours from other methods are considered more reliable than raw data of users' opinions about the user interface [18].

Commonly, users are first allowed to use the prototype and then users fill up or rate pre-defined questions prepared by evaluators. Evaluators collect the data and try to analyze them in some statistics format.

1.3 Evaluated application

In this section we provide an overall description of the AR application that has been used in this project for usability study. The selected application for the usability study is called **Augment**.

The application is targeted to users who want to provide AR solutions for sales and marketing fields. However, the application can be used by anyone for fun or for other ideas that users may come up with. Figure 1.3 shows a screen-shot from the initial view of **Augment**. The developers [5] describe the application in the following manner:
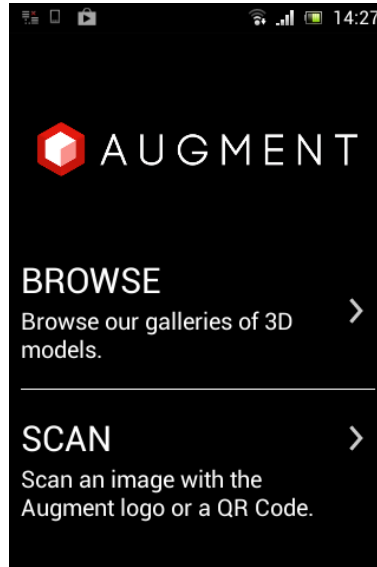
Figure 1.3: Initial view of **Augment**.

> "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. Augment is the perfect Augmented Reality app to boost your sales and bring your print to life in 3 simple steps."

The application has two versions that provide different features. The *free version* has some limited features compared to paid versions. In *paid versions*, users can upload their own 3D models and markers, use a history feature, and so forth. For the study presented in this document, the free version was been chosen. The 3D models can be uploaded from the following websites and software:

1. *Sketchup.*
2. *ArtiosCAD.*
3. *AutoCAD.*
4. *Rhino V5.*

It also supports the plugins for the following software:

1. *3ds Max* plugin (for 3ds Max 2012, 2013, and 2014).
2. *Cinema 4D* plugin (for both Mac and Windows).
3. *Blender* plugin (for Blender 2.68 and above).

Table 1.3: Three standard 3D formats supported in **Augment** that can be exported from most 3D software [5].

3D file format	Extension	Files to upload on Augment Manager
Collada	`.dae` or `.zae`	`.dae` and texture files or `.zae` alone
Wavefront	`.obj`	`.obj` and `.mtl` (materials) and textures
STL	`.stl`	`.stl`

Requirements   **Augment** is an AR application currently available for *Android* and *iOS* mobile operation systems. A Windows version is presently not available. **Augment** supports 3D models in *Collada*, *Wavefront*, and *STL* file formats (these last ones are never textured and appear blue in the application). Table 1.3 provides a more detailed description. All files are uploaded in a compressed (ZIP) format.

Due to the fact that the application has been made by a third party and not by the authors, a detailed description of the architecture of the application cannot be provided in this document. Also, the results obtained in this document cannot be directly applied to improve the design of the application.

1.3.1 Functionality

The application allows users to create AR experiences with their phones and tablets. The application can be used for two purposes.

Sales and design   The sales-and-design functionality of the application is represented by Browse option (see Figure 1.3) on the user interface. This option allows users to select the desired 3D model to use for the augmentation. After the selection of the 3D model, the application view changes to the camera view and the positioning of the 3D model begins (in the cases where the feature is available). From that view, the user is also able to create a marker for the augmentation and use the different options for 3D model manipulation (rotation, translation and scaling) and sharing (using e-mail and/or social networks). Figure 1.4 illustrates this feature.

Interactive print   The interactive-print functionality of the application is represented by Scan option (see Figure 1.3) on the user interface.This option is intended to be used in the cases where the user wants to either scan a QR code or when the user aims to detect one of the predefined markers that the application contains. This option can also be used for visualizing 3D models of a catalogue in 3D. The only requirement is that the image to be used for the augmentation is registered in their site. In order for users to know it,

Figure 1.4: 3D-model user interface [5].

(a) SCAN user interface [5].

(b) A Screen-shot of *marker* usage for placing a 3D model.

Figure 1.5: Interfaces of the application.

each image in the catalogue should contain the **Augment** logo. Figure 1.5a shows how this feature works. The possibility of scanning QR codes option has been analyzed only in the heuristic evaluation due to the limitations of the available time.

Models and markers The application provides several 3D models that can be used to augment the real environment. **Augment** uses image tracking technology for the positioning and orientation of the 3D models. Users can decide to use the predefined images or to use their own images by creating a *marker* within the application. A marker is an image that the application recognizes to place the 3D model in space and at the right scale, in augmented reality. Figure 1.5b shows a screen-shot of the application where a virtual stool has been augmented over a cover image from a magazine which is acting as a marker.

Table 1.4: Specifications of the selected devices.

Device	Aspect	Characteristics
Mobile phone	OS	*Android 4.1.1* (Jelly bean)
	Screen	320 × 480 pixels, 3.5 inches
Sony Xperia E	Video resolution	640 × 480 pixels
	Processor	*Qualcomm MSM7227A Snapdragon* 1 GHz
Tablet	OS	*Android 4.0.4* (Ice cream sandwich)
	Screen	800 × 1,280 pixels, 10.1 inches
Samsung Galaxy Tab 10.1	Video resolution	720p
	Processor	*Nvidia Tegra 2 T20 Dual-core* 1 GHz

1.3.2 Characteristics

Android has been selected as the operating system for the analysis of the four evaluation methods. With the aim of analyzing the consistency of the application through different use cases, two devices have been selected to perform the evaluations. The selected devices are a mobile phone and a tablet (see Table 1.4). The reason for the selection of these devices is that they are different enough in terms of screen, camera resolution and processor to detect inconsistencies in the design of the application. While the mobile phone has a smaller screen and less powerful camera and processor, the tablet provides a large screen with better hardware specifications.

This application has the following two limitations [5]:

Platform limitation This application is currently not available for Windows. Therefore, users having Windows phone or tablet cannot access this application, even if this application interests them.

3D-model limitations Since a mobile device is not (yet) as powerful as a professional computer, the 3D-models for **Augment** have certain limitations on total polygons count and file size. This means that not any 3D model will be compatible with this application. The current polygon-count limit is between 125,000 polygons and 400,000 polygons. The compressed file uploaded to **Augment** must not exceed 15 MB. There is also a limitation of the file size for the textures of 3D models. The detailed limits are shown in Table 1.5.

These limitations somewhat limit the usage of this application. The above mentioned feature limitations are not easy to understand for users having a non-CS background. Even if users are having a computer-science background, they may need to have some knowledge about graphics.

Table 1.5: Maximum number of textures per model according to the selected 3D model's texture resolution [5].

Color space	512×512	$1{,}024 \times 1{,}024$	$2{,}048 \times 2{,}048$
RGB	33 textures	8 textures	2 textures
RGBA	25 textures	6 textures	1 texture

1.4 Usability methods

In this project four usability methods have been chosen to evaluate the AR application **Augment**. The selected usability methods for the proposed study are the following:

Usability inspection #1: Cognitive walk-through.

Usability inspection #2: Heuristic evaluation.

Testing methods: Laboratory observation.

User reports: Questionnaires.

The details of each method have been described below. Each of the subsections below is further sub-divided into experimental design and results. First we describe how the method was adapted and then the outputs obtained from the respective method.

1.4.1 Cognitive walk-through

As explained in Section 1.2.1, this method is divided into two phases: *preparation* and *evaluation*. The details of our experimental design are explained in this section.

Participants   The evaluation is carried out by two evaluators. One of the evaluators has performed the preparation phase and the other evaluator has performed the evaluation phase. Since the application used in this project has not been developed by the authors of this document, the authors have played the evaluators' role one of whom was a doctoral student in AR area and the other was a research assistant of human computer interaction area. Both evaluators were familiar with AR concepts and some concepts of human computer interaction. None of the evaluators had previous experience of using cognitive walk-through method as a usability method for evaluating applications. Both evaluators were provided with lectures on usability evaluation methods, so that the method could be applied in a proper way.

```
(a) Open the application
(b) Choose the desired option
(c) Choose the desired 3D model
        (-) Only for smart-phone in vertical position
        (i) Select CREATE MARKER
        (ii) Read and close help window
        (iii) Perform the scan
(d) Place the model in your environment in a desired way
        (i) Turn on flash (if required)
        (ii) Adjust the scale of the 3D model
                (1) Make it bigger if necessary
                (2) Make it smaller if necessary
        (iii) Rotate it in the desired location
(e) Take a photo of it and save the photo
```

Figure 1.6: Correct steps to be taken to complete the defined task.

Procedure   In the *preparation phase*, one of the evaluators (referred to as evaluator #1) prepares the task that was to be evaluated by another evaluator and the target users of the application.

Task chosen   The first step in cognitive walk-through includes one of the evaluators (referred to as evaluator #1) choosing the task that will be evaluated. Since **Augment** is targeted for purchasing products from catalogues, evaluator #1 has defined a task of choosing a 3D model to imagine how the model would look if placed in the desired surrounding.

Task description   Task description is provided from the point of view of first time users of the application [44]. Evaluator #1 describes a primary task that a user might do with the given application. The described task contains further sub-tasks which required to be done to achieve the higher level task described below [15]. The task prepared by evaluator #1 is:

> Select a Samsung Smart TV 55" and place it in a realistic scale over a table. Then, take a photo and save it. For the augmentation, use the provided image to create a marker. The system will be in a state such that someone could immediately start testing.

Correct sequence of actions   For each task that is analyzed a corresponding correct sequence of actions is described. The correct actions to complete the above defined task are described in Figure 1.6.

Anticipated users   Evaluator #1 describes the targeted users; in this study, the target users are people who have experience with smart-phones but

limited experience with AR applications in smart-phones. They should have basic knowledge of how to use an application in a smart-phone and will have gone through the demo the AR application.

User's initial goal   If there are any other goals that the user may hold at the beginning of the task, these are listed by, evaluator #1.

1. User selects the option to BROWSE the 3D galleries — success.
2. User selects the option to SCAN the 2D catalogues — failure.

According to evaluator #1, if user chooses the first option to BROWSE galleries then user definitely completes the given task and it would be a success. It might also happen that the user chose the second option to SCAN images; in this case, no matter what sub-sequent action the user takes, it is no longer possible to reach the given goal and the result is a failure. Appendix A.1 shows the cognitive start-up sheet used in this study.

Data collection   The data-collection phase is often known as evaluation phase. Data were collected in following way: Evaluator #1 served as *scribe*, who recorded the actions. Evaluator #2 served as a *facilitator* who performed the task and evaluated the user interface. A complete analysis of the interaction between the user and the interface has been done. The evaluation has been done in the following three ways:

1. Facilitator compares the user's goals and the goals required to operate the user interface.

2. Given a goal, facilitator determines the problems a user might have in selecting the appropriate action to go a step further towards the goal.

3. Facilitator evaluates how likely it is that the users' goals might change with the correct user actions and the systems' response.

The facilitator chose an action and recorded answers to the four questions. To assist this process, the scribe used a data collection sheet containing the four questions and the list of possible user actions (cf. Appendix A.2). Second, the scribe took notes individually on corrective step provided by the facilitator.

The cognitive walk-through session took approximately two hours to complete. At the end of the cognitive walk-through, the evaluators expressed their overall conclusions about the AR application according to the task.

Table 1.6: Specific issues identified during the cognitive walk-through. Solutions marked with an asterisk (∗) indicate changes that were discussed before the cognitive walk-through was done, but were also revealed by the method.

Description	Severity
After opening the application user may get confused with two options ∗	Serious
User may get confused in choosing the desired 3D model	Cosmetic
User may not know that swiping the menu bar will show more options	Critical
User may not be able to create the marker in a proper way	Critical
User may not be able to enlarge the 3D model as desired	Critical
Option to rotate the marker is not visible	Cosmetic
User may not use the HELP menu ∗	Serious

Data analysis In the data-analysis phase, for every step of the given task, the facilitator answered the following four questions:

1. Will the user try to achieve the right effect?
2. Will the user notice that the correct action is available?
3. Will the user associate the correct action with the desired effect?
4. If the correct action is performed, will the user see that progress is being made towards the solution of the task?

The scribe recorded the following data:

1. Number of attempts to complete the task.
2. Bugs and usability design issues.
3. Number of times the application crashed.

The evaluations of the facilitator are included in Appendix A. The evaluator has used mobile device to evaluate the AR application.

Number of attempts to complete the task Facilitator had sufficient previous knowledge of using the AR application. Hence, the task was completed at first attempt. This previous knowledge has biased this result. Due to limited time and resources, hiring a separate evaluator (having no previous knowledge of the system) was not possible. If the facilitator did not have any previous knowledge of the AR application this result might have changed.

Software bugs The evaluators have identified two software bugs that had some impact on the evaluators' ability to complete tasks efficiently. The first includes the rate of application getting crashed. The second includes device not working properly. These results are reported in later in this chapter.

Usability-design issues   In addition, the evaluators identified seven areas where the application could be improved to make it easier to use, easier to learn by exploration, and to better support achievement of goals. The design issues judged to have a critical, serious, and cosmetic impact on usability are listed in Table 1.6. The definitions of the criteria used are illustrated below [12]; if at least one sub-criterion is met, the label is applied.

- **Critical** problems

  - Will *prevent* user from completing tasks.
  - Will recur across *majority* of the users.

- **Serious** problems

  - Will significantly increase users' time to complete task.
  - Will recur *frequently* across test subjects.
  - Users will still manage to complete the task eventually.

- **Cosmetic** problems

  - Will slightly increase users' time to complete task.
  - Will recur infrequently across test subjects.
  - Users will complete task easily.

Number of times the application crashed   The facilitator used a mobile phone to evaluate the application. The application crashed five times during the evaluation phase which was for two hours. Provided the facilitator would have used a tablet, the results might have been different in this case.

The scribe used a tablet to prepare the experiment. Though in the tablet the application did not crash while using it, the device stopped working properly since it has been installed. Hence, users might be forced to uninstall the application, even if it served for their purpose of visualizing catalogue in three dimensions.

1.4.2 Heuristic evaluation

Heuristic evaluation method requires more than one evaluator to be reliable, due to the reason that it is not possible that one evaluator can find all design problems. As a result of the limited resources, the authors of this document have acted as evaluators. The evaluation has been based on the ten heuristics described by Nielsen [29], Nielsen and Molich [32]. The evaluators have analyzed the application against the 10 heuristics individually, without any interaction. After the individual evaluation the results have been compared in a group session.

Table 1.7: Summary of the problems found by both evaluators correlated to the ten heuristics. For each evaluator, the number of design problems found for each heuristic is shown. An ID number of each problem is shown in parenthesis.

Heuristic	Evaluator #1	Evaluator #2
1. Visibility of system status	10 (3, 4, 6, 7, 8, 10, 11, 13, 19, 29)	2 (1.i, 1.ii)
2. Match between system and the real world	9 (3, 8, 12, 16, 17, 18, 20, 21, 22)	1 (2)
3. User control and freedom	1 (2)	1 (3)
4. Consistency and standards	4 (1, 2, 7, 8)	1 (4)
5. Error prevention	5 (4, 14, 30, 31, 33)	5 (5.i, 5.ii, 5.iii, 5.iv, 5.v)
6. Recognition rather than recall	2 (7, 10)	2 (6.i, 6.ii)
7. Flexibility and efficiency of use	3 (7, 9, 26)	1 (7)
8. Aesthetic and minimalist design	3 (24, 25, 27)	0
9. Help users recognize, diagnose, and recover from errors	2 (3, 5)	6 (5.i, 5.ii, 5.iii, 5.iv, 9.i, 9.ii)
10. Help and documentation	1 (22)	1 (10)

Each evaluator has used the application individually and with different devices (the devices specifications can be found in Table 1.4). Evaluator #1 has used the mobile phone with Spanish language for the application, while evaluator #2 has used the tablet with the application in English.

Results Each evaluator has provided an individual report of the heuristic evaluation. The reports can be found in Appendix A.3 on page 75. In this section, the final results after both evaluations are discussed. Table 1.7 provides a summary of the errors found by the two evaluators. As evaluations were carried out individually, the reports obtained from both evaluators were not unified. Therefore, a table summarizing the number of design problems found in both cases was needed.

As can be seen in Table 1.7, several design problems have been found in both evaluations. Also, as Nielsen stated, not all problems have been found by both evaluators and therefore the need of more than one evaluator is justified. Some of the problems, though, have been found only by one evaluator as a result of the use of different devices. Also, some problems are related to translation problems only and therefore, they have been found only in the case of evaluator #1 who has used the application in Spanish.

There are several problems that have been found by both evaluators. In fact, there is at least one design problem for every heuristic, which means that an additional effort is needed to enhance the application. One impor-

tant problem stated by both evaluators is the way in which the two main options of the application (Browse and Scan) are displayed. When starting the application, both options without further help are displayed (violation of heuristic #10). As a result, the user may get confused (violation of heuristic #1) and choose the wrong option. Moreover, if that happens, the back button (which by default is used to return to the previous view in Android) exits the application, preventing the user to undo the mistake (violating error prevention heuristic) and violating the standard (heuristic number 4) of the platform.

Some functional problems have been also considered as design problems as they also violate some of the heuristics. The option of scanning a tracker is sometimes confusing as there is not feedback on why a tracker is good or not. Although the application clearly shows when a tracker is valid or when it is not valid, there is no information on why and how to get a valid tracker. Also, when a QR code is not recognized by the application, there is no further information about why it has not been recognized even if the QR code was in the screen. These problems violate several heuristics, such as #1, #2, #5, #6, and #9.

Another common problem found by both evaluators is related to the manipulation of the 3D objects. Rotation and translation of the 3D models is not possible in all axes. Moreover, adding new models to the augmented view resets all changes made by the user in terms of rotation, translation and scaling (without an undo option, which violates heuristic #5).

As it has been explained before, some problems are related to the specific device. From evaluator #1 (using a mobile phone), problems numbered as #3, #7, and #8 are problems that do not appear in tablet. Therefore, a problem of consistency in the application has been found when comparing results from both evaluators. The application should maintain consistency through all screen sizes.

Regarding to language translation, several design problems have been found. Some of them are related to missing of translation, which means that the messages were displayed in English although the application was in Spanish (problems #3, #16 and #22 from evaluator #1). Another kind of language problems is related to a bad translation of the original words, creating non-understandable texts (problems numbered as #17, #18, and #21 from evaluator #1). Therefore, several improvements in the translation of the application are recommended from the heuristic evaluation.

A comprehensive list of all design problems found in both evaluations can be found in Appendix A.

(a) A TV table at home (b) TV table in the laboratory.

Figure 1.7: Comparison of a real-life scenario with laboratory setup.

1.4.3 Laboratory observation

In order to evaluate the AR application from the perspective of the target users, a laboratory evaluation has been performed. In this project, the goal of using laboratory evaluation method is to determine that given the AR application, in this case **Augment**, how well the smart-phone users are able to use it for a desired purpose (described by the authors of this document). The details of the laboratory evaluation method deployed are described below.

The usability laboratory was set up to simulate a part of a room having TV table at home. The laboratory set up was done in order to make it look as realistic as possible, so that users perform the tasks as a customer would have done at home. Due to limited resources, there was a small difference, as shown in Figure 1.7a and Figure 1.7b, between how a TV table would have been placed at home and how it was placed in the laboratory. The test environment was controlled in the context that all the users were provided the same device, all the users performed the experiment in a very quiet environment, all the tests were done using the same network with same speed and the language of the application used was English for all the users. The device that was provided to the users was a *Samsung Galaxy Tab 10.1* (Touchscreen), model GT-P7500 (see Table 1.4 for detailed device specifications).

Seven users (four females and three males) aged between 19 and 24 years were recruited for participating in the experiment. All the users were bachelor or master degree students of Computer Science in University of Helsinki. All the recruited users were smart-phone users but none of them had any experience with AR applications. To avoid biasing of the results, none of the users had any personal relation with the experimenters. None of the users had their mother tongue as English but all were able to read, write, speak and understand English properly. Only English speaking users were recruited as the application's language was set to English. Hence, language

played an important role in understanding the user interface of **Augment**.

All the users were given the same task to perform. The task provided to the users was unstructured. The task was described from the perspective of the purpose of the AR application that is being evaluated. Since **Augment** is targeted for customers who would visualize the 2D image from the catalogue in 3D and try to see how the model looks in their surrounding, the task was described in that manner. We tried to provide the task that a user would perform before purchasing a smart TV for his/her home. Following shows the exact task definition that was provided to the users:

> Use the application to select a 3D model which should be Samsung Smart TV 55" and imagine that you will buy a Samsung Smart TV 55". Before purchasing the Samsung Smart TV 55" you would like to visualize that how Samsung Smart TV 55" would look in your surrounding. So, place the 3D model in such a way as you would have done in real scenario. We will provide you with an image over which you will place the 3D model.

Before each evaluation, the users were given a live demonstration of the application. All the users were asked whether they felt comfortable about using the application. After assuring that the users were confident about using the application, the experiment started. The demo provided to the users included the choosing of the 3D models functionality; how to select the marker; how to place the 3D model over the marker; how to rotate and how to move the 3D model; how to zoom in or zoom out the 3D model; how to use the help menus and how to save the AR desired user interface. We did not specify a time limit to the users. The evaluations lasted for 5–7 minutes followed by a questionnaire session.

The evaluation sessions were conducted by the two authors: one author showed a live demo of the AR application to the users in all sessions and the other author did a video recording of the user interactions in all sessions.

The *data collection* of the laboratory evaluations was done by video and audio recording of the user interactions. All the recordings were done after having permission from all the users. The video and audio recordings were done using a Canon PowerShot A2500 camera. Figure 1.8 shows screen-shots from the video recordings of user interactions from two different users.

The *data analysis* of the video and audio recordings was done accurately so that the maximum number of usability problems could be found. The laboratory tests helped in analyzing the following measures:

- Task-completion time.
- Usability problems encountered by users.
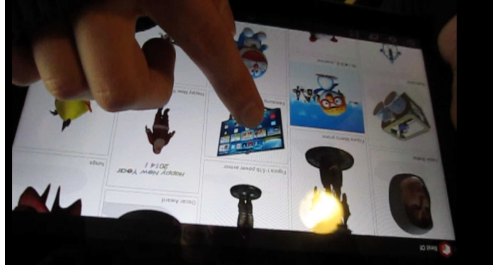
(a) User #1 selecting the 3D model.
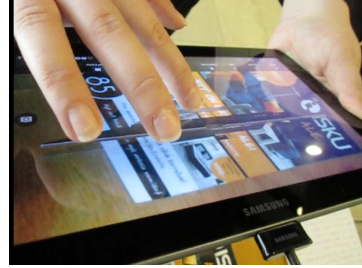
(b) User #2 placing the 3D model in desired position.

Figure 1.8: Screen-shots from the video recordings of user interactions from two different users.
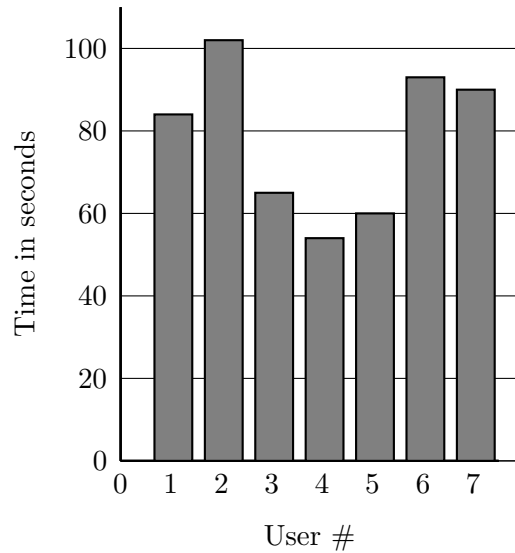
Figure 1.9: Task-completion times of the users.

- Problems that cognitive walk-through evaluator thought might never occur but actually occurred (cf. in Section 1.5.1).

- Success rate of task completion.

- Number of users who used the help menus.

A track of all the video numbers was kept while the above measurements were done. This was crucial because an analysis of number of usability problems faced by each individual user was kept.

The following results were obtained in the laboratory experiments.

Task-completion time   The time taken by each user to complete the given task was measured. The average task completion time of all the seven users

was 1 minute 12 seconds. Though all the users were treated equally, one of the users had some advantages over other users; user #4 not only saw the live demo of the application (given by the experimenters) but also saw the previous user performing the same task. User #4 also did a demonstration herself before performing the experiment. This affected the task completion time significantly. User number #4 took approximately half the time that the other users took to complete the task. This demonstrates that once a user starts using this AR application, the user will get more familiar with the user interface and hence it will be much easier for users to complete the tasks.

All users were novice users in terms of using AR applications. All users were shown a live demo of the application and how to perform the task. User #4 saw the demo once, saw the previous user performing the same task and did a demo herself before performing the experiment. Hence, the task completion time of user #4 was approximately half that of the average of other users but the user made a lot of errors while completing the task. Figure 1.9 demonstrates these findings.

Although user #4 completed the task in approximately half duration than the other users, the user missed few steps in between to complete the task and also took few wrong steps to complete the task. All the details of these errors are described below.

Usability problems encountered by users   In this case, the evaluators have analyzed the problems faced by users by observing the recorded videos. The problems have been categorized according to the standard measures [28] used: critical, severe, and cosmetic problems. The definitions of the criteria used in case of laboratory observation are similar to those used in the cognitive walk-through, reported in Section 1.4.1.

A total of six usability problems (one critical, three severe, and two cosmetic problems) were being experienced by all the seven users in laboratory observation. Since the testing was done using a tablet, most of the problems that would have occurred using mobile device did not come forward. Table 1.8 shows a summary of the problems experienced by the seven users in laboratory observation. The detailed descriptions of the problems have been attached in Appendix A.

Only users #1 and #2 could complete the task successfully, as specified in the task. Both users took exactly the same steps as anticipated by the evaluator (who prepared the cognitive walk-through task). The remaining 5 users completed the task but missed many steps that were required to complete the task.

None of the users utilized the HELP menu option provided in the user interface. Most of them asked for manual help from the experimenter. This

Table 1.8: Number of identified usability problems (number of individual problems found and the total number of common problems experienced by all users).

Usability problems	Individual problems	Common problems
Critical problems	3	1
Serious problems	4	3
Cosmetic problems	2	2

observation clearly demonstrates that designers should try to design the user interface in such a manner that it should have least learnability curve.

1.4.4 Questionnaire

After the laboratory tests, users were asked to fill a questionnaire. The goal of the questionnaire is to evaluate the degree of user satisfaction after the laboratory tests performed by the users. The questionnaire contains 13 statements that have been designed taking into account the results of the heuristic evaluation. However, in this work we are analyzing the results of the questionnaires as an isolated usability evaluation method.

As stated before, the users that filled the questionnaire are the same users that have performed the laboratory tests (i.e., seven university students (four female and three male) with ages between 19 and 24, smart-phone users and with no previous experience in AR applications).

Users have been asked to grade their conformity with the 13 statements on a scale from 1 (totally disagree) to 5 (totally agree) and they have had the opportunity to comment every statement individually in a text area. The questionnaire answers can be found in the A.5. Note that although the laboratory tests and the questionnaires have been carried out by the same users, the results obtained from the questionnaire do not reflect the same results obtained from observation of the laboratory tests. Figure 1.10 shows the frequency of each mark for each of the statements. We discuss each statement individually:

1. *The system provided me with feedback on what I was working. I was not confused or lost while performing the task*

   The majority of the users have felt that they have been in control of the application and that they have not been confused while using it.

2. *The messages that appeared in the application were self-explanatory and I understood what they were trying to explain*

   The messages have been clear for all users.
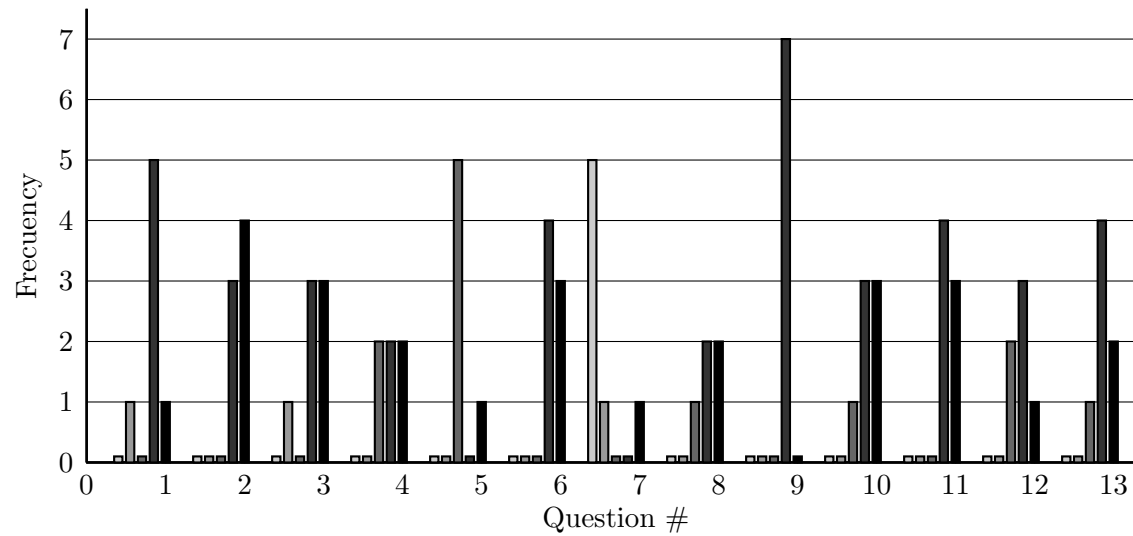
Figure 1.10: Results from the questionnaires. The plot shows the frequency of every mark for each statement; the lightest bar corresponds to one and the darkest to five. A very short bar is drawn for the zero-frequencies as a visual aid. There were four statements (numbers #4, #5, #8, and #12) that were not answered by all users.

3. *I was in control of the application all the time*

The majority of the users have rated positively this statement. However, one user has rated this statement as a 2, showing that not all users are happy with the way that the application is controlled.

4. *I could easily undo/redo any action if I felt to do it*

There is no uniformity in the results of this statement. One user has commented that rotation with fingers was a hard task.

5. *If I mistakenly chose a wrong 3D model, I could easily stop uploading it*

The majority of users have rated this statement as a 3. The reason of this grade is probably that they have not faced such a problem (some have stated this in the comment) as they have felt pleased with the selected model, even if it was not the model that they have been asked to select.

6. *The navigation through the application was easy*

Users have been able to navigate through the application easily. The users were instructed before performing the tests, so the results of this statement are the expected.

7. *The application had many errors and/or crashed*

All users have agreed that this statement is false, as they have not encountered any error or crash (note that one user has rated this statement as a 5, but in the comment of this statement he has written "Didn't crash at all, always a good thing for an app", which means that he misunderstood the marking of this statement).

8. *The application always verified the 3D model before loading*

Users have agreed with this statement. Probably the question should have been redefined to more clearly reflect the problems found in the other usability methods.

9. *The option to select the desired function was always clear and available all the time and the icon images helped me to know the appropriate functionality of the available options*

All users have rated as a 4 this statement. This may lead to think that although they were comfortable with the layout of the options, some improvements can still be done to provide a better experience. One user commented that words are easy to understand even for a non-native English speaker.

10. *It was easy to find the desired options at any time*

The grades of this statement are positive in general. However, as all users have been instructed before using the application, a more unified grading (i.e. a majority of 5s) should have been expected.

11. *The application was well designed visually*

    Users have considered that the application is well designed visually.

12. *If error messages appeared, they were clear in their description and probable steps to recover from it were provided*

    Users have rated this statement either as good (4 or 5) or as neutral (3). The reason for these grades is that users have not encountered errors while performing the task. This has been also reflected in the comments to this statement.

13. *When I needed help, the demo videos and feedback helped me to complete my task successfully*

    The majority of users have rated positively this statement. However, they may have been rating not only the help of the application, but also the instructions presented to them before the tests, as reflected in one comment.

It can be concluded that the users have felt comfortable using the application in general. A larger number of users is required to obtain more robust conclusions. However, due to the restricted conditions in terms of resources and time of this work, the results can be considered appropriated and may open a way for future evaluations. Probably, instructing users has introduced a bias in the way of how a new user would use the application. However, the instructing session has been essential as users were not familiar with AR technology. Also, users have not encountered errors and crashes while using the application. Although this is a good aspect of the application, several errors have been detected while performing the other usability evaluations with the phone case. As these errors have not appeared during the laboratory sessions, the design problems related to errors and their texts have not been found by the users.

In future work, it would be interesting to include tests with users that have not been instructed in the application before using it. The results would probably differ in some statements (e.g. statement #5). Also, another test using a mobile phone instead of a tablet could reveal additional errors or crashes, possibly altering the grades of some of the statements (such as statement #7).

It would be also important to carry out another session with users that are familiar with the concept of AR. From the results obtained in the grades and comments of the questionnaires, the majority of users have been surprised by the novelty of what they have been seeing by first time. Therefore,

users have been more concentrated in understanding AR and how to use it rather than detecting the real design problems. This could have been very different if the application to be evaluated was from a field more familiar to them, such as a messaging application. In that case, they would have had the chance to compare how the application is designed against some previous knowledge that they already have from previous experiences. One positive aspect of this fact, however, is that although they were not familiar with AR, they were able to rapidly use the application, showing once more the fast learning curve of AR technology, as discussed by Sumadio and Rambli [39].

Finally, one important design problem found while performing the other usability evaluations was the translation of text to Spanish. The laboratory tests have been performed with the English version only. In a future laboratory session, it would be good to include some tests with Spanish speakers in order to analyze these language problems. However, due to the restricted resources and time and in order to maintain uniformity in the results, testing with other languages has been left outside of this first laboratory session.

1.5 Comparison of the evaluation results

This section presents a comparison of the results found from the four usability methods. The purpose of this comparison is to find the most appropriated method to be applied while evaluating an AR application. The main reason for analyzing this factor is that, due to limited time and resources, organizations are not able to apply all the usability methods. Hence, choosing the best usability methods is an important issue, so that most of the usability problems could be figured out from the method.

1.5.1 Cognitive walk-through versus laboratory observations

The problems that the users have encountered with respect to the task analysis carried out in cognitive walk-through by the facilitator have been demonstrated. In cognitive walk-through, the facilitator analyzed the same task that has been given to the users to perform in laboratory evaluation. The scribe provided correct steps that a user would have taken to complete the tasks.

Figure 1.6 describes the task divided into steps that the facilitator has used to evaluate the user interface of the AR application **Augment**. At first, the behaviour of the users in those steps has been described. Then, the opinion of the facilitator has been compared with it. The following list shows the steps of the correct actions described in Figure 1.6 and corresponding to it, users' behaviour is described.

(a) Since all users were smart-phone users, they could easily open the application at first attempt.

(b) In the demo session none of the users was shown the possibility to select the second option. So, all users selected exactly the same option which was shown in demo session. As experimenters, we think that if the users were shown the possibility to select the second option then they would have done so. Hence, many of the users would have chosen the wrong option and would have been directed in the wrong path.

(c) Though the exact name of the 3D model to be used was described in the search task, none of the users used the search box option to search for the model. The possibility of using the search option was not shown in the demo session. Hence, it reflects that most users will probably ignore the search option provided at the top of the user interface of this application. Most of the users tried to fetch the 3D model from the catalogue of the 3D models. The raw data obtained from the video recordings is shown in Appendix A.4. One of the serious problems found was finding the 3D model. Few users took some manual help from the experimenters while few users chose entirely different 3D model than the one that was asked in the task to move forward towards task completion. Figure 1.11a presents the findings. Figure 1.11a shows the percentage of users who could easily find the 3D model and percentage of users who had difficulty in finding the 3D model.

(c.-) The demo was shown using the tablet in horizontal direction. Hence, all the users held the tablet in horizontal direction. Therefore, it is not possible to make any evaluation on this task.

(c.i) One of the critical problems that was found was creating the tracker for placing the 3D model. Since tablet was used for usability testing, the 3D model appeared on the user interface even if the tracker was not created. Unfortunately, in case of mobile device, the 3D model will not appear if users are not able to create tracker properly. Hence, this problem was categorized in critical level. If a mobile device would have been used for laboratory testing, then most of the users would not have completed the task. Figure 1.11b summarizes the results of this observation. The raw data obtained from the video recordings is shown in Appendix A.4. Figure 1.11b shows the percentage of users who could easily create the tracker and percentage of users who had difficulty in creating the tracker.

(c.ii) To create the tracker, the user had to press the window button that provides the user with help menu with information about how to use the marker. Though the users who were able to create the marker opened the window, none of the users read the help instructions. They performed as it was shown in the demo session.

(c.iii) Perform the scan: This step is the one where the creation of the marker is performed. Users who remembered the need for creating the

(a) Users who found the desired 3D model easily (43%, dark gray) versus those who could not (57 %, light gray).

(b) Users who easily created a tracker (29 %, dark gray) versus those who did not (71 %, light gray).
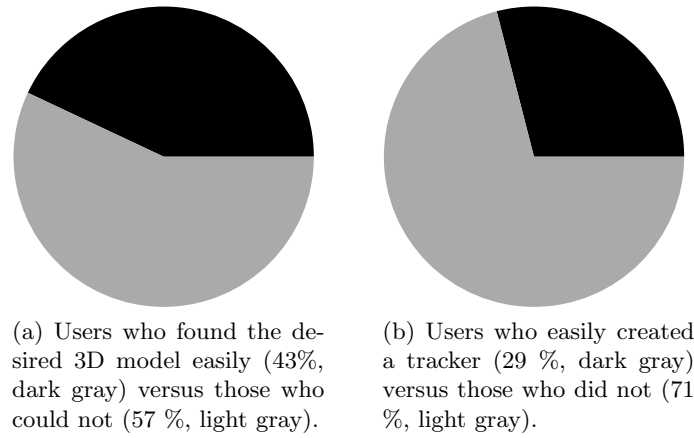
Figure 1.11: Percentages of uses with ease versus difficulty in certain actions with the application.

marker were able to scan the marker as they were explained in the demo session that they had to press the screen when the green light appears on the screen. However, some users needed more time than others to obtain the green color as there was not additional information in the application on how to improve the scanning of the marker.

(d) Most of the users could easily place the model. This is because in laboratory evaluations, tablet was used. In tablet, even if the user is not creating a marker the model will appear in front of the camera.

(d.i) The flash option was not shown to the users in the demo. None of the users used the flash option, though it was one of the option displayed in the bottom menu.

(d.ii.1) All the users could easily enlarge the 3D model. b) All the users could make the 3D model smaller. Both of the above options were shown clearly in demo session to all the users.

(d.iii) Most of the users who used the *rotate* button tried to rotate the model in x and z directions also. Current state of application, allows rotation only in y direction. Figure 1.12 summarizes the results of this observation. The raw data obtained from the video recordings is shown in Appendix A.4.

(e) All the users could perform this step successfully, as it was shown in demo session.

Hence, the results of the two methods demonstrates that the facilitator's opinion is not always accurate due to the large variability of the users' behaviours. Hence, both methods should be used to find a larger number of usability problems.
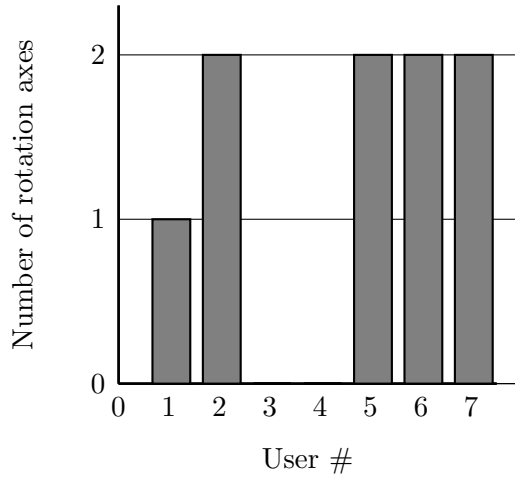
Figure 1.12: The number of unsupported rotation axes that the users attempted to apply; when using the ROTATE button, the users tried to rotate the model in x-axis or z-axis also, even though only y-axis is available for rotation.

1.5.2 Heuristic evaluation versus questionnaire

In this section, the results obtained from heuristic evaluation and questionnaires are compared. Although the number of problems found by evaluators in the heuristic evaluation has been large, the users have not reflected this fact in the questionnaires. The statements presented in the questionnaire have been designed to deal with the heuristic problems found. However, users have rated all statements positively and therefore, the problems found in the heuristic evaluations have not been found in the questionnaires. Therefore, a proper comparison of results cannot be carried out for these two methods.

As it has been mentioned before, several problems found in the heuristic evaluation are related to the specific device (mobile phone) and language of the application (Spanish). As the questionnaires have been filled by the users that carried out the laboratory tests with the tablet and with the application in English, it is obvious that those problems could not have been found in the questionnaires.

Some further tests have been already proposed in order to obtain a larger set of questionnaires that could help to detect design problems. However, with current results, questionnaire method cannot be suggested as a good tool for detecting design problems compared to heuristic method. Therefore, for small scale evaluations, heuristic method is suggested if the number of users for questionnaire method is low.

1.6 Conclusion

In this work, the analysis of four usability evaluation methods has been presented. From all available usability methods, two methods from usability inspection (cognitive walk-through and heuristic evaluation), one from usability testing (laboratory observation) and one from user reports (questionnaire) have been selected. An already available AR application, called **Augment** has been selected as target of the four evaluations. The goal of this work is to analyze the four methods and to detect the suitability of the methods for the proposed aim.

Due to the limited resources and time, the authors of this document have acted as evaluators in the methods were expert evaluators were required while a limited number of users (seven real users) has been used to carry out the user evaluations. Although the application is available in two operating systems, Android has been selected as the target operating system due to the same restrictions.

During the evaluation carried out in this work, several design problems have been found. The results obtained show that there is a need for the developers to revisit the application in order to enhance the design of the interface. Several problems have been found in the laboratory observations, these problems have not been reflected by the users in the questionnaires. This may be due to the fact that users are not familiar neither with the design problems nor with AR applications.

From the four methods, questionnaires appear as the method that has provided a lower number of design problems. Although it is a good practice to include users with no previous knowledge with AR applications, users with prior knowledge could have probably provided better results in the questionnaires. However, the combination of methods has shown good results as not all design problems have been found in all methods. Moreover, using two evaluators in the heuristic evaluation has demonstrated that using more than one evaluator in the method provides better results.

According to the results obtained from the laboratory tests, the learnability curve of the novice users of this application seems high. In order to evaluate the learnability curve of the AR application, further analysis based on the common problems has been performed. The four possible actions by the users are summarized as:

1. User who easily found a 3D model.
2. User who could not easily find a 3D model.
3. User who could easily create a tracker for the 3D model.
4. User who could not easily create a tracker for the 3D model.
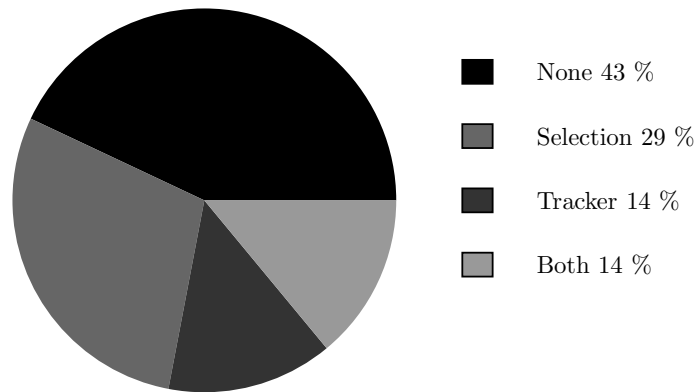
Figure 1.13: Proportions of users who did not face any problem, users who found selecting 3D model a problem but could easily create the tracker, users who could easily select a 3D model but could not create the tracker, and users who faced both problems.

In order to analyze the percentage of users who faced problems in both selecting the 3D model and creating the tracker, users who faced either of the problems and users who faced none of the problems, a graph showing the results is presented in Figure 1.13.

With the aim of detecting the larger number of possible design problems, the methods have been combined for the two different devices (Section 1.3.2) and two languages (English and Spanish). This decision has demonstrated to be a good approach as several design problems have been found only for a specific device and/or for a specific language.

1.6.1 Guidelines

In this section, a series of design guidelines on how to improve the application and how to perform usability evaluations in this kind of AR applications are presented.

Specific recommendations   The **Augment** application has shown several design problems. Some recommendations to enhance the application interface are the following:

- Follow the standards of *Android* platform.
- Provide more intuitive interfaces and a more organized "option dialog".
- Improve the manipulation of 3D objects.

- Provide more information about AR at conceptual level.

- Explain the characteristics of a good marker.

- Provide help tips in the appropriate context.

- Provide better translations of the languages.

Design guidelines for AR evaluation   Evaluating AR interfaces is a relative novel issue compared to other interfaces. Therefore, the combination of several methods, like the approach presented in this document is suggested in order to obtain more accurate results. From the four methods studied in this work, the following conclusions can be summarized:

- Combining one usability inspection method with one usability testing method is recommended to obtain a reliable outcome.

- Using more than one expert in the inspection methods is suggested.

- If questionnaire method is going to be used, the number of users to fill it should be large enough and contain a variety of users, including AR experts.

1.6.2 Future work

The work presented in this document faces one problem of available resources and time due to the restrictions of the context of the work. In order to obtain more robust results, further evaluations need to be done.

One of the first ideas of future work for the proposed study is to create a larger laboratory experiment including more users in order to obtain more statistically consistent results in both, laboratory observations and questionnaires. Including AR experts in the list of users could provide a large advantage as many users have demonstrated to be paying attention to the novelty of AR concept rather than trying to carry out the task in a proper way. Also, the experiments should include both devices to be used by the evaluators and also several users carrying out the tasks in Spanish language would be desirable. If conditions were ideal, more user tests with other devices and also with *iOS* devices as well as other languages should be carried out.

Apart from more user tests, the cognitive walk-through and heuristic evaluation methods should be carried out by more evaluators. It would be desirable to recruit three evaluators for each method in order to obtain more accurate results.

Finally, including further methods apart from the four studied methods would provide a better overview of which usability evaluation method can be more suitable for the evaluation of AR interfaces.

Chapter 2

Design and user evaluation of an AR interface for museums

Maninder Pal Singh & Lei Wang

Traditional visit to museum exhibitions provide limited interaction between the visitor and the museum artifacts. Exhibits labels, booklets and audio guided tour are the common mode of interaction in museums. It does not involve personalized information source to the visitor's and confined to minimal information. It leads to no direct involvement between the visitor and the museum exhibit.

This chapter describes the design and user evaluation of an *augmented-reality* (AR) interface for museums. We evaluate usability evaluation methodologies for museum application. The augmented reality interface for museums focused on providing an audio transcript, a textual information and a video overlay for digitized museum artifacts.

2.1 Introduction

Real-time interaction is a natural way to perform interactions by human beings. Augmented Reality holds the potential to bridge the gap by providing real-time interaction with Augmented Reality environment. Considerable research efforts have been devoted to develop more efficient interactive methods to learn more about certain statues and artifacts display in the museum. In this project involving both the design and the usability evaluation, we focus on an *augmented-reality interface for museums*.

The augmented-reality *Android* smart-phone application aims to combine surrounding reality and the knowledge of the Web together. The application's goal is to provide a deeper museum experience. The application interface is kept simple and functions by scanning QR codes[1] or other pattern markers.

---

[1]A *QR code* is similar to an augmented-reality marker in appearance, but can encode more information than traditional markers [19].

The initial premise of the application is to identify certain statues and artifacts and provide an audio transcript, a textual information and a video overlay for digitized museum artifacts to explore the history as close as possible. The application is a great way to make learning more fun and enhance the knowledge of history already on display in the museum.

This work describes the design and user evaluation analysis for the augmented-reality *Android* smart-phone application. The goals of design and user evaluation analysis of augmented-reality interfaces include assessing the user's ability of application usage, the length of time, and the number of steps takes a user to perform a certain set of predefined tasks.

We utilize these user-performed tasks to measure their performance to draw generic behavior analysis for the application. The analysis also includes an exit interview of the user. The exit interview helps us to identify potential design concerns and it will help us to improve the efficiency and end-user satisfaction of the application. It also helps us to locate the agreeable aspects of the application. The overall analysis helps us to perform initial outline for the improvement of the application before deploying an application into a real time environment.

2.2 Related work

There are not yet many studies in the area of usability evaluation for augmented-reality applications. Nielsen [31] gave a well-accepted definition of *usability* in general, stating that it is a quality attribute that assesses of how easy to use user interfaces are. There are five quality attributes defined in the Nielsen's definition: *learnability*, *efficiency*, *memorability*, *errors*, and *satisfaction*. However, there are also definitions given by others, such as the International Standards Organization (ISO).

With the increasingly application of augmented-reality technology, Kostaras and Xenos [22] called for building high-quality and usable products, which makes people pay more attention on usability testing. So, usability evaluation is gradually accepted as a essential process to build usable products. They describe general guidelines to conduct usability evaluation for augmented-reality based products. They also argue the advantages and weaknesses of three different categories of evaluation methods: *inquiry*, *inspection*, and *testing*. However, evaluation result metrics are not included in their research.

Evaluation with users is a fundamental method in *usability testing*. Although, it is considered as a method which is quite expensive and time consuming, some authors, such as Gould and Lewis [17] along with Rubin et al. [37] affirm that this type of testing has speeded up projects and led to a significant reduction in their costs.

Usability tests can help in the development system or may be directed to the evaluation of the overall quality of interface. Generally this type of test is a composite process in four steps [37]:

1. *Preparation*: all the elements involved such as equipment and software, questionnaires and scripts for testing are appropriate.

2. *Introduction*: tasks required to execute are shown to the user.

3. *Test*: an evaluator assists each user test, done in isolation.

4. *Final session*: after performing the task, each participant is invited to make comments or general suggestions through interviews and questionnaires.

Another recent research by Zainuddin et al. [47] employs *heuristic evaluation*, which is one of inspection methods. This evaluation technique involves experts of this field instead of the end users. The authors do not consider the metrics to measure the evaluation results, either.

Martins et al. [26] propose usability-evaluation metrics for augmented-reality applications based on the heuristics of Nielsen, but does not involve a cognitive walk-through for an augmented-reality application. The *cognitive walk-through* is an inspection method that mainly focuses on the ease use of system for a first time user.

The ARCO project [45] attends the possibility of proposing the museum curators the creation of virtual replicas of the museum cultural objects. The project focused on a web interface to provide real time experience for museum visitors.

2.3 Prototype implementation

In this work, we demonstrate a usability evaluation based on cognitive walk-through for an augmented-reality *Android* application for museums with some evaluation metrics. The application allows museum curators to present a information regarding the museum artifacts. The curators can build and manage interactive information; the cultural artifacts can be digitized by means of images, texts, or audio transcripts [9], as well as by the means of 3D models that are accompanied by short video information. Such information can be managed through the use of content management applications which allows museum curators to provide the information to the visitors on the fly based on demand. The interactive smart-phone application aims to transform museum visitors from passive viewers to active participants with more sources of information.

The proposed application can be divided into two main components: *client* interface and *server* interface.

Client interface The client interface of the application is implemented using *Android* application framework. The *Android* smart-phone application provides quick access to QR scan and AR-scan capabilities to the application.

The artifacts explored by the user — either by QR scan or AR scan — will be considered as recent activities of the user. The recent activities are listed in the application as per descending order of exploration of museum artifacts.

The museum artifact information retrieved by QR scan allows users to download an audio transcript to receive more information on the museum artifacts. The AR scan allows users to augment the museum artifacts with video information projected as an overlay on the artifacts. A single artifact can be bound to the several possible overlays of video information.

The application also allows users to *share* their recent activities or recently explored artifacts information with social-media applications. The inherit capabilities of *Android* framework has been used to share the information across any application in the smart-phone which is having shareable feature. The application also provides an interface to manage the listed recent activities. The client interface is created as a simpler application with focus on the museum visitors with little experience.

Server interface The server interface of the application allows museum curators to build and manage interactive information on museum artifacts. The museum curators manage the database of the representations of cultural objects. The associated media objects, such as images, texts, audio transcript, and video will be hosted on servers in the museum.

The client smart-phone application can request for information for these cultural objects. The information is accessible via QR scan or AR scan. The server-client interaction is based on the *pull* model. The pull request made by a user retrieves the required cultural object information in the form of text, images, audio, or video media.

Mobile application The interface for museums is built as an *Android* smart-phone application. The application interface is shown in Figure 1 and Figure 2. The interface follows an application design pattern provided by *Android* framework. The interface aimed to be simple so that novice users could get used to the application easily. The artifacts explored by museum visitors using an AR scan or a QR scan is handled in an abstract and descriptive data representation in the application. Several screen-shots of the application are included in Figure 2.1.

(a) Recent activities
(b) Activity view
(c) Activity context menu

(d) QR Scanner
(e) Audio transcript downloading
(f) Audio transcript playing
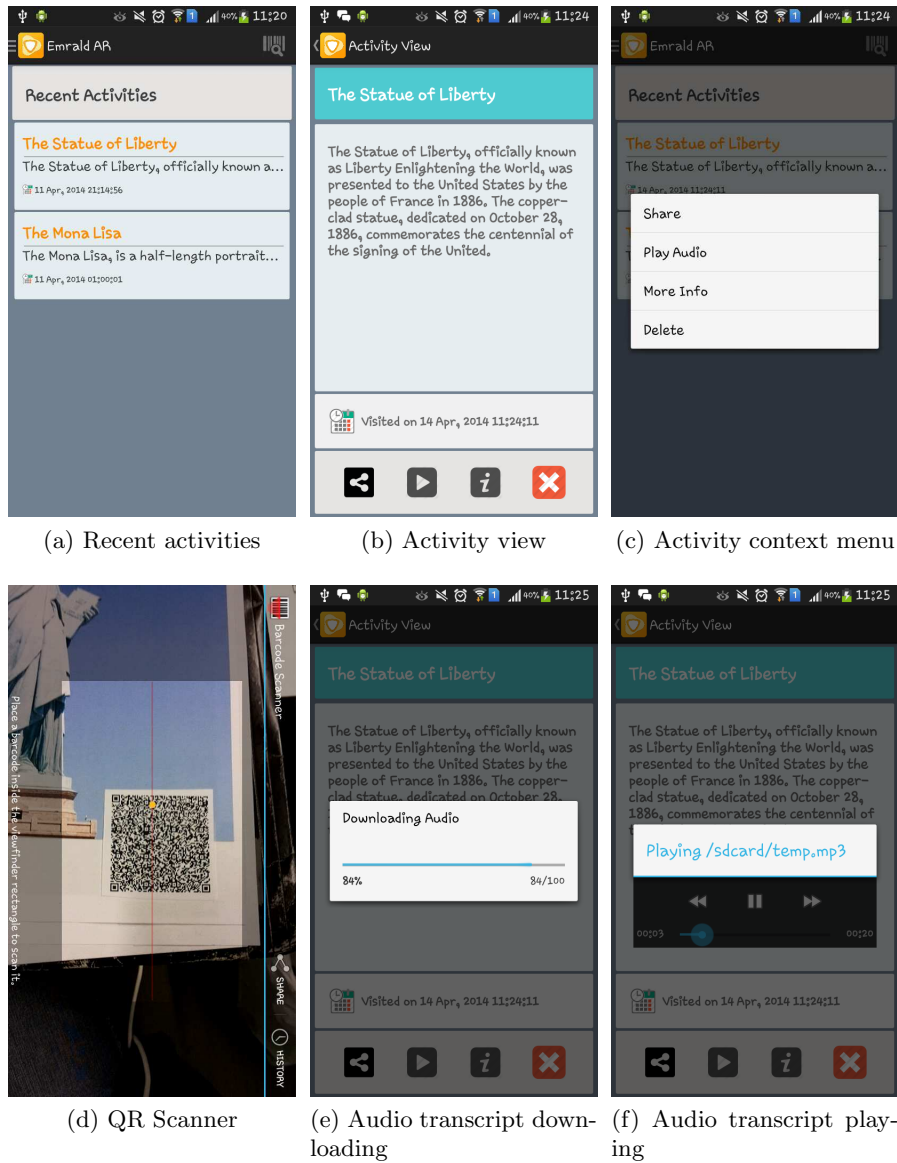
Figure 2.1: Screen-shots of the proposed mobile application.

2.4 Usability evaluation

Usability involves number of aspects and consists of various factors such as:

Functionality: how the application supports the actual tasks of the users.

Easy to learn: how easy is learning about the application for various user groups.

Easy to remember: the effort require to remember how to interact with interface interactive system.

Effectiveness: the effectiveness of the application to reach user goal.

Efficiency: the resources necessary for the user to interact with the application.

Satisfaction: is the user satisfied with the application.

Understanding: how easy it is to understand what the application performs; this is important in unusual situations such as errors while carrying out a task with the application.

Design standards: how the system can adapted to a standard design interface.

Currently, there are several evaluation methods usability in general vary in the step of development that are applied in the manner of collecting data and characteristics. The type of analysis is performed based on the data obtained.

Usability testing refers to activities performed with users or methods of observation, involving the participation of users to perform the evaluation, whereas *usability inspection* is carried out by specialists without mandating the direct participation of users. Compared to usability testing with users, the use of usability inspection techniques is an option of less expensive in terms of cost and time. Inspections can be carried without the presence of the target user. However, the inspection does not replace user evaluation and should preferably be applied prior to evaluation by the user end, so that problems can be identified resolved before the product's release to the general public [33]. Thus, usability testing is important for collecting not only the opinions of users on the system, but also issues related time to the user play activity, number of clicks needed to achieve activity, among others.

The proposed AR interface for museums is evaluated in two sessions: the first being a cognitive walk-through and the other the application of a questionnaire. The questionnaire-based session and cognitive walk-through were performed by simulated museum visitors in laboratory conditions.

Cognitive walk-through   In cognitive walk-through, one or several evaluators inspect a user interface by assigning set of tasks and evaluate whether it is comprehensible and assess the ease of learning. A pre-defined set of questions is presented to each evaluator to obtain detailed information on their interaction with the interface. In our study, The evaluators going use a working prototype of the application.

The inputs to the cognitive walk-through rely on user profile, users̀ knowledge of handling application on *Android* mobile application and the task assigned to the user. This technique is useful in the design and development stages. It also holds potential to be applied in other stages of development such as coding and testing. Cognitive walk-through also works for early mock-ups of interfaces [20], as it does not require a fully functioning prototype.

Methodology   Each evaluator interacts with the AR interface and perform a set of associated tasks. Each task involves the action that seems likely to make progress toward the goal. Each action will be evaluated for feedback of the progress which is being made toward the current goal. Evaluation is perform for the desired information and whether they found the correct task. The steps are repeated for multiple samples to achieve a series of goals that define the completion of the task.

For the evaluation, the following activities are carried out:

- Record the correctness of the assigned mentioned tasks.
- Record the progress which is being made toward the current goal.
- Provide satisfaction assessments.
- Provide suggestions for improvement.

Participants   The evaluation was performed with four participants, aged between twenty to thirty years old. All of them had an experience of visiting the museum in the past. No participants were involved in the technical development of the AR interface for museums. The usability evaluation does not require participants to have any previous experience with an AR *Android* smart-phone application. Most of the participants were average users and are familiar with *Android* smart-phone.

The participants were asked to complete a set of representative task scenarios presented to them in an efficient and timely manner as possible; they were also requested to provide feedback regarding the usability and acceptability of the user interface. The participants were addressed to provide sincere opinions regarding the usability of the application, and to participate in post-session subjective questionnaires.

Instrumentation   In the present study, a *questionnaire* and *snap-shots* are the main instruments of analysis of the application. The questionnaire was designed to record their general experience with the AR interface for museums, the overall user reactions, as well as the learnability of the application interface. The information collected from users enables identifying strengths and weaknesses of the application.

Questionnaires allow to collect data on the basis of profile of the users (pre-test) and the thoughts and emotions about the interface (post-test). Method is a relatively cheap to evaluate an interface. Questions should be as accurate as possible in order to achieve clear answers. To obtain greater precision from the responses of users, it is desired to standardize the answers to some scale of measurement.

Snap-shots are recordings of all user actions on video files (capture screens and the audio) for posterior analysis. Video equipment, relatively low cost, which allow to record the direct observation of user sessions for later analysis. However, the complete video analysis requires specific equipment with many cameras focused on the user and screen. However, the write operations of a selected user can be useful to detect problems.

Sometimes, notes are useful to record user behavior while executing task on the application. Notes require special attention which should be recorded without knowledge of user. None of these techniques is completely safe collections and are probably sufficient to detect all user interface problems. However, they can be used on a group of volunteers before the application is released without a large expenditure of resources.

Variables and hypotheses   The application prototype has been created to perform a usability evaluation on the AR interface for museums. This study focuses merely on the comparison of the assessments of different participants. This study is an initial analysis to get concrete suggestions for the improvement of the AR interface for museums. The application prototype has been evaluated using fewer samples assumed as a museum artifacts. The study requires more extensive time and resources to be evaluated on real time cultural objects of museums. All participants considered in the evaluation are average users with basic experience with *Android* smart-phone.

2.5   Results

During the usability evaluation of the application prototype with users, a multitude of data was collected from questionnaires and snap-shots. The consent of the users for filming was requested in written form (cf. Appendix B.1 on page 90). Considering the diversity among users, each user provided a unique set of traits which reflects behavior on the application prototype
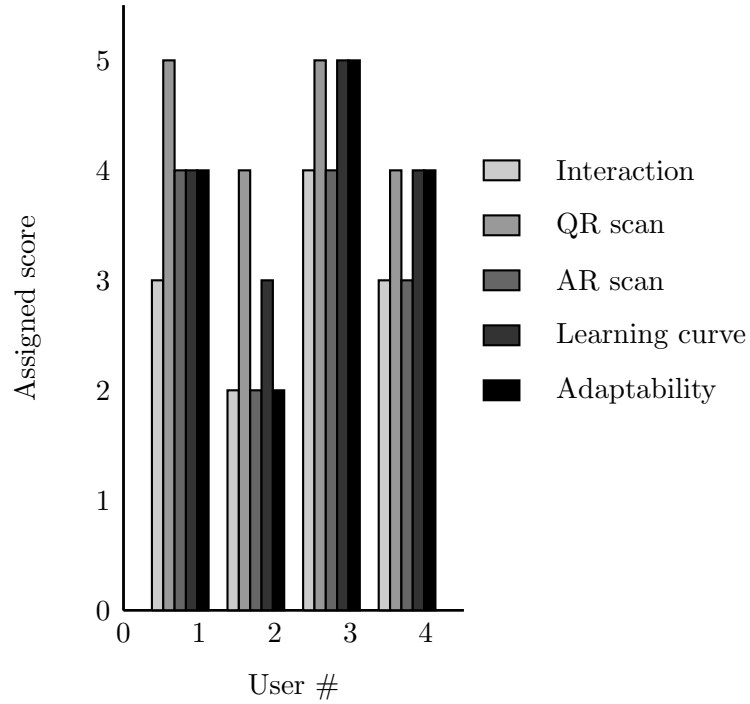
Figure 2.2: Results of the snap-shot analysis per user and metric.

analysis.

The snap-shots were analyzed and the information was classified on the scale of one to five. Figure 2.2 reflects the graphical analysis of snap-shot analysis. The scale considers the factor as zero being the lowest and five being the highest. The classification has been performed based on user's pattern toward application interaction, the ease of QR-scan usage, the ease of AR-scan usage, the learning curve towards application and adaptability.

Application interaction depends upon the use case of smart-phone in the daily life of each user. It has been observed from the snap-shots, the users who used smart-phone actively in their daily life scored high in the application interaction evaluation. One out of four users scored well considering the application interaction and its design standard as per the *Android* framework. Application interaction subjected to various other factors which is hard to comment considering the small set of test users. Considering the each of QR scan usage, three out of four users score well in this classification. The QR scan is quite known to the users and everyone know the usage of QR code. It reflects the easy adaptation of usage of QR scan in the application.

Comparing the pattern observed in the ease of AR scan-usage, the pattern are more contrasting in context to QR scan. The AR scan was new to the every user. The difference in the scale of each user reflects the property

46

of being a new feature to the application. The user was amazed to see the overlay video on the top of artifacts. Each user behaved differently on the scale while observing ease of AR-scan usage. The design of the application was kept simple to decrease the learning curve. Three out of the four users showed a quick learning curve to adapt to the application and scored above four on the scale. The easy access to the recent activities and quick menu to interact with different options was the plus point for the application learning curve. Considering the overall performance of each user, it is hard to mark the scale on adaptability towards application. As per the snap-shots, most of the users quickly adapted to the application design and use case of the application constructs. Most of the users score well in this classification as well.

The usage of the smart-phone and the pattern to interact with application of similar design considered as the major factors which determine user adaptability towards the application. The standard design guidelines as per the *Android* framework in the application reflects the ease of use of the application. A good application design can also reduce mental and physical stress, reduce learning curve, improve operability in using the device and thus improve overall product quality [12].

The questionnaires provided significant statistics for the use case of the application (cf. Appendix B: Section B.2 is the background questionnaire and Section B.3 the post-test questionnaire). The post-test questionnaire was based on the different aspect of application and questions of different context were asked to under the user requirements. The questions involve evaluation of different features provided in the application, some suggestions and recommendations from the user's perspective. The users provided a constructive analysis which was organized into graphical results in Figure 2.3.

The classification has been based on feature provided in the application and user evaluation of feature as per their daily life pattern. The audio transcripts and video overlay was a main feature liked by all the users. Three out of four users strongly liked the features in the application. Half of the users believed the medium usability of sharing feature in the application, whereas other two users accepted it as a useful feature. Sharing depends on social network usage by the user in their daily life. The handling of recent activities and deleting history was considered as good by three out of four users. The more information feature let users explore more about artifacts on the Wikipedia website. It is considered as a relatively useful feature as per the user pattern. Overall, the provided menu options liked by most of the users. The features provided in the menu options considered as an optimum and required as per user pattern.

The pattern obtained from the questionnaires and snap-shots are useful to refine the application. The inputs and suggestions obtained from the users

(a) Menu.

(b) Share.

(c) Audio and video.

(d) More information.

(e) Delete recent activities.
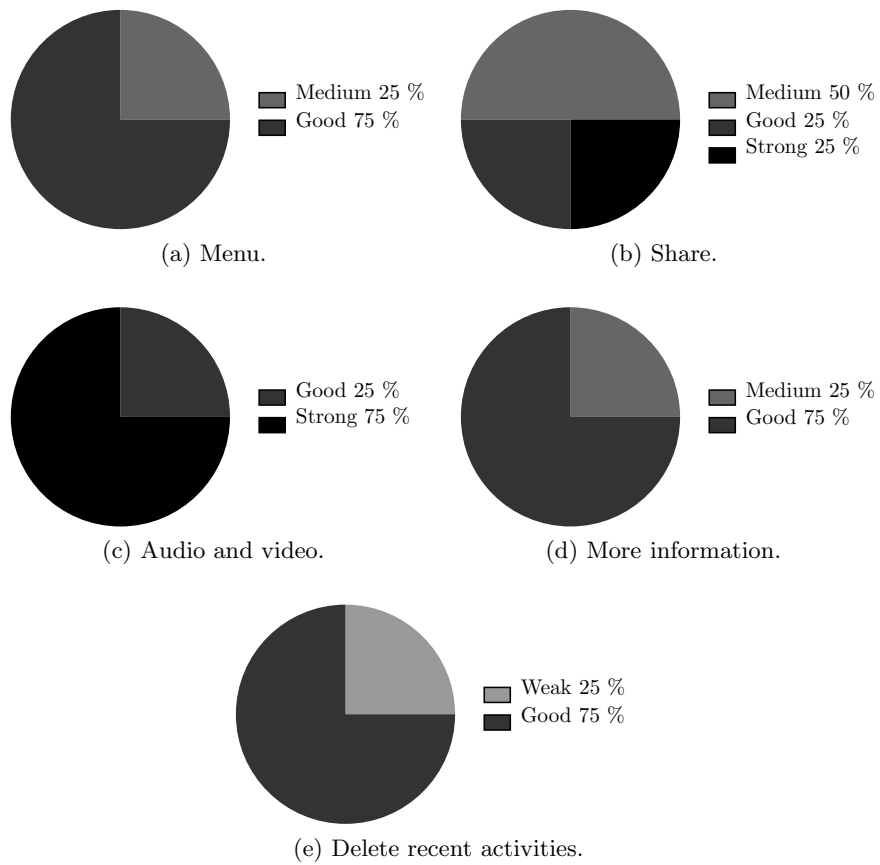
Figure 2.3: Feature ratings on the following scale: poor (lightest color), weak, medium, good, and strong (darkest color).

are used for the early evaluation of the application. The more subjective evaluation is required to make a constructive view of complete usability analysis of the application.

2.6 Conclusions

In this work, we have performed user evaluation of Augmented Reality interface for museums. We created an Augmented Reality interface for museums for the visitors across the world. Our main contributions are showing how the information retrieval concepts can be put to work for serving visitors to ease their visit to museum. The information in the form of text, audio and video streamline the usability of application by users of different age groups. We provided a simple yet practical option of scanning a QR code and Augmented Reality around artifacts. The usability evaluation provided the required feedback of handling of application for the visit to the museum. The empirical study of usability evaluation methods justify the use of these methods to incorporate into different stages of product development.

2.6.1 Discussion

Some of the observations of the study need further discussion. The museum visitors have different aspects of accepting information in different modes of multimedia. Some prefer video over audio whereas others prefer audio and text. The application requires personalization of the multimedia according to a user profile. The results are very encouraging considering the scanning approaches of QR and Augmented Reality markers. All of the participants projected a moderate or a strong positive attitude towards ease of feature usage in the application. One of the user, suggested to have a single scanner for both QR scan and Augmented Reality marker scan. Since, holding a device towards a cultural object for a long time is a cumbersome task, some users suggested a feature to pin the video or download the video to the smart-phone. Most of the users agreed on the concept of providing a utility which eases their learning experience of a museum visit. The observations are quite significant to provide the best experience to the museum visitors.

2.6.2 Future work

Future work includes the full analysis of Augmented Reality interface for museums. It includes formalization of the qualitative aspects by adding suggested features to the application. The application requires more enhancements to address the user requirements of rich experience of a museum visit. The server interaction is subject to testing in context to museum curators. Future work also includes adding a more advance server interaction

to feed dynamic information relevant to the cultural objects. Considering the advent of mobile technologies in all aspects of our everyday life and exponential growth of Augmented Reality technologies, the future holds the potential to provide a rich multimedia experience to the museum visitors.

Acknowledgments

Chapter 3

Usability testing and heuristic evaluation of the Wikitude navigational application

Ondrej Perutka

Since augmented reality mobile applications depend on surrounding environment, usability testing in a laboratory environment is not always applicable for them. In addition, expert usability evaluation can be used to discover general usability problems which may be ignored by test subjects. This paper describes modified heuristic evaluation and a real environment usability testing method applicable for augmented reality navigation applications. Both methods are applied on application called Wikitude. Results of these evaluation methods are presented.

3.1 Introduction

Augmented reality (AR) applications for mobile devices became popular in the last few years. Mostly thanks to the growing market with highly available smart devices. Standard usability evaluation techniques are not always suitable for these applications because of their often context dependency. This paper starts with a brief description of a representative of these applications – Wikitude. In the following sections there is a description of the used evaluation methods, their modifications, testing scenarios and test results.

Wikitude is an AR navigation for Android platform. It has many categories of points of interest (e.g. sights, hotels, restaurants, shops). The application allows users to list a set of nearby points of interest (list-view), show them on the top of a local map (map-view) and show them as labels on the top of a live video feed (live-view). There is also a search functionality which allows to search for a particular point of interest. The application depends on external web services such as Wikipedia, Yelp.com and Hotels.com. It uses GPS and mobile networks to determine user's location and electronic compass to get user's orientation.

3.2 Evaluation

Two methods have been chosen for purpose of usability evaluation of Wikitude – heuristic evaluation and testing in real environment by observing. The heuristic evaluation is targeted to discover general usability problems that might be ignored by real users but are important for their overall satisfaction. The real environment testing should discover common user habits and patterns of using the application. It should verify the efficiency of using this application. The evaluations were designed based on the works of Van Elzakker et al. [42] and Wichansky [46].

An adaptation of the original Jakob Nielsen's heuristic evaluation [30] was made to modify the heuristics for purposes of evaluation of the Wikitude user interface. The modifications are shown in the appendix for this chapter as Table C.1 on page 94.

Two testing scenarios for testing in a real environment were prepared for a representative group of databases and/or points of interest categories. The scenarios cover neighborhood discovery (e.g. sightseeing, looking for a restaurant) as well as searching for concrete points of interest (e.g. shops, bus stops). The goal of this testing was to analyze overall efficiency of use and usage patterns. Data from testing were recorded in form of written notes.

The objective of the first scenario was to compare efficiency of Wikitude and printed tourist guides and maps (publicly available at airports, railway stations etc.). The task was divided into two parts.

In the first part the test subject had five minutes to go through a given set of tourist information in paper form (see below) and another five minutes for using Wikitude. The order of these two sessions was randomized by flipping a coin. Each time the test subject had to find as many sights in a reasonable distance (approximately two kilometres) as possible. The initial position of the test subject was at Central railway station in Helsinki.

In the next part the test subject went from the Central railway station to the Helsinki Cathedral and from the Helsinki Cathedral to the Uspenski Cathedral (approximately same walking distance in both cases). In both cases he had to use both – the printed map and Wikitude. The printed map was used for one half of the path and Wikitude was used for the other half. The order was also randomized by flipping a coin. The order for the path from the Helsinki Cathedral to the Uspenski Cathedral was reversed (i.e. since the order for the path from the Central railway station to the Helsinki Cathedral was: *first use the printed map and then Wikitude*, the order for the second path was: *first use Wikitude and then the printed map*).

These are the exact instructions for the test subject that were presented in oral form:

1. Take this map and this tourist guide and try to find as many sights as possible. The sights should not be further than approximately two kilometres (current location was shown). Tell me each time you find something. You have five minutes.

2. Take this cell phone and open Wikitude. Try to find as many sights as you can. They should not be further than approximately two kilometres. Tell me each time you find something. You have five minutes.

3. Take this map and go to the Helsinki Cathedral (current location and location of the cathedral were shown).

4. *After one half of the path* – Give me the map and use Wikitude for the rest of the path (location of the cathedral was preset).

5. Go to the Uspenski Cathedral, use Wikitude for navigation (location of the cathedral was preset).

6. *After one half of the path* – Give me the cell phone and use the map for the rest of the path (current location and location of the cathedral were shown).

The evaluation metrics for this experiment were:

1. Number of sights discovered using the printed tourist information and using Wikitude.

2. Amount of time required for navigating from one point to another using the printed map and using Wikitude.

3. Number of mistakes done while navigating using the printed map and using Wikitude.

4. Number of times the test subject consulted the map or Wikitude while navigating from one point to another.

The following resources in paper form were available (both obtained at the Central railway station): a *Helsinki City Tourist Map* by City-Opas and a *Helsinki Visitors Guide.*

The objective of the second scenario was to determine efficiency of Wikitude search. The test subject was intended to find the nearest stationery. The subject's initial location was Kumpula Campus (in front of the Physicum building). The evaluation metric was number of search results explored before finding the required result.

These are the exact instructions for the test subject, presented in oral form:

1. Take this cell phone and open Wikitude. Use the search bar to find the nearest stationery. Imagine you want to buy a notebook and a

pen. Tell me each time you access a search result to get a detailed info until you find a shop where you can buy these things.

3.3 Results

In this section we document the problems identified in the inspection as well as the results of the user tests.

3.3.1 Results of the heuristic evaluation

Violations of the rule 1. *The application should always keep user informed about what is going on. The information about the used database or a category of points of interest should be always visible as well as location tracking status and orientation. All information should be given through appropriate feedback within reasonable time.*

1. The graphic symbol used to show user's current location in the map-view can be covered by labels related to points of interest (see Figure 3.1). It is hard to find the user's location in such case.

2. The application shows only camera live-view after selecting some categories of points of interest (e.g. Hotels.com or Yelp.com). There are no controls or system status visible.

3. Very slow response times in some cases (mostly while switching between various categories of points of interest).

Violations of the rule 8. *Screens should not contain information which is irrelevant or unneeded.*

1. There are buttons for following the application on Facebook and Twitter mixed into featured points of interest categories. It may be confusing.

Violations of the rule 9. *All errors should be in a clear, user-readable form and they should suggest a reasonable solution. If an error happens, user should be notified. There should be no "silent" errors.*

1. The application stops responding randomly when switching between live-view, list-view and map-view. This happens also after returning from idle mode.

Violations of the rule 10. *Provide some help and/or documentation. Both a quick start help and a detailed documentation should be available.*

(a) Covered.   (b) Visible.
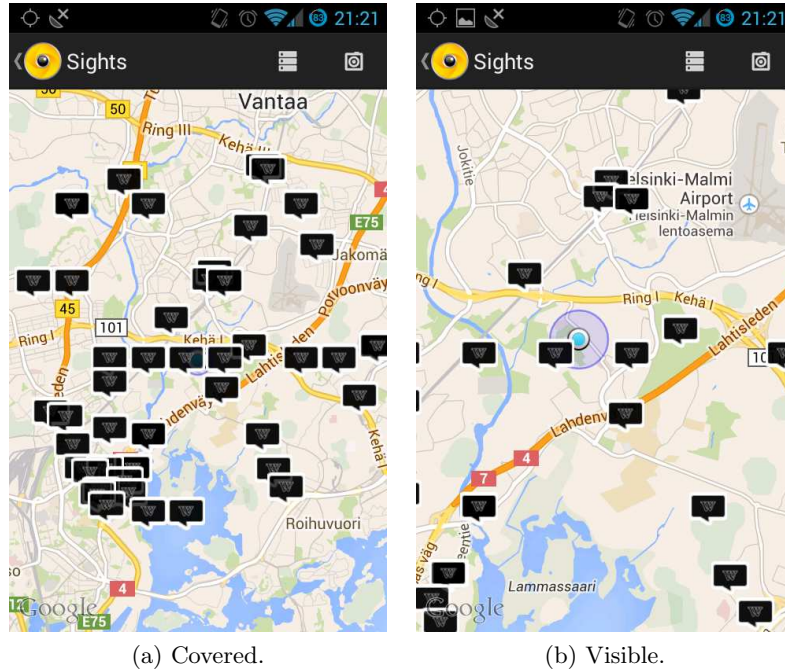
Figure 3.1: Location indicator covered by labels and uncovered after zooming.

1. There is no help. The provided demo does not work. In addition its usage requires to open a demo web page on a screen of another device and scan it using the phone camera. The online demos describes only general features of AR interfaces, not the Wikitude itself.

3.3.2 Results of the testing scenarios

Table 3.1 shows the sights discovered by the test subject during the first part of the tourist testing scenario. Surprisingly, the difference between the numbers of discovered sights is not big. There were basically three reasons:

1. Not all results shown by Wikitude could be considered as sights from a tourist point of view.

2. There were some duplicities between the results (same sight, different name or different language mutation of a corresponding Wikipedia page).

3. Wikitude has limited amount of points of interest it can show. The maximal amount is 50, so the application shows 50 closest sights. It is hard to find more distant and possibly interesting sights in areas with high density of sights (such as Helsinki city center).

55

Table 3.1: Sights discovered by the test subject.

Name of the sight	Method of discovery	
	Wikitude	Printed map
Amos Anderson Art Museum	•	−
City Hall	−	•
Finnish Museum of Natural History	•	−
Finnish National Theatre	•	−
Freemason's Grave	•	−
Glass Palace	•	•
Helsinki Cathedral	−	•
Helsinki Central railway station	•	•
Helsinki City Art Museum	•	−
Helsinki Old Church	•	−
Kaisaniemi Park	•	−
Kiasma Art Museum	•	−
Market Square	−	•
National Museum of Finland	•	−
Old Market Hall	−	•
Parliament House	•	•
Presidential Palace	−	•
Saint Barbara Altarpiece	•	−
Senate Square	−	•
Sibelius Academy	•	−
Temppeliaukio Church	−	•
Three Smiths Statue	•	−
University of Helsinki	−	•
Uspenski Cathedral	−	•
Virka Gallery	−	•
Total #	15	13

Another interesting observation is that the sets of sights discovered using the printed map and sights discovered using Wikitude are almost disjoint.

Tables 3.2a and 3.2b show metrics captured while navigating from the Central railway station to the Helsinki Cathedral and from the Helsinki Cathedral to the Uspenski Cathedral. All values are normalized by distance in hundreds of meters, corresponding raw values are in parentheses. The actual paths of the test subject are shown in Appendix C.2 on page 97.

It is worth noticing that in the first case (i.e. the way from the Central railway station to the Helsinki Cathedral) the length of the path navigated using the printed map is twice as long as the rest of the path. It is because of a mistake done by the test subject. He chose a longer path (see the map in Figure C.1a in the appendix on page 97). The number of consultations is also higher in case of the printed map because of the same reason. The normalized time is almost equal in both cases. In the second case (i.e. the way from the Helsinki Cathedral to the Uspenski Cathedral), all metrics are almost equal for both the printed map and Wikitude. Interesting fact about this path is that the test subject was actually using Google Maps instead of

Table 3.2: Results of navigation.

(a) Path from the Central railway station to the Helsinki Cathedral.

Metric	Value
Map navigation (first half)	
Number of consultations	0.57 (4)
Number of mistakes	0.14 (1)
Total time (m:ss)	1:22 (9:32)
Length of the path	700 m
Navigation using Wikitude (second half)	
Number of consultations	0.29 (1)
Number of mistakes	0
Total time (m:ss)	1:18 (4:33)
Length of the path	350 m

(b) Path from the Helsinki Cathedral to the Uspenski Cathedral.

Metric	Value
Navigation using Wikitude (first half)	
Number of consultations	0.69 (2)
Number of mistakes	0
Total time (m:ss)	(1:24) 4:05
Length of the path	290 m
Map navigation (second half)	
Number of consultations	0.36 (1)
Number of mistakes	0
Total time (m:ss)	(1:29) 4:10
Length of the path	280 m

Wikitude. He used a feature of Wikitude which allows to visualize a path to some destination using Google Maps.

Unfortunately it was not possible to complete the second test scenario due to an unexpected problem with the Yelp.com database in Wikitude which rendered useless the search results as there was no shop among them.

3.4 Conclusions

The testing revealed several usability problems of Wikitude. The most noticeable one was the overall instability of the application causing application crashes in many cases and some not working databases. Another significant problem was that the application has limited number of points of interest it can show. The number cannot be adjusted, so the application shows only a small surrounding in places with high density of points of interest from a selected category. There were also several minor usability issues.

The testing scenarios were designed for educational purposes. It would require additional testing in a more complicated environment to state some conclusions about the efficiency of navigation using Wikitude.

Chapter 4

Tactile navigation belt: Wearable device for sensing the direction

Aki Kesulahti & Jorma Nieminen

In this work we introduce our prototype device which gives tactile feedback for navigation. We show our user test scenarios and evaluate the test results. The device is a mobile and hands-free alternative to other displays like visual displays in hand held navigators. Our goal is to find out useful use cases and to research if user can feel the direction correctly. Secondary goal is to collect feedback on how to improve the idea of indicate direction or other data through vibration.

4.1 Introduction

Satellite based positioning technology such as GPS (Global Positioning System) has made location awareness based on navigating popular and widely used [2]. Modern navigators are based on maps and other features which helps users to find their goal. Most of these devices uses visual displays to show all the information on screen and possibly some information are repeated with audio. This setup is simple but if one is moving by walking that usually means that one has to hold the device in order to see the information. However in some mobile use cases one might need both hands for other tasks or just feel more comfortable if it is not necessary to carry the device continuously. Properly implemented tactile feedback such as vibration can give this needed information without visual display.

There are several different forms of wearable interfaces. The easiest way to use a device is to implement the device to some clothing or equipment one is already wearing such as glasses, gloves, belt, watches, shoes, caps etc. But not all of these are very suitable for giving tactile feedback. The main requirement to sense direction is that user normally will be physically in the position that it is possible to feel every horizontal direction. In some devices like watches the hand movement is too major to give reasonable feedback. So if the user is standing or walking or sitting the best option would be some

kind of band that covers all directions [41].

A second requirement is that if the feedback is given by vibration motors it is not trivial to feel very precisely the spot on skin [8]. For example we tested attaching the device to ankle, waist and chest and found out that out of these three the chest was the most responsive to sense vibration. In our work we evaluate the usability of a tactile navigation belt. We test the belt with real users as they perform various navigation related tasks. The navigation belt prototype that we use in our tests consists of four vibration motors and a console-game controller connected to the motors. The usability evaluation methods we use consists of usability testing, talk aloud protocol and questionnaires. During the tests we also gather some relevant metrics. After the tests we analyze and discuss about the results and propose some future work.

4.2 Previous work

There are several previous works that uses wearable device such as 5DT Data Gloves [1] for motion capture, Tactual Wearable Display [10] in vest and SuperShoes [10] for tickling interface. One of the really similar work is that of Tsukada and Yasumura [41] who evaluated the performance and usability of a tactile belt system prototype called ActiveBelt. Their prototype unit consisted of a direction sensor, a GPS, eight vibration motors and a microcomputer for controlling the other devices.

4.3 Tactile navigation belt

We call our device a tactile navigation belt. We thought of some shorter and more memorable names but it seems that most of great names like NaviBelt [41] and have already been taken. We ended up to keeping the most describing name: tactile navigation belt.

4.3.1 Concept

The idea of this belt is that a user can feel the direction or other haptic coded information. The simplest way is just to vibrate the motor which is closest to the goal. Optionally the motors can indicate distance with accelerating vibrates-per-minute ratio when approaching the goal. The ideal belt could have these features:

- Wearable belt-like device.
- Tactile feedback to give user intuitive information, most cases direction.
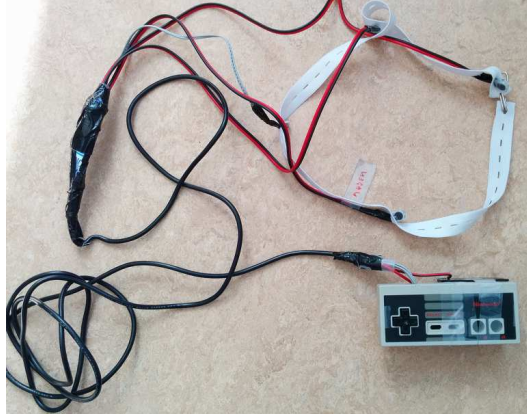
Figure 4.1: Working prototype.

- Multiple vibration motors. Could be four, eight, or even 16.
- Built-in battery.
- Optionally built-in data link like WiFi or Bluetooth — if used with other devices like mobile phone.
- Optionally built-in compass, GPS, accelerometer, or gyroscope — if used as standalone device.

4.3.2 Prototype architecture and development

In the test scenario a vibration belt first prototype is used. Unlike the concept, this prototype does not have any data link or build in battery. It contains a controller unit which is connected to belt via cable (Figure 4.1). The belt unit has four coin-like vibration motors. The controller unit is used by another person to simulate an intelligent software that would know which way to go.

In order to let someone give tactile feedback to the belt we chose a *D-pad* (short for directional pad) [38]. It is a four-way (up, down, left, right) directional control with one button on each point, very common in game consoles. These d-pads buttons have just on-off states so one can't signal any different values like modify motors voltage through the button.

In addition to just four-ways it is possible to use combinations of two near by directions (up and left, for example) to provide diagonal directions. The prototype controller with D-pad is a modified Nintendo Entertainment System controller. In evaluation process this controller can be used by one hand so the operator can video record the pressings with another hand.

The vibration motors we are using are removed from another device. We measure that this device uses 4.16 volts to operate motors. We tested

Figure 4.2: A user wearing the prototype on chest.

different battery setups and notice that the most noticeable vibration can be reach with three AA-sized batteries in series, 4.5 volts in total (Figure 4.3). In early testing we discovered that the vibration motors that we were using was somewhat too weak to be recognized properly if the belt was attached to waist. Also we notice that regular inflexible belt didn't give enough pressure between the skin and the motors. So we ended up improving the sensing by attaching motors to an elastic band and attaching it to chest, like heart rate monitoring belts (Figure 4.2).

4.4 Evaluation

We use usability testing as our main method for the evaluation. To carry out the usability evaluation we need the working prototype, two video cameras and two operators. There are two printed forms that test user fills: one before test scenario and the other after testing. First the user is given the belt to wear. After this the first evaluator gives a general introduction about the test and its purpose and also a short description of an easy example. After the introduction the user performs the example task. The purpose of the example task is to test that the belt is working correctly and that the user has understood the guidance.

Next the user starts to perform the real tasks one by one. While the user is performing the tasks, the first evaluator walks behind him or her and controls the motors. Vibration information are given in pulses. Pulses are supposed to be stable and between 50 and 200 ms. Longer durations may perceived as being irritating [24].

The second evaluator captures the overall performance on video with a smart phone. The evaluators do not give any guidance on how to perform
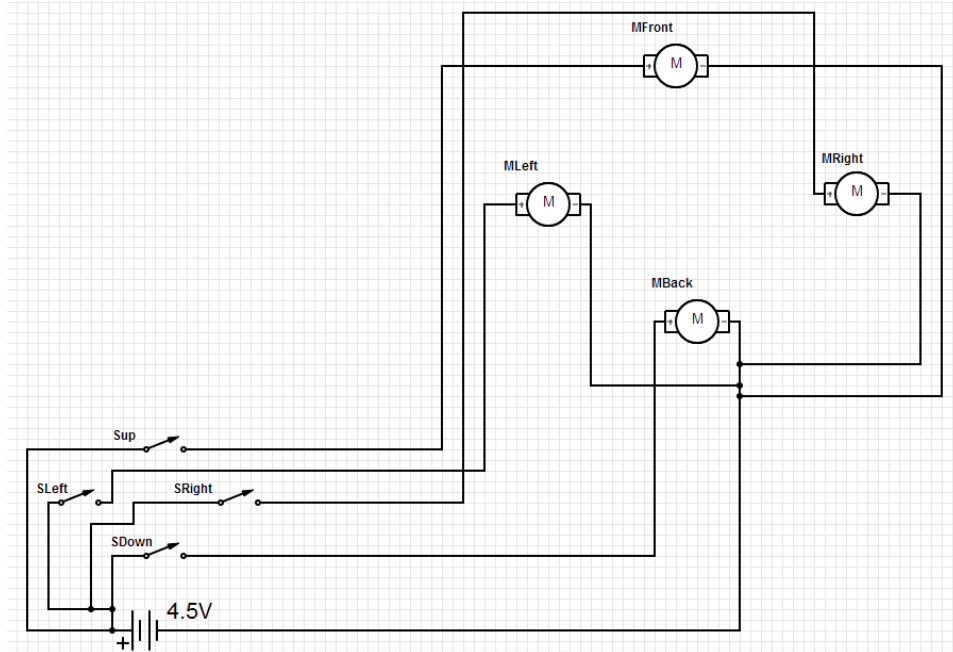
Figure 4.3: Electrical schema of prototype.

the task while the test is going on. While the user is performing a given task, it is possible for the user or either one of the evaluators to stop the task if it seems that finishing the task has become too difficult or impossible.

4.4.1 Tasks

Each test subject perform twos tasks. Tasks descriptions are given on a paper so the tasks would be as similar as possible to each user. The task info sheet is written in Finnish. Here are short translations for tasks:

Task #1: Find a room in a conference

> You are in a conference held at Exactum building. The last session has just ended and you need to find the correct room for the next session. The device is smart; it knows precisely when to turn.

Task #2: Find a rental car

> You have rented a car. It is located somewhere near this building. It is a green Volkswagen Polo. Find the car with the

help of this device. The device is not smart; it acts like a compass that points to the car.

These tasks represent different use cases. In the first task, a more advanced prototype could use special hot-spots inside building to know its location in pre-installed map. The second task simulates a use case where device would know only one direction and keep indicate that direction no matter possible obstacles. Our goal is to figure out how user will react different kind of haptic guidance and are those directional information useful.

4.4.2 Protocols

For evaluation of the usability of the developed prototype, we mainly focus on two protocols: think-aloud protocol and questionnaires. In addition to this we also observe test users reactions during tests and afterward from video records.

The *think-aloud protocol* is a usability testing protocol that is quite popularly used for its simplicity and effectiveness [14]. In this protocol real users are thinking aloud when testing the product meanwhile developers observer their process. Users are asked to say whatever they are looking at, thinking, doing, and feeling as they go about their task. This enables observers to see first-hand the process of task completion Protocol is used to gather data in product design and development [14]. This data allows developers to understand the thought process of a subject as they use the product. The think-aloud protocol was introduced in 1982 by Lewis [23].

In our evaluation process we give an introduction to the think-aloud protocol to each user before the test. In order for the test scenario and guidelines to be as identical to everyone as possible, this introduction is in printed form.

A *questionnaire* is a tool for research which consists of a series of questions [16]. Questionnaires are often designed for statistical analysis but we are using it more like written form of interview. Questionnaires are simple, cheap and easy to create and fill. Numerical data is comparable and free field questions are easier to analyze than interviews because it is already in written form [16]. In our test process we give two questionnaire forms to test users: pre-test and post-test questionnaires.

4.4.3 Performance metrics

Performance metrics are used to obtain quantitative data about test participants performance when they perform the tasks during usability test [7]. Performance measurements procedure can contains define the goals for the usability testing. Our goal is to test if this system works or not and if it does

Table 4.1: Navigational aids that each user normally uses, as answers to the question "When I need to find my way to a new location I use...".

	User #1	User #2	User #3	User #4
Paper map	✓		✓	
Phone navigator	✓		✓	
Compass			✓	
Journey planner		✓		✓

is there any improvements. To gather comparable data one must quantify usability issues by measurements. The metrics we gather during the tests are:

1. Time to complete the task.

2. Time it takes for the user to react to the vibration.

3. The number of times the user doesn't react to vibration.

4. The number of times the user turns to a wrong direction.

5. The number of times the user fails to complete a task.

We use two video cameras to capture the test scenario and to measure and analyze these metrics after the actual tests.

4.5 Results

There are four test users, one female and three males, who have never used this device before. Three of the test users are computer science students and one a math student. Test users are aged from 21 to 37.

4.5.1 Pre-test questionnaire

A pre-test questionnaire is given to test person before we give more details about our device and the rest of the test. The classification information we gather in person's gender and year of birth. Also this questionnaire includes some questions about user current habits about what kind of devices and how often he or she uses them in everyday life.

Table 4.1 shows that half of the users prefer a paper map and phone navigator whereas the other half prefer the journey planner[1] Unfortunately there was no related question are they normally using a vehicle; there might

---

[1]Helsinki Region Transport provides a web service called *Journey Planner* at http://www.reittiopas.fi/en/.

Table 4.2: Performance metrics. The times are indicated in minutes and seconds (m:ss); the last column is the average over the set of users.

Performance metric	User #1	User #2	User #3	User #4	Avg.
Task #1 completion time	1:47	1:03	1:58	2:08	1:44
Task #2 completion time	2:29	1:43	2:37	2:42	2:43
Number of failures to reach to a vibration	0	0	0	0	0
Number of wrong-direction turns	2	0	1	0	0.75
Number of task-completion failures	0	0	0	0	0

be correlation between paper map and phone navigator of the user who prefer own car over public transportation.

"Properties and features I wish a navigation system has:"

- It should show my current location on a map.
- Easy to use.
- Works smoothly.
- Speech-control functionality.

The field of wanted features does not give any new innovations or out-of-the-box thinking before the test. It just list features that are already covered in systems there are in markets.

4.5.2 Performance metrics

Performance metrics listed in Section 4.4.3 are measured after actual tests from video records (cf. Table 4.2).

First thing to notice is that every user successfully completed the given tasks. The completion time of two different tasks were in line with user normal walking speed. Everyone finished first task faster (Table 4.4). Also there was no case when the user would not react to one motor vibrations at all. The average time to react vibration turned out too hard measure reliably from the video records. The operator who used the controller also video recorded the buttons he pressed. However half of the times we cannot determine from the footage the precise time of when the different directions are pressed. It was too much for one operator to record that measure with adequate precision. The error rate of analyzed information was too high so
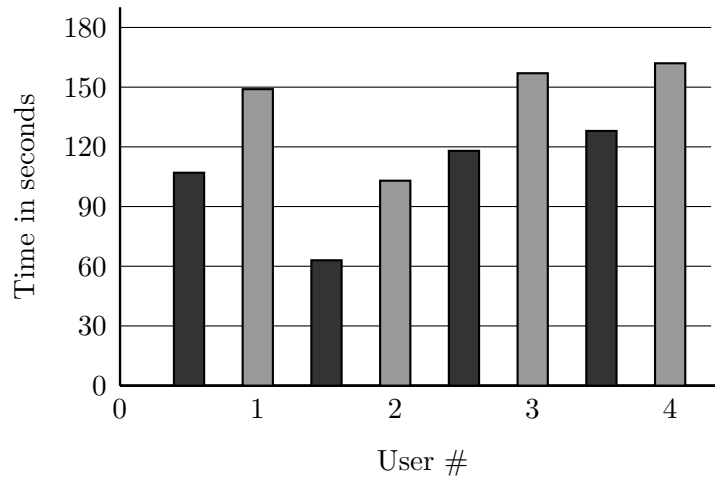
Figure 4.4: Time to complete each task for each user. Task #1 is shown on the left side (dark bar) and task #2 on the right (light bar) for each user.

we end up not measuring that delay at all. User turns to a wrong direction row (Table 4.2) is the most interesting part. Half of the users had hard time to figure direction in one spot during first task. There is lobby with multiple direction to go. The confusing part was stairs in the lobby, right next to correct path. Even now when the operator simulate intelligence it is hard to know how far you should bypass misleading paths that are nearly in the same direction. The last of those three wrong turns were in situation when user thought he or she found the correct car in second task even when the device told otherwise. Correct colored car as of "easy target" was too strong to mislead user for a while.

4.5.3 Post-test questionnaire reviews

Post-test questionnaire contains some review questions that are answered with scale one to five, with the following meanings:

1. Strongly disagree / Never.
2. Slightly different view / Randomly.
3. Do not know / Neutral / Moderate.
4. Slightly agree / Usually.
5. Completely agree / Always.

Table 4.3 shows that most of users felt that four motors are enough to determine the correct direction. User #4 was hoping more motors especially in second task when the device worked like compass. There are is quite a lot of deviation among the answers. However, for some reason, user #3 was

Table 4.3: Post-test questionnaire.

	User #1	User #2	User #3	User #4
I was able to sense the vibration well.	2	4	5	4
Four motors was enough to determine the correct direction.	5	5	5	2
Choosing the correct direction was easy.	4	4	5	3
The belt worked as I expected.	5	3	5	2
The grade I give to the usability of the belt.	3	4	5	2

really satisfied how this prototype worked even after experiencing an issue with finding the correct direction (see Table 4.2).

4.5.4 Issues and observations

Most of users reported some problems with sensing some direction (Table 4.4). The issues are quite different and most likely the cause is prototype's limits to maintain correct attachment near skin. We noticed also that back motor was near backbone and human physical structure affects as well how user can feel the vibration in different spots [8].

In one case user had to stand still to feel the direction when got lost. With one user the front sensor was difficult to sense while walking. Less than 90 degrees turns were reported difficult since the vibrations came in pulses. Two users felt vibration on the left side while the correct direction was to the front right, however we can't be sure if these are phantom feels or flaws from prototype or from its operator. During test we observed the subjects by eye and also evaluated the footage afterward from video records. The observations gave some notably issues. Two users were considering the stairs. One user slowed down when passing a door. On the outside task two users relied on the task description about the green Volkswagen Polo. Those same users first walked to a green Škoda which was a wrong target in easy spot.

Table 4.4: Problems with sensing direction.

	User #1	User #2	User #3	User #4
Front			✗	
Back	✗	✗		
Left				✗
Right				

4.5.5 Suggestions for improvements and other applications

In last part of post-test questionnaire was field question about users toughs to device improvements or possible applications where and when vibration based information might be useful. Users requested the following improvements to current prototype; the number in parenthesis indicate how many person requested it out of our total of four users:

- Stronger vibrations especially to the back (2).
- Different amplitudes for different environments (1).
- User should be able to adjust the motor levels (1).
- Error signal when one is lost (1).
- More motors / "linear" belt (1).
- Static vibration (1).
- The pulse interval for the front motor could be longer when walking straight ahead (1).

Users' suggestion for applications were the following:

- Bachelor party.
- For blind people.
- Biking or motorcycling.
- Generally in situations where looking at a navigator is impossible.
- To alarm when one must look at a visual navigator.
- In a forest, for example orienteering or hunting.
- In Tokyo where navigating around is hard because of a different address system.

4.6 Conclusions

Our prototype did what it was planned to do. Some good suggestion for improvements should be implemented to future work for example bigger motors. Also testing would need better equipment's and possible one more observer to measure users reaction time more specifically. In future work the next prototype should be more advanced for test how the real software can guide users.

Some difficulties were faced during tests that could be avoided in future work. For the operator who uses the controller it was hard to keep same density of pulses. Also it was hard to control motors correctly in fast turnings. And while he at the same time used video camera the footage

quality was poor. Third evaluator could be useful to record this video and further indicators that shows to camera when specific motor is activated like led-lights would make the measurements of performance metrics a lot easier.

As mentioned in results (Section 4.5) there were occassional problems with sensing directions correctly. And because there was human involved in belt behavior we can not be sure if misbehave appear because of "phantom sensation" or flaw in device or use error of person who is using the controller. In order to eliminate the use error the prototype should actually contain a computer guided controller unit. In post-test questionnaire a lot used route search was missing at first in the list of devices one may use. Post-test two last field questions are too large and complex. Those questions could be divided into more specific questions.

Appendix A

Appendices to Chapter 1

A.1 The cognitive start-up sheet

Interface: Augment-3D
Task: Visualize your furniture in your room before purchasing it.
Evaluator(s):
Date:

Task Description: Describe the task from the point of view of the first time user. Include any special assumptions about the state of the system assumed when the user begins to work.

Select a furniture 3D model and place it in your room in a desired place and save it. The system will be in a state such that someone could immediately start testing.

Action Sequence. Make a numbered list of the atomic actions that the user should perform to accomplish this task.

 (a) Open the application
 (b) Choose the desired option
 (c) Choose the desired 3D model
       (-) *Only for smart-phone in vertical position*
       (i) Select CREATE MARKER
       (ii) Read and close help window
       (iii) Perform the scan
 (d) Place the model in your environment in a desired way
       (i) Turn on flash (if required)
       (ii) Adjust the scale of the 3D model
           (1) Make it bigger if necessary
           (2) Make it smaller if necessary
       (iii) Rotate it in the desired location
 (e) Take a photo of it and save the photo

Anticipated Users. Briefly describe the class of users who will use this system. Note what experience they are expected to have with systems similar to this one, or with earlier versions of this system.

People who have experience with smartphones but limited experience with Augmented Reality applications in smartphones. They should have basic knowledge of how to use an application in a smartphone and will have gone through the demonstration of the augmented-reality application.

User's Initial Goals. List the goals the user is likely to form when starting the task. If there are likely goal structures list them.

1. User selects the option to Browse the 3D galleries — success.

2. User selects the option to Scan the 2D catalogues — failure.

A.2 Data collection for the cognitive walkthrough

We assume that the user understands the goal of the app and some has some basic knowledge of how AR works. The steps are named according to Figure 1.6 on page 17. We do not consider here the problem of not getting the appropriate image and the failures due to technological issues as they do not concern to design problems. The scores have the following interpretation:

1. Totally wrong
2. Quite wrong
3. Confusing
4. Quite obvious
5. Totally obvious

The symbol ✗ in Table A.1 stands for "not applicable", meaning that it was necessary to perform the step. In addition to the scores in Table A.1, the observations for each step and question are recorded in Table A.2. Also note that the step (d.iii) can be carried out either by rotating the paper (not related to the app) or by using the rotation function of the app.

Table A.1: Data-capture format used in the cognitive walk-through for the score assignment per question (Q) and step (S).

		(a)	(b)	(c)	(c.-)	(c.i)	(c.ii)	(c.iii)	(d)	(d.i)	(d.ii)	(d.ii.1)	(d.ii.2)	(d.iii)	(e)
1	Will the user try to achieve the right effect?	5	5	4	2	3	5	5	5	✗	✗	3	✗	4	4
2	Will the user notice that the correct action is available?	5	4	5	1	5	5	5	5	✗	✗	1	✗	1	4
3	Will the user associate the correct action with the desired effect?	5	3	5	1	4	5	5	5	✗	✗	3	✗	3	4
4	If the correct action is performed, will the user see that progress is being made towards the solution of the task?	5	4	4	5	5	5	5	5	✗	✗	4	✗	5	5

Table A.2: Notes on the steps (S) and the questions (Q) in Table A.1.

S	Q	Observations
(a)	4	If the user selects the correct option, the selection of models appear which may lead the user to think it is the right direction. However, some users may think that this is not the right way to proceed.
(b)	1	It is clear that the user needs to select one option.
	2	The user may understand that the option is one of the two available options.
	3	There are two options: BROWSE 3D models (the desired option) may be chosen as the task asks for select a furniture 3D model, but also the SCAN option may be chosen as it says that this option is suited to scan an image.
(c)	1	The user will probably scroll down and find the 3D model. The user may also use the search box for finding the model.
	3	When the right model is found, the user will understand the need to click on it.
(c.-)	1	The user has been asked to create a marker and hence may directly try to point the camera to the image that will act as marker.
	2	The option is hidden in the tool bar as only three icons are visible in the vertical position of the phone.
	3	It is very unlikely that the user will realize that it is necessary to scroll the toolbar to find the right option.
	4	The user will see that the next screen is a video capture screen.
(c.i)	1	When the create option appears, the user will probably select it.
	3	It is likely that user will associate the option CREATE MARKER with the correct action.
	4	If the user scrolls the toolbar, the CREATE MARKER option appears in first place.
(c.ii)	3	It is clear that user needs to press ACCEPT to continue.
	4	A help message appears explaining the procedure to create the marker.
(c.iii)	1	The user will try to scan the image.
	3	Colors and text are self explanatory.
	4	A new window with a frame appears.
(d)	1	The user will take the marker to the desired place.
	3	Movement of the marker will show a movement in the 3D model.
	4	The next view is the actual augmented scene.
(d.i)	1	No flash is available.
(d.ii)	3	If users are familiarized with zooming images in smart-phones, they may understand that they can scale with this same feature.
	4	If user scales the model and can probably see that it is being scaled, although small scales may not be noticed. If users are familiarized with zooming images in smart-phones, they may understand that they can scale with this same feature.
(d.ii.1)	1	User may not understand that the possibility to scale the model by pinching the screen.
	2	The option for re-sizing is only explained in the help; unless the user is using the help feature, the option is not visible.
(d.iii)	1	User may rotate the marker so that the object will be rotated.
	2	The option to rotate by using the software is neither visible nor intuitive.
	3	If user finds the option for rotating, it will be possible to understand that it will rotate the model. However, rotation is not the desired.
(e)	1	User may understand that the photo can be taken directly from this view.
	2	The icon is available with an image of a camera.
	3	If the user recognizes the icon, it will be easy to know that it is the button to press.
	4	A message confirms that image has been saved.

A.3 Results of the heuristic evaluation

As instructed in the literature, each evaluator worked independently and recorded the problems found together with the violated heuristics.

A.3.1 Problems reported by evaluator #1

The problems are listed one by one, indicating for each problem the heuristics it violates.

Table A.3: Problems found by evaluator #1 with the violated heuristics.

ID	Problem	Heuristic	Proposed solution
1	Options are not provided when pressing the default button of Android devices	4	Provide options when pressing the button.
2	When pressing back from the "selecting model" window, the application exists instead of going back to the previous window (the initial window). This happens in more places.	3, 4	Provide history navigation through the application.
3	A message appears when trying to use the Scan feature saying that the app is not going to work, but it still works. The same message appears in the AR feature. The message and the options provided create confusion. Also, the app works in Spanish, but this message appears in English.	1, 2, 9	Do not show the error message or clarify the message. Also, provide translation for the message.

Table A.3: Problems found by evaluator #1 with the violated heuristics.

ID	Problem	Heuristic	Proposed solution
4	When scanning a QR code, If QR code is detected but not recognized, it gets a loading dialog and after that a new interface with the view of the camera, a button for taking a photo and a button with an unknown behaviour (when pushing it nothing happens).	1, 5	Provide an error message instead of a new interface.
5	A message appears after some time when SCAN feature has found no codes or images to scan. It says it was not found and offers to try again, but the only option is to push accept and that takes the user back to the main window, but not to really "try again" option.	9	Implement the "try again" option.
6	Saving images from the app is working but the application says that the image has been saved to the gallery before the image appears in the gallery.	1	Either wait until the image is saved to display the message or replace the message with a text saying "the image will appear soon in the gallery".
7	When using the AR feature, the icon bar or tool bar shows 3 icons if using the phone vertically and there is no sign of more icons. However, there are more icons that can be accessed by sliding the finger over the icon bar. When using horizontally, 5 icons appear, but there is no evidence of the availability of more icons (same as before).	1, 4, 6, 7	Show hints for the user to know that there are more available icons.

Table A.3: Problems found by evaluator #1 with the violated heuristics.

ID	Problem	Heuristic	Proposed solution
8	The text of the icons is cropped for some icons.	1, 2, 4	Reduce text or avoid cropping.
9	When trying to create a new marker, the help message appears every time, regardless the feature has been used before and the user is experienced with the app.	7	Show help message only the first time.
10	When trying to create a new marker with an image that the system considers not appropriated, a message saying that it is not a good marker appears, but there is no way to know why it is not a good marker.	1, 6	Provide information on why it is not a good marker.
11	Sometimes, the app switches between "not a good marker" and "too close" which creates confusion as user may still think that the image is valid as a marker.	1	Provide more precise information about the suitability of the marker.
12	"Too close" message is displayed, but sometimes getting closer takes the user to a "good marker". Therefore, the "too close" message creates confusion.	2	Provide more precise information about the suitability of the marker.

Table A.3: Problems found by evaluator #1 with the violated heuristics.

ID	Problem	Heuristic	Proposed solution
13	When using the sharing function, it is not clear what it is going to be shared. The user may think that he/she is sharing an image, but what it is shared is a link to a webpage where an image of the 3D model and a QR code is displayed. They ask the user to scan the QR code to visualize the 3D model in AR, but when scanning the code (with Augment app), it takes you to download Augment app.	1	Provide more info about what is going to be shared.
14	Sharing by e-mail is not always working (it might be problem of a third party app).	5	Check for errors.
15	After trying to share twice by e-mail, the app crashed. After the crash, when trying to open the app again, it keeps crashing.	Bug	Handle crash events.
16	The app works in Spanish, but the model categories are in English.	2	Translate categories.
17	Some messages are using not "natural language". For example: "Usted requiere estar firmado al sistema para disfrutar de esta característica". The Word "firmado" is never used in this context. The appropriate word would be "registrado".	2	Improve translation.

Table A.3: Problems found by evaluator #1 with the violated heuristics.

ID	Problem	Heuristic	Proposed solution
18	Help has been translated to Spanish, but the sentences lack of meaning. For example, "fresco" has no sense here and probably comes from a bad translation of "cool". "Genial"' could be a more appropriated word.	2	Improve translation.
19	When trying to send a marker to an e-mail account, the app asks for the e-mail address, but keyboard shows up very late.	1	Improve behaviour.
20	When sending the marker by e-mail, the user feels he/she is sending his/her own marker, but what is sent is a link to a webpage where they explain the functioning of the application.	2	Explain clearly what is going to be shared.
21	The words "marcador" and "rastreador" are used for the same purpose. It needs coherency here.	2	Use always same notation.
22	When using the help in Spanish, the external help is in English.	2, 10	Provide external Spanish help.
23	There is no option to change the language.	Functional	Provide the feature of changing language.

Table A.3: Problems found by evaluator #1 with the violated heuristics.

ID	Problem	Heuristic	Proposed solution
24	When the app starts, it offers two options (AR and SCAN). Later, in the application options, the SCAN feature is clearly visible, but the AR option is not visible and there is a need to scroll down (even below HELP section) to go to CATEGORIES and enter one of these categories to access the same feature that was easily accessible when starting the app.	8	Reorganize the application options.
25	The list of categories is displayed in the same context as the other options. It may be more suitable to provide an AR option together with the scan option and inside the AR option to display the categories.	8	Reorganize the application options.
26	There are shortcuts, but they require to be registered. For some it may have sense (own models) but for others it should be possible to access without being registered (History of 3D models, favourites).	7	Allow some shortcuts for non-registered users.
27	There are three interaction options: translate, scale, and rotate. In the toolbar, only rotation is available.	8	Provide the other two interaction options as buttons. Another option would be eliminate the rotation from the tool bar.
28	Rotation is possible in one axis only. Translation is possible only along the plane defined by the marker.	Functional	Provide interaction in more axes/planes.

Table A.3: Problems found by evaluator #1 with the violated heuristics.

ID	Problem	Heuristic	Proposed solution
29	There is one button in the main AR view that is not clear what it does (actually, the button resets the model, so it is only possible to understand what is does until some modifications have been made to the model).	1	Provide a hint or better icon for the user to understand the behaviour.
30	The button can be pressed by error (as user is not aware of the behaviour of the button) and it will eliminate the modifications in the model without asking the user or without any "undo" option.	5	Provide an undo option for the action.
31	When adding a second model to the AR view, the changes in translating, rotating and scaling the first model disappear and the model is restarted to the original position.	5	Maintain the position and orientation of previous models.
32	Scaling models when more than one is visible is more complicated than when having only one.	Bug	Improve the control.
33	Adding models to the scene is possible, but removing models is not possible.	5	Provide undo and/or delete option.

A.3.2 Problems reported by evaluator #2

The problems are listed by heuristic.

1. **Visibility of system status:**

   (a) The application does not show what should be an ideal tracker. The user uses the tracker and then does not knows which is the ideal way to create a tracker.

   (b) The application does not why your tablet cannot detect the QR code or **Augment** logo even after you are scanning the right image.

2. **Match between system and the real world**: When inn SCAN, no error message is provided if the tablet camera is unable to detect the QR code or the **Augment** logo. This confuses the user, as whether the user is doing the correct action or not.

3. **User control and freedom**: Very poor. The user does not have any control of the system. At the start of the application there are two options that a user can choose: BROWSE or SCAN. If the user chooses BROWSE then if the user wants to go back to the SCAN and presses the BACK button, then the application just closes and does not show the user with the interface with the two original options. Everytime the user wants to navigate from one option to the other, the user has to close the application and then again relaunch it.

4. **Consistency and standards**: Consistency is not maintained. Since, it is a 3D model it should have possibility of rotation in all three axes. The ROTATE button is not consistent. The 3D model can be rotated only in y-axis whereas a 3D model should be such that it can be rotated in three directions, so that users can place the model in right direction.

5. **Error Prevention**: No error messages are shown. If the user performs a task no confirmation is shown. For example:

   (a) If the user mistakenly selects the wrong 3D model, no confirmation is showned on whether the user wants to select this model or not.

   (b) If the user mistakenly chooses the wrong tracker (this can happen because even if the tracker is not in correct position, the user interface provides a green colored feedback), no confirmation of whether the user is satisfied with the tracker or not.

   (c) No confirmation is required before user clicks the SAVE button. User can mistakenly click the SAVE button even if user did not get the desired position of the model.

(d) In the SMALL CAPS Email Tracker button, even if the tracker is not created, if the user presses the Email Tracker button, the user is requested to enter the email address. No feedback is given to the user that the tracker has not been create and therefore nothing can be e-mailed either.

(e) In the Share button, it directly asks the user to choose the mode of sharing. After the user completes all the steps of sharing (such as providing an e-mail address), then the user is shown the feedback that user first needs to upload the 3D model in website. If the error was shown before, the user would have warned beforehand.

6. **Recognition rather than recall**: Instructions for using the system is not visible or easily retrievable whenever appropriate.

(a) When user opens the application, the user is directly shown with two options. User is confused on what to do.

(b) When user clicks the Browse button, the user is again displayed with a catalogue of articles. Information on how to use the catalogues or how to select the catalogues is not provided.

7. **Flexibility and efficiency of use**: No such feature is found in this prototype.

8. **Aesthetic and minimalist design**: Information is relevant and to the point.

9. **Help users recognize, diagnose, and recover from errors**:

(a) In Scan, the user does not understand why the camera is not recognizing the QR code or the **Augment** logo even when the QR code and the **Augment** logo are present. No feedback is given.

(b) In Browse, the user is not given any feedback on how to place the background image so that it can be used as a marker.

10. **Help and documentation**: No help or documentation is provided on how to use the application at the beginning when user launches the application for the first time. Every user has to see the video on **Augment**'s website to know how to use the system.

Miscellaneous remarks

- In case of tablet, most of the time the model did not place in the correct place and in correct size.

- The application provides limited functionality in free version and all functionality in paid version. When a user user uses the application, no such feedback is shown. For example, that this particular feature can be accessed in paid version or some feedback like such as "you are currently using a free version which has limited functionality".

A.4 Raw data of experiments

Raw data of Figure 1.9

1. The time required to complete the task by User #1: 1 minutes 24 seconds..

2. The time required to complete the task by User #2: 1 minutes 42 seconds.

3. The time required to complete the task by User #3: 1 minutes 5 seconds.

4. The time required to complete the task by User #4: 54 seconds.

5. The time required to complete the task by User #5: 1 minute.

6. The time required to complete the task by User #6: 1 minute 33 seconds.

7. The time required to complete the task by User #7: 1 minute 30 seconds.

Raw data of Figure 1.11a

1. User #1 easily found out the 3D model described in the search task.

2. User #2 easily found out the 3D model.

3. User #3 could not find out the 3D model.

4. User #4 could not find out the 3D model.

5. User #5 could not find out the 3D model.

6. User #6 was at first confused seeing a list of 3D models in the catalogue but later on could find the exact 3D model.

7. User #7 could easily find out the 3D model.

Raw data of Figure 1.11b

1. User #1 could easily create the tracker.

2. User #2 could easily create the tracker.

3. User #3 was confused regarding how to create the tracker. He asked for manual help from one of the experimenters to move forward towards the goal.

4. User #4 (had seen the demo once, seen the previous user performing the same task, did a demo herself before performing the experiment) could easily create the tracker.

5. User #5 could not even remember that he/she had to create a tracker.

6. User #6 was totally confused on how to create the tracker. User asked for manual help.

7. User #7 could not even remember that he/she had to create a tracker.

Raw data of Figure 1.13

- User #1 used the rotate button to rotate the 3D model. Since the rotate button only provides the option of rotating the 3D model in y-axis, the user tried to unsuccessfully rotate the 3D model in x-axis. User #1 also used the tracker to rotate the 3D model.

- User #2 used the rotate button to rotate the 3D model. Since the rotate button only provides the option of rotating the 3D model in y-axis, the user tried to unsuccessfully rotate the 3D model in x-axis and z-axis. User #2 was a bit confused while using the rotate button because rotating the TV on y-axis would put in the TV in top view position with respect to the table which is not an obvious way to put the TV. User #2 also used the tracker to rotate the 3D model.

- User #3 used the rotate button to rotate the 3D model. User #3 did not use the tracker to rotate the 3D model.

- User #4 used the rotate button to rotate the 3D model. User #4 did not use the tracker to rotate the 3D model.

- User #5 used the rotate button to rotate the 3D model. Since the rotate button only provides the option of rotating the 3D model in y-axis, the user tried to unsuccessfully rotate the 3D model in x-axis and z-axis. User #5 did not use the tracker to rotate the 3D model.

- User #6 used the rotate button to rotate the 3D model. Since the rotate button only provides the option of rotating the 3D model in y-axis, the user tried to unsuccessfully rotate the 3D model in x-axis and z-axis. User #6 could create tracker but did not use the tracker to rotate the 3D model.

- User #7 used the rotate button to rotate the 3D model. Since the rotate button only provides the option of rotating the 3D model in y-axis, the user tried to unsuccessfully rotate the 3D model in x-axis and z-axis. User #7 could create tracker and used the tracker to rotate the 3D model.

A.5 Questionnaire for the Augment application

Place: University of Helsinki

We request your help for the design evaluation of Augment application. Please complete the following questionnaire based on your experience when trying to complete the requested task.

Thank you for your time.

Device

☐ Provided tablet
☐ Provided smartphone
☐ Other: _____

Language of the app

☐ English
☐ Spanish
☐ Other: _____

Age (optional) _____

For what purpose do you use your smartphone?

☐ Receive/make calls
☐ Messaging
☐ Calls and messaging
☐ All of the above as well as apps

Have you used any augmented-reality application before?

☐ Yes
☐ No

Answer the questionnaire rating the statements from 1 (totally disagree) to 5 (totally agree). After each question, you can add a comment to specify the reasons of your answer.

1. The system provided me with feedback on what I was working. I was not confused or lost while performing the task.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

2. The messages that appeared in the application were self-explanatory and I understood what they were trying to explain.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

3. I was in control of the application all the time.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

4. I could easily undo/redo any action if I felt to do it.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

5. If I mistakenly chose a wrong 3D model, I could easily stop uploading it.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

6. The navigation through the application was easy.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

7. The application had many errors and/or crashed.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

8. The application always verified the 3D model before loading.

   □ 1   □ 2   □ 3   □ 4   □ 5

   Comments:

9. The option to select the desired function was always clear and available all the time and the icon images helped me to know the appropriate functionality of the available options.

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5

Comments:

10. t was easy to find the desired options at any time.

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5

Comments:

11. The application was visually well-designed.

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5

Comments:

12. If error messages appeared, they were clear in their description and probable steps to recover from it were provided.

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5

Comments:

13. When I needed help, the demo videos and feedback helped me to complete my task successfully.

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5

Comments:

Comments / testimonial:

Thank you very much for taking the time to complete this questionnaire. Your feedback is valued and very much appreciated!

Appendix B

Appendices to Chapter 2

B.1 Participant consent and waiver form

<div align="center">

University of Helsinki, Helsinki

Consent And Waiver

TO WHOM IT MAY CONCERN:

</div>

I hereby grant full permission to the University of Helsinki at Helsinki, Finland to take notes of my comments during the usability test for Augmented Reality Android Application. I understand that other University of Helsinki students involved with user's Android Application may review these usability notes. I understand that I may be quoted directly in internal University of Helsinki documents, with or without reference to my name. I understand that usability notes, my name, address, phone number and email address will not be shared with external third parties. This consent and waiver will not be made on the basis of a future claim of any kind against the University of Helsinki at Helsinki, Finland and any of its agencies.

In witness whereof I have hereunto set my hand and seal this on the _____ day of _____ .

Name: _____

Signature: _____

Address: _____

_____

Witness: _____

Date: _____

B.2 Demographic information questionnaire

1. Full name

   _____

2. Age Group
   □ 11 - 20   □ 21 - 30   □ 31 - 40 □ 41 - 50   □ 51 - 60   □ over 60

3. *Android* smartphone make and model

   _____

4. Years of using *Android* smartphones

   _____

5. Would you consider yourself comfortable with using an *Android* smartphone?
   □ Yes   □ No

6. Gender
   □ Male   □ Female

7. What language do you use to interact with your *Android* smartphone?
   □ Finnish   □ English   □ Other

8. How often do you visit a museum?
   □ Monthly   □ Occasionally   □ Never

9. What is for you the major purpose of using an *Android* smartphone?

   _____

10. Which type of information do you find more useful?
    □ Text   □ Audio

11. Would you like us to contact you about your input at a later date ?
    If yes, please provide your e-mail address.

    _____

B.3 Usability testing questionnaire

1. Please rate the following features based on your evaluation:

1.a) Option Menu
☐ Poor   ☐ Weak   ☐ Medium   ☐ Good   ☐ Strong

1.b) Sharing Audio/Video/Text about museum to your social networks
☐ Poor   ☐ Weak   ☐ Medium   ☐ Good   ☐ Strong

1.c) Audio and video transcript about interesting museum artifacts
☐ Poor   ☐ Weak   ☐ Medium   ☐ Good   ☐ Strong

1.d) More information on Wikipedia
☐ Poor   ☐ Weak   ☐ Medium   ☐ Good   ☐ Strong

1.e) Delete recent activities
☐ Poor   ☐ Weak   ☐ Medium   ☐ Good   ☐ Strong

2. Which kind of information about museums do you find more accessible?
☐ Audio transcript   ☐ Video media   ☐ Text transcript

3. What features of the augmented-reality museum application are vague or confusing to you; add comments if any?

---

4. Does the menu operation seem easy to operation or difficult? What makes it that way?

---

5. What other features do you think should be included in the augmented-reality museum Application?

---

6. What do you like best about the augmented-reality museum application?

---

7. What part or feature do you like the least?

---

8. Scale the rating as per usefulness of Augmented Reality Museum Application (1 being lowest and 5 being highest rating):

☐ 1   ☐ 2   ☐ 3   ☐ 4   ☐ 5

9.Would you like to use the augmented-reality museum application for your next trip to any museum?

☐ Yes   ☐ No

10. Any suggestions to improve the augmented-reality museum application?

---

11. Would you like to make any other comments about the augmented-reality museum application?

Appendix C

# Appendices to Chapter 3

C.1 Adapted heuristics

#	Original heuristic	Modified version
1	Visibility of system status	
	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	The application should always keep user informed about what is going on. The information about the used database or a category of points of interest should be always visible as well as location tracking status and orientation. All information should be given through appropriate feedback within reasonable time.
2	Match between system and the real world	
	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	The application should show correct user's location and orientation and also correct location and distance to visible points of interest. The distance units should respect user's locale.
3	User control and freedom	

| #	Original heuristic	Modified version
	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.	Allow user to switch freely between different databases and categories of points of interest. Allow also to choose between the map navigation and the live video overlay. Return to the same state after switching between applications. Consider returning into a default state after a long interruption.
4	Consistency and standards	
	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.	The user interface controls should respect conventions used for mobile devices. All used graphic symbols should be clear and intuitive.
5	Error prevention	
	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	The application should be designed in such way, that it prevents errors from happening. If it is not possible, the application should be able to recover.
6	Recognition rather than recall	
	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	Show user "where" he is in the application and make navigation throughout the application simple. The most important functions (like search and switching between the map navigation and the live video overlay) should be always easily accessible.
7	Flexibility and efficiency of use	

#	Original heuristic	Modified version
	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.	Make user's favorite functions/databases easily accessible. Allow user to configure shortcuts for these options.
8	Aesthetic and minimalist design	
	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	Screens should not contain information which is irrelevant or unneeded.
9	Help users recognize, diagnose, and recover from errors	
	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.	All errors should be in a clear, user-readable form and they should suggest a reasonable solution. If an error happens, user should be notified. There should be no "silent" errors.
10	Help and documentation	
	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.	Provide some help and/or documentation. Both a quick start help and a detailed documentation should be available.

C.2 Maps for the user-test scenarios

In both figures is a red cross indicates the location of switching between the printed map and Wikitude.
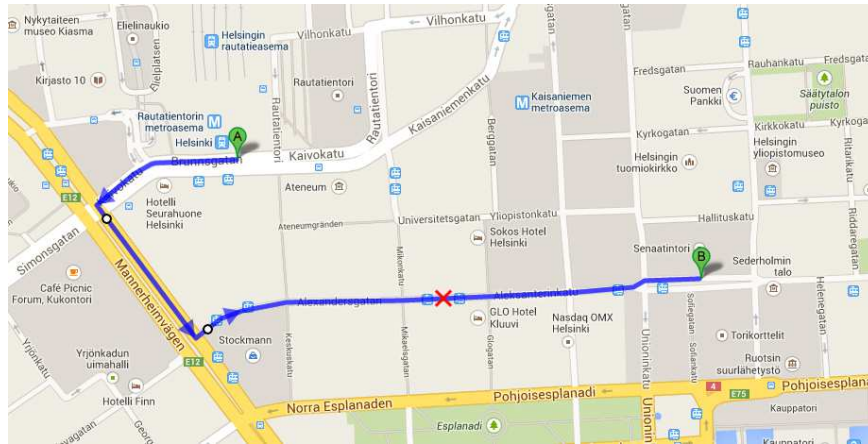
(a) The path from the Central railway station to the Helsinki Cathedral.
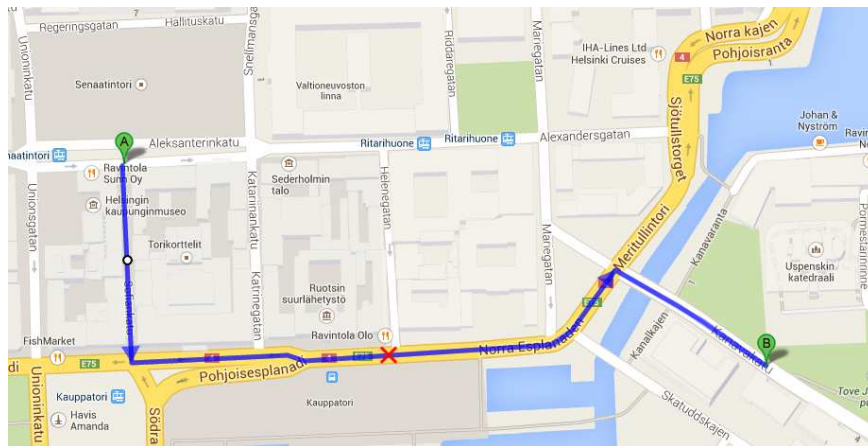
(b) The path from the Helsinki Cathedral to the Uspenski Cathedral.

Figure C.1: Maps used in the test cases.

Author information

Satu Elisa Schaeffer `elisa.schaeffer@uanl.edu.mx`
Studied Computer Science and Engineering at the Helsinki University of Technology (now Aalto University), Finland, 1996–2006. Presently she is associate professor at Universidad Autónoma de Nuevo León and is a member of the National Research System of Mexico. Her research interests include complex adaptive systems and diverse applications of optimization methods. In 2013–2014, she spent a two-semester sabbatical leave at Helsinki Institute for Information Technology HIIT, at the Kumpula Campus of the University of Helsinki.

Héctor Martínez Landa `hector.martinez@sensetrix.com`
Researcher at SenseTrix and Ph.D. student at Tampere University of Technology. Received a M.Eng. in Industrial Engineering specializing in Automation and Robotics in 2009 and a M.Eng. specializing in System Engineering and Computer Science in 2011, both at University of Zaragoza. Between 2008 and 2011, worked as researcher at BSH Home appliances, University of Zaragoza, and Instituto Tecnológico de Aragón. His research topics during these periods were computer vision, augmented reality, and virtual agents, among other topics. Recipient of a Marie Curie fellowship to move to Finland to join SenseTrix in July 2011. His main current area of interest is augmented reality, especially oriented to applications in maintenance and education.

Payel Bandyopadhyay `payel.bandyopadhyay@helsinki.fi`
Studied Bachelor's in Information Technology from Jalpaiguri Government Engineering College (Autonomous), which is under West Bengal University of Technology, India, 2008–2012. Currently studying Master's in Computer Science and Engineering at University of Helsinki, Finland, 2012 to present. Presently she is also a member of Helsinki Institute of Information Technology HIIT (Ubiquitous Interaction group) in Helsinki, Finland. Her research interests include human-computer interaction, mixed reality, and information retrieval.

Maninder Pal Singh `maninder.singh@cs.helsinki.fi`

Received a Bachelor's degree in Computer Science and Engineering from National Institute of Technology, Durapur, India, 2005–2009. Between 2009 and 2013, he worked in Android research and development labs at Samsung Electronics, India. He is currently working towards a Master's degree in Computer Science and Engineering at University of Helsinki, Finland, 2013 to present. His research interests include large-scale distributed computing, data mining, and information retrieval.

Lei Wang `lei.z.wang@helsinki.fi`

Obtained B.Eng. in Communication Engineering at the Jilin University in China, 2006–2010. Currently studying for M.Sc. in Computer Science at the University of Helsinki, 2013 to present. A research assistant in the Nodes Lab at the University of Helsinki. His research interests include networking protocols, SDN, distributed systems, data mining, and machine learning.

Ondřej Perutka `perutond@fit.cvut.cz`

Studied Bachelor's in Computer Science at Faculty of Information Technology of Czech Technical University in Prague, Czech Republic, 2009–2012. Currently studying Master's in Computer Science at the same institution, 2012 to present. His research interests include stringology and data compression.

Aki Kesulahti `aki@kesulahti.fi`

Studied Bachelor's in Computer Science at University of Helsinki, Finland, 2009–2014. Currently studying Master's in Computer Science at the same institution. His main current areas of interests include wearable technology, augmented reality, and anonymity networks.

Jorma Nieminen `Jorma.Nieminen@cs.helsinki.fi`

Currently a student at University of Helsinki.

Bibliography

[1] 5DT. Data gloves. URL `http://www.5dt.com/?page_id=34`.

[2] G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide: a mobile context aware tour guide. *Wireless Networks*, 3 (5):421–433, 1997.

[3] M. Alshamari and P. Mayhew. Task design: Its impact on usability testing. In *Proceedings of the Third International Conference on Internet and Web Applications and Services*, pages 583–589, Washington, DC, USA, 2008. IEEE.

[4] M.S. Andreasen, H.V. Nielsen, S.O. Schrøder, and J. Stage. What happened to remote usability testing?: An empirical study of three methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1405–1414, New York, NY, USA, 2007. ACM.

[5] Augment. Company website. URL `http://augmentedev.com/`.

[6] R.T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, 1997.

[7] R. Bailey, C. Wolfson, J. Nall, and S. Koyani. Performance-based usability testing: Metrics that have the greatest impact for improving a system's usability. In M. Kurosu, editor, *Human Centered Design*, volume 5619 of *Lecture Notes in Computer Science*, pages 3–12. Springer, Berlin/Heidelberg, Germany, 2009.

[8] R.W. Cholewiak and A.A. Collins. Sensory and physiological bases of touch. In M.A. Heller and W. Schiff, editors, *The Psychology of Touch*, pages 23–60. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, USA, 1991.

[9] A. Damala, P. Cubaud, A. Bationo, P. Houlier, and I. Marchal. Bridging the gap between the digital and the physical: Design and evaluation of a mobile augmented reality guide for the museum visit. In *Proceedings of the Third International Conference on Digital Interactive Media in*

Entertainment and Arts, pages 120–127, New York, NY, USA, 2008. ACM.

[10] D. Dand. SuperShoes. Tokyo Design Week, 2013. URL `http://dhairyadand.com/sec/page=projects&id=supershoes`.

[11] H. Desurvire, J. Kondziela, and M.E. Atwood. What is gained and lost when using methods other than empirical testing. In *Posters and Short Talks of the 1992 SIGCHI Conference on Human Factors in Computing Systems*, pages 125–126, New York, NY, USA, 1992. ACM.

[12] H.B.-L. Duh, G.C.B. Tan, and V. H. Chen. Usability evaluation for mobile device: A comparison of laboratory and field tests. In *Proceedings of the Eighth Conference on Human-computer Interaction with Mobile Devices and Services*, pages 181–186, New York, NY, USA, 2006. ACM.

[13] A. Dünser, R. Grasset, H. Seichter, and M. Billinghurst. Applying hci principles in ar systems design. In *Proceedings of the Second International Workshop on Mixed Reality User Interfaces: Specification, Authoring, Adaptation*, Washington, DC, USA, 2007. IEEE.

[14] K. Ericsson and H. Simon. *Protocol Analysis: Verbal Reports as Data.* MIT Press, Boston, MA, USA, 2nd edition, 1993.

[15] L. Fang. Usability evaluation on websites. In *Proceedings of the Ninth International Conference on Computer-Aided Industrial Design and Conceptual Design*, pages 141–144, Washington, DC, USA, 2008. IEEE.

[16] B. Gillham. *Developing a questionnaire.* Continuum International Publishing Group Ltd., London, England, UK, 2nd edition, 2008.

[17] J.D. Gould and C. Lewis. Designing for usability: Key principles and what designers think. *Commununications of the ACM*, 28(3):300–311, 1985.

[18] A. Holzinger. Usability engineering methods for software developers. *Communications of the ACM*, 48(1):71–74, 2005.

[19] T.-W. Kan, C.-H. Teng, and W.-S. Chou. Applying QR code in augmented reality applications. In *Proceedings of the Eighth International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 253–257, New York, NY, USA, 2009.

[20] A. Karoulis, S. Sylaiou, and M. White. Usability evaluation of a virtual museum interface. *Informatica*, 17(3):363–380, 2006.

[21] J. Kjeldskov and G. Connor. A review of mobile HCI research methods. In *Human-computer interaction with mobile devices and services*, pages 317–335, Berlin/Heidelberg, Germany, 2003. Springer.

[22] N.N. Kostaras and M.N. Xenos. Assessing the usability of augmented reality systems. In *Proceedings of the 13th Panhellenic Conference on Informatics*, pages 197–201, Washington, DC, USA, 2009. IEEE Computer Society.

[23] C. Lewis. Using the "thinking aloud" method in cognitive interface design. Technical Report RC-9265, IBM, 1982.

[24] J. Linjama and T. Kaaresoja. Perception of short tactile pulses generated by a vibration motor in a mobile phone. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 471–472, Washington, DC, USA, 2005. IEEE.

[25] M.N. Mahrin, P. Strooper, and D. Carrington. Selecting usability evaluation methods for software process descriptions. In *Proceedings of the Asia-Pacific Software Engineering Conference*, pages 523–529, Washington, DC, USA, 2009. IEEE.

[26] V.F. Martins, E.L. Makihara, L.V. Soares, R.N. Pinto, V.B. Mattos, and M. de Paiva Guimaraes. Usability metrics for augmented reality applications. In *Proceedings of the 2012 XXXVIII Conferencia Latinoamericana en,Informatica*, pages 1–10, 2012.

[27] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, E77-D(12):1321–1329, 1994.

[28] R. Molich. *Usable web design*. Nyt Teknisk Forlag, Odense, Denmark, 2008.

[29] J. Nielsen. How to conduct a heuristic evaluation. Nielsen Norman Group, 1995. URL http://www.nngroup.com/.

[30] J. Nielsen. 10 heuristics for user interface design. Nielsen Norman Group, 1995. URL http://www.nngroup.com/.

[31] J. Nielsen. Usability 101: Introduction to usability. Nielsen Normam Group, January 2012. URL http://www.nngroup.com/.

[32] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 249–256, New York, NY, USA, 1990. ACM.

[33] J. Preece, Y. Rogers, and H. Sharp. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley and Sons, New York, NY, USA, 2007.

[34] F.H.A. Razak, H. Hafit, N. Sedi, N.A. Zubaidi, and H. Haron. Usability testing with children: Laboratory vs field studies. In *Proceedings of the International Conference on User Science and Engineering*, pages 104–109, Washington, DC, USA, 2010. IEEE.

[35] J. Rieman, M. Franzke, and D. Redmiles. Usability evaluation with the cognitive walkthrough. In *Conference Companion on Human Factors in Computing Systems*, pages 387–388, New York, NY, USA, 1995. ACM.

[36] S. Rosenbaum. Usability evaluations versus usability testing: when and why? *IEEE Transactions on Professional Communication*, 32(4): 210–216, 1989.

[37] J. Rubin, D. Chisnell, and J. Spool. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. John Wiley and Sons, New York, NY, USA, 2008.

[38] I. Shirai. Multi-directional switch. US Patent US 4687200 A, 1987. URL http://www.google.com/patents/US4687200. Nintendo Co, Ltd.

[39] D.D. Sumadio and D.R.A. Rambli. Preliminary evaluation on user acceptance of the augmented reality use for education. In *Proceedings of the Second International Conference on Computer Engineering and Applications*, pages 461–465, Washington, DC, USA, 2010. IEEE.

[40] I.E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the Fall Joint Computer Conference, Part I*, pages 757–764, New York, NY, USA, 1968. ACM.

[41] K. Tsukada and M. Yasumura. ActiveBelt: Belt-type wearable tactile display for directional navigation. In *UbiComp 2004: Ubiquitous Computing*, volume 3205 of *Lecture Notes in Computer Science*, pages 384–399, Berlin/Heidelberg, Germany, 2004. Springer.

[42] C.P.J.M. Van Elzakker, I. Delikostidis, and P.J.M. Van Oosterom. Field-based usability evaluation methodology for mobile geo-applications. *The Cartographic Journal*, 45(2):139–149, 2008.

[43] X. Wang. Using cognitive walkthrough procedure to prototype and evaluate dynamic menu interfaces: A design improvement. In *Proceedings of the International Conference on Computer Supported Cooperative Work in Design*, pages 76–80, Washington, DC, USA, 2008. IEEE.

[44] C. Wharton, J. Rieman, C. Lewis, and P. Polson. The cognitive walkthrough method: A practitioner's guide. In J. Nielsen and R.L. Mack, editors, *Usability Inspection Methods*, pages 105–140. John Wiley and Sons, New York, NY, USA, 1994.

[45] M. White, F. Liarokapsis, J. Darcy, N. Mourkoussis, P. Petridis, and P.F. Lister. Augmented reality for museum artefact visualization. In *Proceedings of the Fourth Irish Workshop on Computer Graphics, Eurographics Ireland Chapter*, pages 75–80, Coleraine, Northern Ireland, UK, 2003. Eurographics Ireland.

[46] A.M. Wichansky. Usability testing in 2000 and beyond. *Ergonomics*, 43(7):998–1006, 2000.

[47] N.M.M. Zainuddin, H.B. Zaman, and A. Ahmad. Heuristic evaluation on augmented reality courseware for the deaf. In *Proceedings of the 2011 International Conference on User Science and Engineering*, pages 183–188, Washington, DC, USA, 2011. IEEE.