Journal of Universal Computer Science, vol. 19, no. 8 (2013), 1175-1196
submitted: 30/11/12, accepted: 29/3/13, appeared: 28/4/13 © J.UCS

A Case Study on User Experience (UX) Evaluation of Mobile Augmented Reality Prototypes

Amandeep Dhir
(Department of Computer Science and Engineering, Aalto University, Finland
amandeep.dhir@aalto.fi)

Mohammed Al-kahtani
(Computer Engineering Department, Salman Bin Abdulaziz University
Al Kharj, Saudi Arabia
alkahtani@sau.edu.sa)

Abstract: Mobile Augmented Reality (MAR) blends the real world with digital objects especially in ubiquitous devices such as smartphones. The MAR applications provide an intelligent interface for users. In this, valuable digital information is advertised in physical spaces. However, the success of these applications is tied directly to the degree of user acceptance. This makes understanding the needs and expectations of the MAR's potential users of paramount importance for designing and building the proper application. The objective of the paper is to expose an important gap in the development of novel applications in the virtual world. Previous research has shown that it is essential to study and understand the needs and expectations of the potential users of the upcoming application or system. Studying user needs and expectations before offering the developed application ensures a minimum level of acceptance and, of course, success. This paper presents a detailed study comprising of a user-experience (UX) evaluation of different prototypes through the use of three different UX evaluation methods. This kind of evaluation allows new developments to offer systems, which do not fail. The main contributions of this study are that it: 1) solicits expectations when consumers use MAR applications, 2) assesses the UX over different prototypes using three different metrics, 3) provides methodological insights on UX evaluation experiments and, 4) is useful for anyone who wants to develop handheld applications after understanding user expectations and how his experience should progress. The results of the study show that users value concreteness, realizability, personalization, novelty, intuitiveness and the usefulness of presented information. Paying attention to these factors can help develop more acceptable MAR applications and lead to more novel future designs.

Keywords: End-user application, mobile mixed reality, mobile services, user experience, user expectations, user experience evaluations
Categories: H.5.0, H.5.1, H.5.2, L.2.1, L.3.1, L.3.4, L.3.6, L.3.8

1 Introduction

The early 1990's witnessed the emergence of Ubiquitous Computing (UbiComp) that has made interaction between users and their environment possible anywhere and anytime [Weiser, 1991]. Mark Weiser emphasized that UbiComp stands for the seamless integration of computing into the users' environment so that computing becomes invisible at a certain point [Weiser, 1993]. One of the well-known and important practical realizations of UbiComp is Augmented Reality (AR). The AR

based concepts are founded on the principle that real world objects are augmented with the digital objects and everything eventually appears part of one environment [Azuma, 1997; Bach and Scapin, 2004; Milgram and Kishino, 1994]. Due to the presence of interaction and integration of digital and real-world objects, AR presents an excellent use case of UbiComp as coined by Mark Weiser. AR is a two-decade-old concept, but the Mobile Augmented Reality (MAR) concept is relatively new due to the advent of mobile devices, equipped with numerous capabilities such as high connectivity, multiple sensors and a wide range of multimedia features. The MAR applications present an immersive view by making digital objects supplement and mix with the real world-view [Azuma, 1997; Bach and Scapin, 2004; Milgram and Kishino, 1994]. They tend to increase the user's understanding of their surrounding environment by presenting contextual and relevant information to the mobile users [Olsson et. al., 2009].

With the introduction of the mobile version of AR, plenty of newer opportunities have opened up for software developers, designers and business enterprises and above all for the ordinary mobile user. Some of the known businesses prospects strengthened by MAR are available from the different field such as tourism, personalized shopping and different augmented events [Friedrich, 2002; Höllerer and Feiner, 2004; Stanney, 1995; Zhang et. al., 2000].

Despite the incessant rise in the number of commercially available MAR applications; such as layer; Wikitude and Junaio, the user needs, expectations and experiences are understudied. This is partially due to the nature of MAR applications that constantly face temporal changes in contexts due to the user mobility and movements [Olsson et. al., 2009; Vaittinen et. al., 2010]. During the review of existing literature, it was found that there is scarcity of research that investigates the needs and expectations of users from MAR and how to evaluate the user experience of any potential MAR application with real users. Understanding the needs, expectations and requirements of the potential users from any application/software, which is yet to be developed, is considered important and essential. Previous research from the related subject area has shown that a study of early user needs and expectations could potentially help in the approximation of the actual user experience even before the application, product or service is realized [Heikkinen et al., 2009]. Similarly, user experience researchers and practitioners have concluded that the basis of any positive user experience is already defined when the application, product or service is still at the conceptual level, even when no actual interaction between the application or product and the user exists [Roto et al., 2009].

Interestingly, we found another practical phenomenon during the review of previous literature, our interaction and observation exercises with application developers and software specialists in start-up companies in Finland during Spring 2011. It was found that the majority of the application developers and software specialists are fully aware of performing usability and user experience tests of the fully functional and implemented systems. However, less is known if it is possible to evaluate both usability and user experience, even when the software or application is not yet developed i.e. application is non-functional and available at concept level. This clearly shows the presence of a gap in the related literature, as well as in the understanding of when and how to evaluate the usability and user experience of the application, software or product, which is currently under-development.

Considering this need and existing gap in the literature, the current paper presents a detailed evaluative study comprising user-experience (UX) evaluation of different prototypes (developed using user-centered design (UCD) process) through the use of three different UX evaluation methods. This kind of evaluation allows new developments to offer systems and a lower probability to fail. Additionally, this study also aims to explore, examine and present those factors and attributes that can positively build user experience of MAR applications. We argue that examining those aforementioned factors and attributes are important and are of worth considering the practical needs of application and software developers. Furthermore, these factors might directly influence the business potential of intended MAR based applications and services.

The rest of the paper is organized as follows: Section 2 presents background literature on UX and UCD covering discussion on the emergence of UX research since the previous decade. Additionally, previous work related to UX design in MAR has been presented. In section 3, research methodology of the study comprising main research questions, different used MAR prototypes and three employed UX evaluation methods have been discussed. In section 4, the UX evaluation study setup, its design and process has been thoroughly explained. In section 5, the evaluation results of this study have been presented. In the next section, the study results have been discussed in the light of main research questions and associated objectives because consideration has been given to the validity and reliability of the study results. Finally, the paper concludes in Section 7

2 User Experience (UX) and MAR Applications

The last decade witnessed an ever-growing interest in studying and understanding the UX of different products and services [Jordan, 2000; Kuutti, 2010]. UX governs the broad range of perspectives involving user's pre-use expectations and interaction experiences with product [Forlizzi and Battarbee, 2004; Hassenzahl and Tractinsky, 2006; Kankainen, 2003; Karapanos et. al., 2009]. The concept of UX is not even a decade old phenomenon and, due to this reason, there has been no agreement so far on one particular definition of UX. In this direction, most UX researchers and practitioners prefer ISO 9241-210 standard on UX. ISO 9241-210 defines UX as "*a person's perceptions and responses that result from the use or anticipated use of a product, system or service"* [ISO 9241-210, 2009].

UX research agenda is also seen as an extension of the much studied research topic referred as UCD. UCD is defined as an interactive process with the sole goal to develop a usable system, software or application. The UCD process strongly advocates the need to involve potential users of the intended system or product (which is yet to be developed) into the system design [Karat et. al., 1996]. The main benefits of UCD methodology are that: 1) active user involvement enables software developers and specialists to get a clear understanding of the potential users' needs and expectations 2) UCD does not only cover user involvement but a philosophy towards interactive designing and evaluation [Vredenburg et. al., 2002]. UCD has several definitions, but, for this paper, we adopted the definition given by ISO 13407 on UCD. This standard defines UCD as a four stage interactive process that can be explained as follows: 1) First, understand and specify the context of use with a strong

focus on finding user goals, insights, tasks and behavior 2) Specify the requirements from the user and organization's point of view 3) Based on the collected information and previous knowledge, the product is designed 4) Finally, evaluate the design against the user needs and requirements [ISO/IEC. 13407, 1999].

In the field of designing pleasurable products, services or systems, work by Jordan is most cited and refereed to [Jordan, 2000]. Jordan created a three-phase roadmap on how the field of Human-Computer Interaction (HCI) has evolved since 2000 (see Figure 1) [Jordan, 2000]. The first phase is situated in the early 1980's when HCI specialists were underappreciated. The second phase is situated in the early 1990's when HCI specialists began to be acknowledged and, were asked to design interfaces for a few products. The third phase begins at the start of the new millennium when more HCI specialists were hired, usability became essential for any type of product and the UCD methodology became part of any agile product development. Based on our experience, we adopted Jordan's three-phase model and augmented it with two more phases in order to match it with the recent developments in the HCI field. Figure 1 shows the full spectrum of phases including the two phases, which we added. The newly added fourth phase is the "Emergence of UX and overlapping with Usability", which describes what UX is, how UX emerged, how to study it, how to evaluate it and so on. At this time, the UX is often overlapped with traditional usability to the extent that UX and usability are used interchangeably.

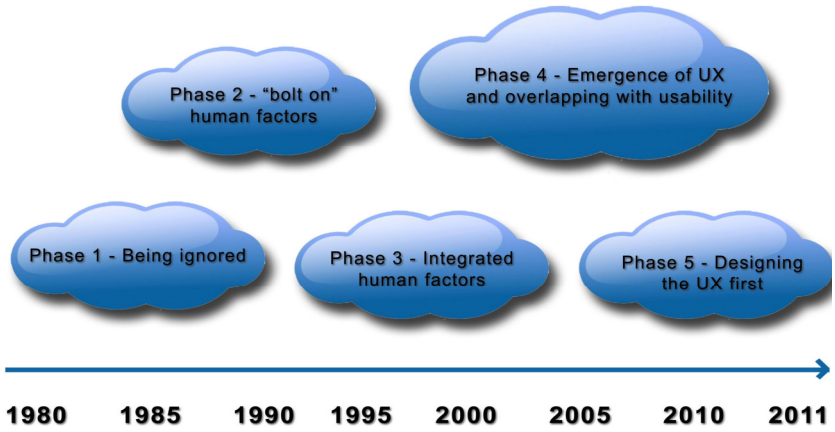


Figure 1: The Roadmap of UX over time

Finally, the fifth phase began at the end of the previous decade, which is "UX comes first!" and this it is still in progress. It alludes to the significance and real need for conducting UX, but the field is still surrounded by confusion and unanswered questions germane to the methodologies that should be adopted. UX is different from usability as it focuses on the emotional and temporal nature of experiences that users go through when testing a product or service. Lately, different models, frameworks and tools have been developed in order to examine the UX of products and services and evaluate their design.

The UX of any product or service is believed to evolve and change continuously over time due to its temporal nature [Kankainen, 2003]. Furthermore, the UX is affected by the users' previous experiences and present expectations [Hassenzahl and Tractinsky, 2006; Kankainen, 2003; Karapanos et. al., 2009]. However, despite the increasing number of studies pertaining to UX of products and services, its research agenda remains immature. On examining previous published research works on UX design and evaluation in context to MAR applications, it was found that despite the large body of existing research on MAR, there is still a lack of understanding of the users' needs and expectations from the UX perspective.

As users' expectations of novel technologies grow, it becomes imperative to evaluate the UX of MAR services in order achieve mass adoption. Heikkinen et al., [2009] suggested that studying user expectations could potentially help in the approximation of the actual UX even before the product or service is realized. Furthermore, Roto et al. [2009] emphasized that the basis of any positive UX is already defined at the conceptual level of the product, even when no actual interaction between the product and the user exists. Olsson et. al. [2009] conducted user studies to gather user needs and expectations through focus discussions using different sets of use cases and scenarios on MAR. In another study, Vaittinen et. al. [2010] performed diary studies to collect user needs in the context of MAR services. However, like most of the existing investigations, these studies were based on gathering user expectations through a limited set of use cases, supporting almost no actual interaction between the user and the MAR prototypes. Therefore, we argue for the urgent need to examine the UCD and UX aspects of MAR. In contrast, our study herein evaluates the UX by assessing the user needs and expectations of some MAR prototypes that we developed. Our MAR prototypes are visual and physical through which users can interact, formulate opinions and express their user experience rigorously.

Considering these scientific facts on designing potential UX for any system or application, we identified two main research objectives in context to the development of potential MAR applications. These research objectives were: 1) understand the users' needs and expectations from any MAR service or application 2) ensure that the MAR application fits user contexts and enhances their daily lives. In this study, UX has been evaluated in context to the MAR prototypes, which was developed after practicing UCD methodology. The current paper solely focuses on UX design and evaluation and, UCD methodology practiced to develop MAR prototypes has been explained in a previous paper [Dhir et. al., 2012]

3 Research Methodology

In the current study, UX evaluation of the different semi-functional MAR prototypes has been presented through the use of three different UX evaluation methods namely AttrakDiff [Hassenzahl, 2003a], Emocard [Desmet et. al., 2001] and the SUXES [Turunen et. al., 2009]. The prototypes used in this evaluation were developed using the UCD. The detailed UCD process for constructing these prototypes is not described in this paper as the current paper is solely about UX evaluation while the UCD methodology practiced to develop MAR prototypes has been explained in a previous paper [Dhir et. al., 2012].

In this section, the research methodology behind our study is explained by presenting associated research questions, MAR prototypes and research methods for the purpose of the UX evaluation.

3.1 Research Questions

The present study is based on three main research questions, which are addressed in this paper:

1. Which research methods are most suitable for evaluating the UX?
2. How are the hedonic and pragmatic aspects of the UX evaluated in the context of MAR prototypes?
3. Which elements affect the users' expectations and perceptions when interacting with MAR applications?

3.2 The MAR Prototypes

We have developed four semi-functional MAR prototypes through the UCD process which includes focus group discussions using written scenarios, heuristic evaluations of the 25 available AR-MAR applications and prototyping with 10 users. The developed prototypes represent an extensive set of use cases and developed scenarios based on user needs and expectations during our user research. The UCD methodology practiced to develop MAR prototypes has been explained in a previous paper [Dhir et. al., 2012].

The created prototypes differ from and complement each other at the same time. For instance, the first prototype presents digital icons augment streets and locations on maps and the second prototype displays digital information clouds while navigating indoor or outdoor. The third prototype aims to help users to overcome the language barrier. The fourth prototype is based on the playful aspects of MAR services. While creating the MAR prototypes, we decided to focus on different factors such as different types of interactions, different MAR contents, and the mode of MAR use. Furthermore, we wanted to explore if the level of completion i.e. maturity of the prototype affects the UX evaluation results.

3.2.1 The MAR Street

As Figure 2 shows, the MAR Street app provides an immersive view of streets by tagging the desired locations with different icons for dinners, restaurants and bars. The app is built using Google Maps API, HTML and XML technologies.

3.2.2 The MAR Navigation

The MAR navigation app augments the surroundings, such as building, object or a location, with information tags like the weather forecast and current temperature, real time traffic broadcasts, news headlines and location names (see Figure 2). This app was created using short video clips displaying annotated text information.

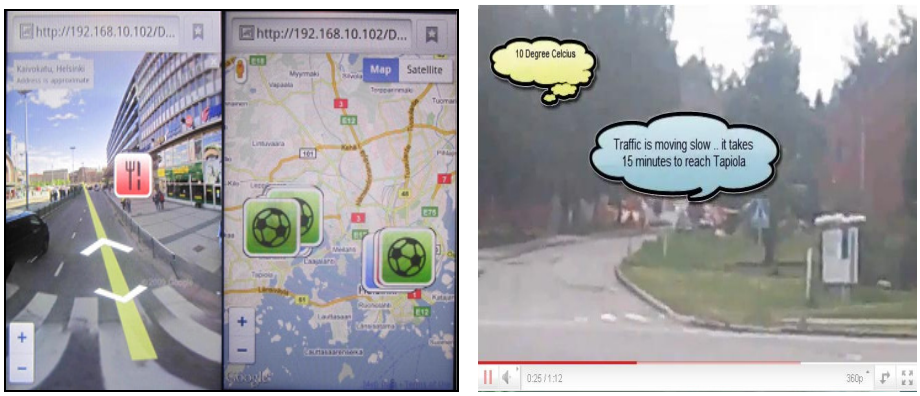


Figure 2: Overview of the MAR Street and MAR Navigation Apps

3.2.3 The MAR Barrier App

This app demonstrates how the MAR can potentially be used to overcome the language barrier. Short video clips were used to show markets and streets with large signs with Chinese text (see Figure 3). Later, digital icons representing "places to eat and drink" appeared over the different signs.

Figure 3: The MAR Language Barrier App

3.2.4 The MAR Playfulness App

Through this app, participants are introduced to the concept of the virtual pet in order to motivate them to engage in information sharing (see Figure 4). The virtual pet is a MAR avatar and invisible friend with whom a participant can chat and share feelings. Based on the amount of time spent with the pet, a participant can earn points which can be used to buys food for their virtual pet. It creates social, gaming and sporting experiences and motivates MAR users to share information.

Figure 4: Virtual pets from the MAR Playfulness App

3.3 Research Methods

We assessed the UX using three different methods: the SUXES, Emocard, and AttrakDiff. In addition to testing each method individually, we combined the three together. This combination enabled us to collect quantitative, as well as qualitative user feedback. The process of experimentation using three different UX evaluation methods is explained in the section 4 on "*Our study*". Next, we briefly explain the three research methods used in this study.

3.3.1 The SUXES Method

SUXES is a subjective method that is commonly used to evaluate the UX of speech-based multimodal systems [Turunen et. al., 2009]. SUXES is an adapted version of the famous marketing tool SERVQUAL that is commonly used for service quality evaluation and, is recognized as an effective research method to capture subjective metrics, such as the users' pre-expectations and post experiences. SUXES is helpful in locating the strong features, as well as those areas that require further improvement in any product or service [Turunen et. al., 2009].

Since the SUXES methodology does not have a strict and well-defined structure for its implementation we opted to modify the method, thus it consisted of 6 phases: background questionnaire, briefing about the prototype, expectation questionnaire, testing the prototype, experience questionnaire and, finally, the user feedback. The SUXES gathers nine types of statements and service feedbacks from users, namely, speed, pleasantness, clearness, ease of use, robustness, learning curve, naturalness, usefulness and future use [Turunen et. al., 2009].

For each MAR prototype, the participants are required to evaluate all seven attributes on a scale from one to seven. Participants have to choose a value between *acceptable level*, denoting the *lowest level* of quality that is adequate for the participant, and the *desired level,* denoting the *highest level* of quality after which the subject cannot advance.

3.3.2 The Emocard Method

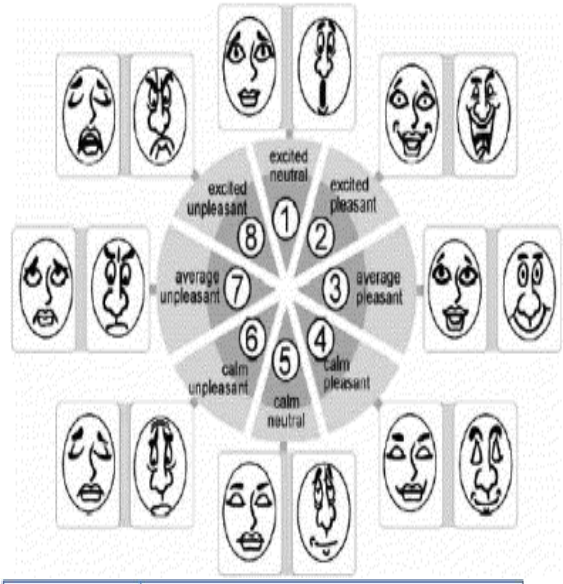
Due to the presence of a direct relationship between emotional responses and product acceptability, UX researchers and practitioners are interested in understanding the emotional response of users when they interact with any product or service, i.e. emotional assessment helps to improve understanding of the UX [Tahti et. al., 2004].

Emotional assessment and evaluation is a complex and challenging subject as users often face problems while explaining their feelings [Tahti et. al., 2004]. At present, there are many emotion evaluation tools that help to capture and interpret emotions. These tools are broadly classified as verbal and non-verbal evaluation tools. The former refers to self-reports where subjects use a scale to record their emotions while the latter makes use of visual representations of emotions from which the subject chooses how they feel. In this study, we use a non-verbal tool, Emocard, for the emotional assessment of our MAR prototypes. The Emocard method is relatively reliable in its interpretation as it consists of human like deceptions that are cross-culturally validated [Desmet et. al., 2001]. Furthermore, non-verbal tools are applicable over a wider audience compared to verbal tools because non-verbal tools are efficient in capturing the conscious state of the human mind unlike their counterparts [Tahti et. al., 2004]. However, it was found that Emocard is widely used and it is empirically supported tool whose validity is known in the usability domain, but its validity in evaluating the UX of MAR applications is unknown.

Emocard consists of 16 different cartoon faces (8 Male and 8 Female faces). These faces represent different emotions and every face depicts a combination of two emotion dimensions, e.g., pleasure and excitement. Emocard is divided into four quadrants namely calm-pleasant, clam-unpleasant, excited-pleasant and excited-unpleasant [Desmet et. al., 2001]. If the user's reaction is more pleasant and higher in arousal then the product is refereed as desirable and neat. Similarly, if the user's reaction is calm-pleasant and excited-pleasant then it shows positive results. Participants were given training on how to interpret Emocard and enter the corresponding answer. To that end, every participant performed two dummy tasks using Emocard and was interviewed before performing the actual test. During the pilot study, it was found that some participants still faced problems in interpreting the original Emocard. Therefore, we decided write a more descriptive text with every image as shown in Figure 5.

3.3.3 AttrakDiff

AttrakDiff is a UX evaluation metric for evaluating the pragmatic and hedonic qualities of UX for any product or service [Hassenzahl, 2003a]. AttrakDiff metric contains 28 different attributes, which are categorized into three main groups: perceived hedonic quality identification (HQI), perceived hedonic quality stimulation (HQS) and perceived pragmatic quality (PQ) [Hassenzahl, 2003a]. All the attributes are evaluated using the bipolar semantic differential 7-scale method. In any typical study, the AttrakDiff questionnaire is performed in four phases: the first phase is introduces the AttrakDiff method and its 28 attributes. In the second phase, the technology itself is introduced, and in the third phase, the participants are instructed to answer AttrakDiff. Finally in the fourth phase, the participants are introduced to the actual technology in the form of product or prototype and asked to evaluate it through the AttrakDiff metric.



Emocard	Explanations
	Excited Neutral Quite ok, Useful but slow, No idea what I am doing
	Excited Pleasant Positive, Nice to test and Funny
	Average Pleasant Nice to test, Easy to use. Nice experience
	Clam Pleasant Easy even for me, Quite good, Usage as expected
	Clam Neutral Getting tired, Complex and problematic, Neutral
	Clam Unpleasant Difficult to use, unfinished demo, less than expected
	Average Unpleasant Not working properly, Difficult to use, Don't know how to use
	Excited Unpleasant Causing anxiety, not working as expected, jammed

Figure 5: Overview of Original Emocard along with the modified version of Emocard

All three methods fulfill our research questions pertaining to gathering user expectations and evaluating user experiences. Since the UX inherently has a temporal nature, we opted to collect pre-use expectations and post-use perceptions in order to help analyze the overall UX of our MAR prototypes. Specifically, we used the SUXES method to capture subjective metrics on pre-use expectations and post-use experiences; we used AttrakDiff to evaluate both hedonic and pragmatic qualities,

while Emocard was used as a non-verbal self-reporting method to evaluate momentary emotions.

4 Our Study

Ten subjects participated in our UX evaluation through one-on-one sessions. Demographically, they were 9 Males and 1 Female and they are Master candidates with an IT background. Their ages ranged from 22 to 28 with a mean of 24.4 years and a standard deviation of 2.01 years. The study was advertised through a common mailing list targeted at master-level students at a local university. The invitation contained the required background information of our UX evaluation study. A total of 16 students responded and showed interest in participating. After scheduling the experiment sessions, only 10 students showed up. We definitely did not seek an unbalanced gender participation where Female participation was limited to one person only, but unfortunately this conforms to the general trend of fewer Female students in IT. On average, each test session took one hour and 15 minutes. The participants received two university lunch coupons as a compensation for their time.

4.1 The Experiments

All tests took place in June 2011 on a local university campus. We used two mobile phones: Nokia N900 and Samsung Nexus S, one laptop computer and an LCD TV. The test participants were given blank sheets, pencils and pens to record their responses and draw their ideas and related meanings. All 10-test sessions were audio taped and later coded to interpret the participants' responses and feedback. Before the actual experiment, we conducted pilot tests with two subjects in order to pinpoint any pitfalls that we may encounter as early as possible. To that end, the results collected from these rehearsals were not included in the final UX evaluation results. Each pilot test lasted for 50 minutes, which enabled us to spot unclear and confusing statements in the questionnaire. For example, the participants initially found the scales, dimensions, adjectives and facial patterns difficult to interpret. We redressed these problems by adding additional explanations and training components prior to the commencement of the real study. The pilot users also faced slight problems while interacting with one of the prototype. Consequently, we improved the way they drag and drop the virtual objects. In order to control the participants' learning pace and fatigue, we randomized the sequence in which the MAR prototypes were introduced to them. Each experiment consisted of six phases (see Figure 6):

Phase 1: In this phase, the participants were briefed on the study setup, ethical information and the various phases of the UX evaluation. The participants' written consent had also been obtained in advance. Later, the participants took the SUXES background questionnaire, which asked them about their usage of the Internet (3G, 4G, WLAN), installed mobile applications, use of social networking sites, location based reminder systems, Map and VoIP services.

Phase 2: The concept of MAR was introduced through a power point presentation, which included different use cases and scenarios on MAR and its potential daily use.

Phase 3: The participants were asked to take the SUXES expectation questionnaire and the AttrakDiff questionnaire after introducing the MAR as a notion. In this phase, the participants were briefly interviewed in order to gain their qualitative feedback. The structured interview contained three questions that explained why the participants selected a particular answer in the distributed questionnaires.

Phase 4: One out of the four prototypes was shown to the participants in order for them to test its functionalities. Afterwards, the participants were asked to answer Emocard in order to evaluate their emotional response towards the prototype. Later on, the participants were interviewed on similar questions as in the previous phase. The above process was repeated for the four prototypes (selected and revealed in a random order to each subject).

Phase 5: After testing each MAR prototype, the participants again answered the SUXES and AttrakDiff questionnaires.

Phase 6: In the last phase, the participants answered the AttrakDiff questionnaire in order to gather their experience of all prototypes altogether.

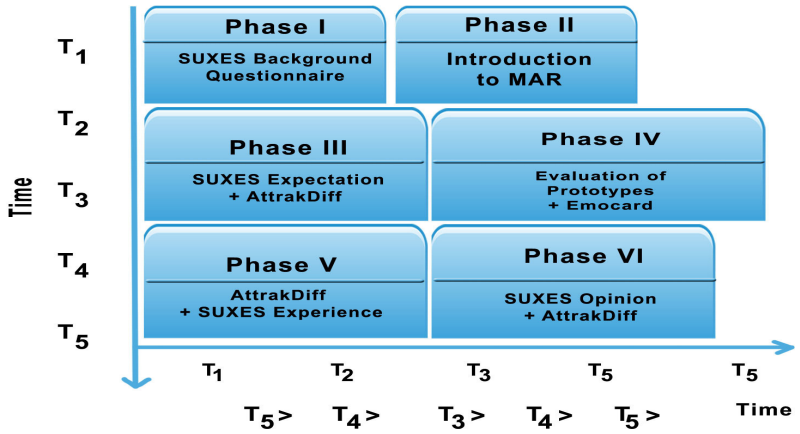


Figure 6: Experimentation Phases

5 Study Evaluation Results

During the UX evaluation study, we gathered both qualitative and quantitative data from the three different UX evaluation methods. The qualitative data enabled us to assess the participants' feedback on different prototypes and enabled us to understand how the participants assessed the usefulness of the different stimulus material.

The UX evaluation results are presented in three parts. The first part of the results reveals the differences between the different prototypes. The second part reflects on the overall experience with the prototypes in general. The third part shows the discrepancies found between the expectations of the users versus their experiences after testing the MAR prototypes. The following subsections provide more details.

5.1 Instant Responses to Prototypes

All the prototypes were appreciated by the participants and were regarded as pleasant, easy to use, enjoyable and nice. The only exception was the "*Playfulness"* prototype, which was discerned as "*unwanted*". The *"language barrier"* was highly appreciated and received the maximum positive responses compared to all other prototypes (see Figure 7). The "*MAR navigation"* received higher *"pleasant"* responses in comparison to the "*MAR Street view"*. The qualitative data gathered from the interview revealed other interesting facts about the UX evaluation that were not captured by Emocard due to its non-verbal nature. The quantitative results on the *MAR "Street View"* show that the app requires further improvement. The participants expected more than the simple features of location tagging and address display.

"*Quite good. I like the concept (male, 24)*"
"It should be more informative. The address thing is good. It gives the exact address (Female, 26)"
"Intuitive and gives a lot of good experience (Male, 28)"

The qualitative feedback received on the "*MAR Navigation"* showed that the overall concept was appreciated, but that participants expected more than just a simple presentation of the information. Several participants mentioned that MAR should provide customization and personalization while its content. Some of the main responses were:

"*Digital information on the buildings is informative, but it should also provide further information such as a website of that building, direction, etc. Tags could include some more information (Female 26)*"

The concept of a virtual pet supported through the "*Playfulness*" prototype was rejected by the majority of the participants. The participants did not regard the virtual pet as a utility, nor did they think it was necessary to have and use this MAR feature.

"*Instead of having virtual pet, I would like to see a virtual tutor who helps me in finding places where I can go for lunch, help me in my studies and talks to me when I feel bored (26 Male)*"
"*It is difficult for me to visualize how it is going to work, and how it will be used in daily routines (Male, 27)*"

The qualitative results shows that the *MAR Language barrier* prototype represents a true utility value for any prospective MAR users. Although not based on a large number of users, the quantitative responses to Emocard (see Figure 7) further confirm the general overall approval; the participants found this prototype to be the most pleasant and exciting of them all.

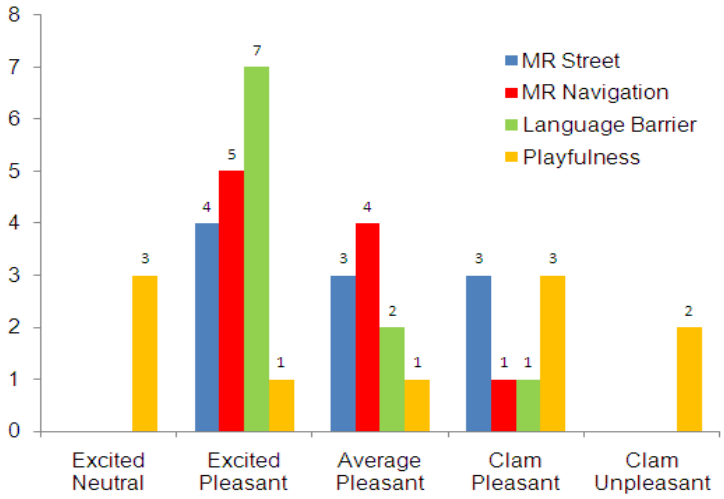


Figure 7: Emotional responses to different MAR prototypes

"*Very much information, especially when you are going to China (Female, 26)*"
"*Informative as language is a big barrier in this rising world so it would be very much useful (Male, 24)*".

The Emocard based UX evaluation and the qualitative feedback based on a short structured interview imply that the users appreciate those prototypes or the core technology behind them.

5.2 Pragmatic and Hedonic Aspects of Prototypes

In this section, we summarize the UX evaluation for all the prototypes in general. We have outlined the "AttrakDiff-3" results from the participants' responses after evaluating the MAR prototypes. We found that AttrakDiff 2 (i.e. answered after evaluating *MAR Street View*) and AttrakDiff 3 do not present any significant difference in the PQ, HQS and HQI distributions. This indicates that the distribution between the different types of prototypes was similar enough to reach this conclusion. The UX evaluation based on AttrakDiff 3 is also a metric for the overall AR as the PQ, HQS and HQI distribution in AttrakDiff 3 also represent the evaluation of all MAR prototypes.

5.2.1 Pragmatic Quality (PQ)

The quantitative results reveal that MAR prototypes received high ratings in different PQ such as *appeal, structure, manageable, captivating, novel, innovative* and *motivating* (see Figure 8). This implies that the participants reacted positively to the MAR prototypes and that they appreciated the content, as well as the overall utility of MAR. PQ such as "*undemanding-challenging"* and "*bold-cautious"* received moderate or in-between ratings. It could be possible that either the participants did not

understand the "*undemanding-challenging*" and "*bold-cautious*" adjective pairs in AttrakDiff, or that they truly thought that the MAR prototypes lie in the middle of the adjective pairs. The quantitative results were further validated from qualitative comments such as:

"Informative…Language is a big barrier. It would be very useful", (Male, 24)
"Core idea is very good but the MAR must possess ease of use" (Male, 22)

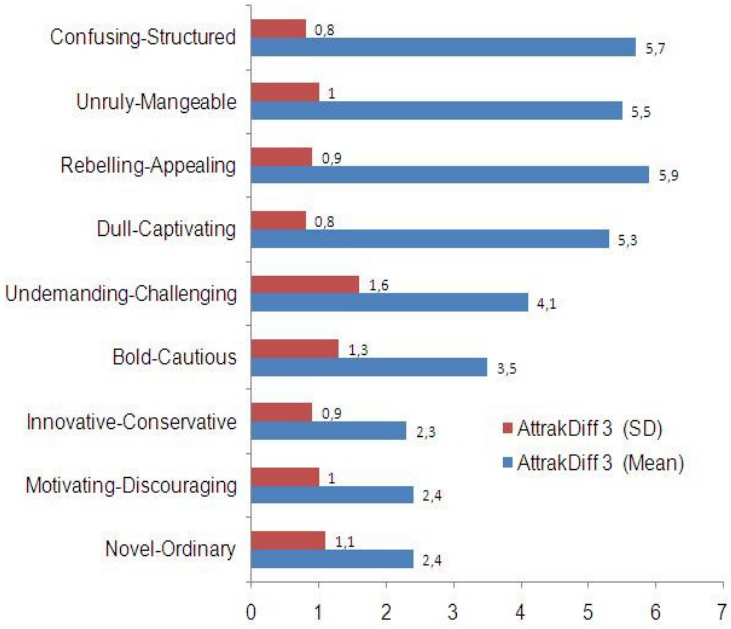


Figure 8: PQ in AttrakDiff 3

5.2.2 Hedonic Quality Identification (HQI)

The quantitative analysis of the HQI adjectives shows that the MAR prototypes were considered *presentable, integrating, creative, inviting, good, stylish*, and *predictable* (see Figure 9). This shows that the MAR, as a whole, was considered good, stylish and also aided social integration. However, the HQI adjectives such as "*cheap-premium*" and "*bring closer-separate*" received in-between ratings in the participants' responses. A possible reason could be that the *"cheap-premium"* adjective was unclear to the participants, and hence this adjective received an in-between response rating. We revisited the transcribed interview results that also show that terms "*cheap-premium"* and the "*bring closer-separate"* were confusing to some users. Some of the notable qualitative comments were:

"This technology takes me away from real world. It's as if I get addicted to MAR", (Female, 26)
"MAR is cool and it can be a style statement", (Male, 24)

5.2.3 Hedonic Quality Stimulation (HQS)

The quantitative analysis of the HQS response rate shows that MAR prototypes were recognized as *attractive, straightforward, connective, professional, inventive, simple, likeable, pleasant* and *practical* (see Figure 10). However, the participants believe that MAR services in general are in-between *human* and *technical*. This shows that users appreciate MAR apps as pleasing and straightforward but still consider MAR services as *technical* rather than natural. This statement is also confirmed by the qualitative data extracted from the participants' comments.

"The interface is pretty much OK and it is pleasant to use", (Male, 27)
"Mobile crazy people might be using it daily", (Male, 27)
"MAR seems like a lot of technical stuff", (Male, 22)

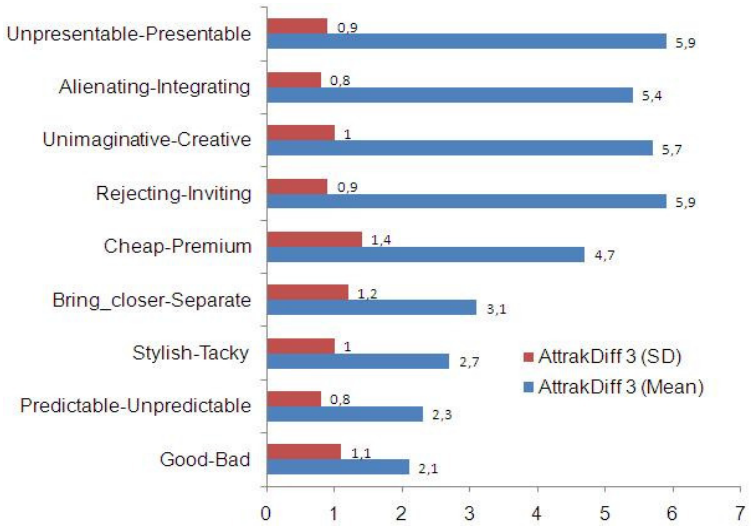


Figure 9: HQI in AttrakDiff 3

The AttrakDiff questionnaire's pragmatic and hedonic quality evaluation of the MAR prototypes that were shown reflects the fact that the MAR service, in general, has the potential to serve user needs and become part of their daily routine. This is because MAR prototypes cater to the users' pragmatic, hedonic identification and stimulation needs. The pragmatic and hedonic quality evaluation also shows that the UX of MAR prototypes can be enhanced by improving the MAR prototypes based on user feedback and resolving their concerns about MAR.

5.3 Expectations vs. Actual User Experience

We used the SUXES questionnaire to analyse user expectations and post-use perceptions. Our objective was to answer the question of *how the actual use experience differed from the pre-use expectations.* We found that expectations did not ultimately differ much from the use experiences. We also carried out comparison of

AttrakDiff 1 (i.e. expectations before evaluating prototypes) versus AttrakDiff 3 (i.e. use experience after evaluating MAR prototypes). The comparisons did not show any significant differences and, hence there is no need to present the AttrakDiff 1 vs. AttrakDiff 3 comparison here. Therefore, this research question remains open for discussion in our future research, along with the research settings, which will allow us to perform such comparisons with greater reliability.

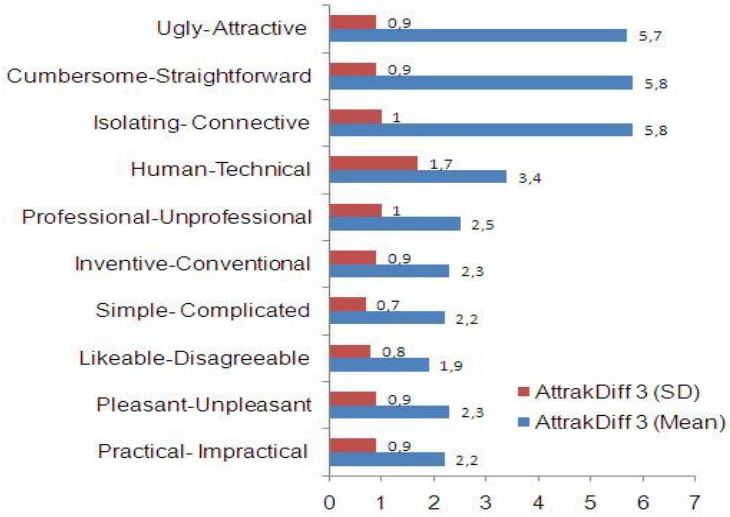


Figure 10: HQS in AttrakDiff 3

5.3.1 Expectations vs. User Experience based on SUXES

After considering various UX evaluation tools, it was concluded that SUXES is a known tool for measuring the temporal UX through the evaluation of pre-use expectations and post-use experiences. Through SUXES, we want to evaluate *how well participant expectations have been met* in regard to the MAR prototypes. The SUXES method contains two questionnaires, namely expectation and perception, both containing the same set of statements but differing in two respects. First, the expectation questionnaire is answered before the evaluation of the prototypes, while the perception questionnaire is answered after the evaluation. Second, the expectation questionnaire is answered with two types of ratings, "acceptable" and "desirable" for every statement, while the perception questionnaire is answered by one rating called "use experience".

The difference between the perceived and the desired ratings is referred to as the Measure of Service Superiority (MSS), the difference between perceived and acceptable ranking is called the Measure of Service Adequacy (MSA), and the range between acceptable and desirable ratings is called the Zone of Tolerance (ZOT). If the experiences are located inside the ZOT scale, MSA is positive (i.e. the perceived ranking is higher than the acceptable rating) and if MSS is negative (i.e. the desired rating is higher than perceived ranking) then it can be concluded that the participants' expectations have been met very well. Figure 11 shows the results obtained from the

SUXES questionnaires. The MSA value is positive for all the SUXES statements, the UX value lies in the ZOT scale and the MSS is negative. This shows that the expectations of the participants are met in all ten of the SUXES statements.

Metric	Acceptable	Desirable	UX	ZOT	MSS	MSA
Diff. possibilities of interaction	4,8	6,3	5,9	<48 - 63>	-0,4	1,1
Fast to use	4,6	6,2	5,1	<46 - 62>	-1,1	0,5
Useful in daily routine	4,4	6	5,3	<44 - 6>	-0,7	0,9
Intuitive & easy	5,1	6,4	5,9	<51 - 64>	-0,5	0,8
MMR is simple to use	4,7	6,4	6,1	<47 - 64>	-0,3	1,4
Usage at public places & acceptance	4,8	6	5,9	<4,8 - 6>	-0,1	1,1
Stylish & style statement	4,4	5,3	4,9	<4,4 - 5,3>	-0,4	0,5
Fascinating to use	5,1	6,8	5,2	<5,1 - 6,8>	-1,6	0,1
Connecting with other people	4	5,7	5,4	<4 - 5,7>	-0,3	1,4
Innovative technology	4,7	6,4	5,3	<4,7 - 6,4>	-1,1	0,6

Figure 11: SUXES expectation versus actual use experience Pragmatic Quality (PQ)

6 Discussion

The UX evaluation results suggest that concreteness, realizability, level of interaction supported, personalization, novelty, intuitiveness and utility are some of the deciding factors in the users' expectations and use perceptions in the context of MAR applications. These findings further support the earlier findings in the domain of the UX of MAR. Olsson et. al. [2009] presented new insights into users' expectations of MAR services. They found usefulness, personalization and reliability of information content as deciding factors in designing the UX of MAR services. Our UX evaluation study has also shown that these factors also define UX to some extent although these needs are mostly pragmatic in nature. Understanding user expectations potentially helps in designing UX at the concept level implementations and prototyping.

Heikkinen et al. [2009] suggested studying user expectations when technology's focus in the future will be to give an approximation of the actual user experience even before the practical realization of the MAR-based services. Our UX evaluation results further validate this previous finding in the domain of UX. We found that our MAR prototypes received better results in the UX evaluation because our prototypes were based on user needs and expectations gathered during the user research. Furthermore, iterating MAR prototypes based on existing UX evaluation results can further enhance the UX of MAR prototypes.

The Emocard results suggest that test users appreciated all the MAR prototypes, with the exception of the MAR Playfulness app, as they were pleasant, easy to use, to some extent intuitive and showed usefulness. The AttrakDiff PQ, HQI and HQS adjectives and qualitative interview have stated the underlying reasons for the

pragmatic and hedonic experience of the MAR prototypes. Finally, SUXES helped in analysing the difference between the user expectations and post-use perceptions of the prototypes.

The MAR Street View and *MAR Navigation* apps were recognized as pleasant and easy-to-use. However, the participants in their qualitative opinions mentioned that both prototypes required improvements in their functionality and supported content. The *MAR Language barrier* was the most highly appreciated among all the MAR prototypes. It received the maximum positive quantitative responses in Emocard. Furthermore, in the qualitative interview comments, the users mentioned that they appreciated the *MAR Language barrier* as they could see its usefulness and need in their daily routines. Most users underappreciated the *"MAR Playfulness"* because it did not meet their needs and did not manifest its usefulness in their daily life.

In the UX evaluation study, we found that the majority of the user needs were pragmatic in nature when they first interacted with the MAR prototypes. For example, during the Emocard based evaluations, most of the participants made such qualitative comments as *easy-to-use possess usefulness* and *can be used in daily life*. However, the hedonic needs were only elicited when the participants responded to the AttrakDiff questionnaire adjectives, especially in HQS and HQI. As the participants had only a short-term interaction with the MAR prototypes, the hedonic needs were discussed less often than the more pragmatic and immediate needs of the participants.

Our study results can be exploited as design guidelines to create intelligent user interfaces for MAR applications. For example, the study results suggest that any application concept should be intuitive, novel, concrete, and realizable, support different levels of interaction, personalized, and possess usefulness in order to build up a good UX of the product when it is eventually delivered to the consumer. Factually, the product designers could benefit from our UX evaluation study and the underlying process as our UX evaluation assists in evaluating the difference between user expectations and post-use experiences. We chose SUXES, Emocard and AttrakDiff because they complemented each other and offered rich insights into different user expectations and post-use perceptions.

All our prototypes were based on AR in the smartphones context, which helped our prototype, had a appreciable impact on the users' daily routines. The AR's basic nature was considered to be novel, intuitive, structured, attractive, creative, stylish, appealing and pleasant to use. This very basic nature of AR had an impact on the UX evaluation results too, which was manifested by some of the participants' qualitative comments shown below.

"MAR is very high tech and intuitive", (Male, 24)
"When will this MAR become available in common market?" (Male, 23)

It was noticed that as the UX evaluation progressed, service features namely concreteness, personalization, level of interaction and usefulness became paramount. This study is valuable as it provides an insight into how to perform a triangulation method in case of UX evaluation. Triangulation strategy argues for combining qualitative and quantitative methods in order to examine a particular phenomenon in further detail [Jick, 1979]. The validity of the results can be questioned as the test participants were technology biased and may not reflect the experiences of non-

technical users. Additionally, social or other related aspects to a real usage of the prototypes were not considered as tests took place in a laboratory setting.

7 Conclusions

The main contribution of this paper is the new insight into the user expectations and use perceptions for early producers of MAR. Through our UX evaluation study, we were able to elicit the technology and design level needs in regard to MAR services. Understanding user expectations and use experiences for a novel and immature technology like MAR is challenging. We consider our UX evaluation study a necessary step towards understanding user expectations versus experience in the MAR applications context. So far the research on UX in MAR services is limited only to understanding user needs and expectations through the use of limited use cases supporting limited interaction. However, in our study, we have extended the already existing work by practicing three different UX evaluation methods against early MAR demonstrations.

We trust that our choice of research methods, study process and UX evaluation can also be used for the UX evaluation of other technology concepts that are futuristic and novel. Mobile application developers, UX researchers and practitioners, and consumer companies who are interested in the UX evaluation of novel and futuristic technologies may adopt our process model of evaluation which included: 1) Providing concrete and visually realizable prototypes for UX evaluation 2) Giving enough training to test participants in UX methods such as AttrakDiff and Emocard, and 3) Practicing the triangulation research principles consisting of subjective, objective and emotional UX evaluation methods.

We believe that all the above procedures and precautions enabled us to access and improve the UX of early MAR demonstrations. This study helped us reach the conclusion that MAR apps can provide pleasing, enjoyable, rich and positively surprising experiences to its users. However, and per our findings in this research, it is imperative for MAR application developers and service companies to understand the design implications on the user experiences at the early phases of the conceptual design and prototyping.

References

[Azuma, 1997] Azuma, R. T. 1997 A Survey of Augmented Reality. Presence: Tele-operators and Virtual Environments. vol. 6, No. 4, pp. 355-385.

[Bach and Scapin, 2004] Bach, C., and Scapin, D. L. 2004 Obstacles and Perspectives for Evaluating Mixed Reality Systems Usability. Acte du Workshop MIXER,IUI-CADUI'04

[Desmet et al., 2001] Desmet, P.M.A., Overbeeke, C.J. & Tax, S. J. E. T. (2001). Designing Products with Added Emotional Value; Development and Application of an Approach for Research through Design. The Design Journal, 4(1), 32-47.

[Dhir et. al., 2012] Dhir A, Olsson T, Elnaffar S. Developing Mobile Mixed Reality Application Based on User Needs and Expectations. Proceedings of 2012 IEEE International Conference on Innovations in Information Technology (IIT), United Arab Emirates, 18-21 March, 2012. pp. 83-88.

[Forlizzi and Battarbee, 2004] Forlizzi, J. and Battarbee, K. (2004) Understanding experience in interactive systems. In Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (DIS '04). ACM, New York, NY, USA, 261-268

[Friedrich, 2002] W. Friedrich. 2002. ARVIKA Augmented Reality for Development, Production and Service. In Proceedings of the 1st International Symposium on Mixed and Augmented Reality (ISMAR '02). IEEE Computer Society, Washington, DC, USA, 3-16.

[Hassenzahl and Tractinsky, 2006] Hassenzahl, M. and Tractinsky, N. (2006) User Experience – A Research Agenda. Behaviour and Information Technology, Vol. 25, No. 2, pp. 91-97.

[Hassenzahl, 2003a] Hassenzahl, M. B. M. K. F. Attrakdiff: Ein fragebogen zur mes-sung wahrgenommener hedonischer und pragmatischer qualitat. B. G. Teubner, Ed., Proceedings of Mensch & Computer 2003: Interaktion in Bewewgung. Stuttgart, pp. 187-196.

[Hassenzahl, 2003b] Hassenzahl, M. The Thing and I: Understanding the Relationship Between User and Product. In Blythe, M., Overbeeke, C. J., Monk, A.F., & Wright, P.C. (Eds.), Funology. From Usability to Enjoyment, Kluwer Academic Publishers, Dordrecht, 2003, 31-42.

[Heikkinen et al., 2009] Heikkinen, J., Olsson, T., Väänänen-Vainio-Mattila, K. Expectations for User Experience in Haptic Communication with Mobile Devices. In Proc. Of MobileHCI 2009, Bonn, Germany, Sep. 15-18, 2009.ACM Press.

[Höllerer and Feiner, 2004] Höllerer,T.H., and Feiner,S.K. 2004. Mobile Augmented Reality. Tele-geoinformatics: Location-Based Computing and Services. Taylor and Francis Books Ltd., London, UK

[ISO 9241-210, 2009] ISO 9241-210, 2009. Ergonomics of Human System Interaction – Part 210: Human Centred Design for Interactive Systems (formerly known as ISO 13407), Switzerland International Organization for Standardization (ISO).

[ISO/IEC 13407, 1999] ISO/IEC. 13407 Human-Centred Design Processes for Interactive Systems, 1999, Switzerland International Organisation for Standardiza- tion(ISO)

[Jick, 1979] Jick, T. D. Mixing qualitative and quantitative methods: Triangulation in action. Administrative Science Quarterly vol. 24, no. 4, Qualitative Methodology (December, 1979), pp. 602-611.

[Jordan, 2000] Jordan, P. W. Designing Pleasurable Products: An Introduction to the New Human Factors. Taylor & Francis, London, UK, 2000. 216p

[Kuutti, 2010] Kuutti, K. 2010. Where are the Ionians of user experience research? In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10). ACM, New York, NY, USA, 715-718.

[Kankainen, 2003] Kankainen, A. 2003. UCPCD: user-centered product concept design. In Proceedings of the 2003 conference on Designing for user experiences (DUX '03). ACM, New York, NY, USA, 1-13.

[Karapanos et. al., 2009] Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J. M. 2009. User experience over time: an initial framework. In Proceedings of the 27th international conference on Human factors in computing systems (CHI '09). ACM, New York, NY, USA, 729-738.

[Karat et. al., 1996] Karat, J., Atwood, M. E., Dray, S. M., Rantzer, M., and Wixon, D. R. User centered design: quality or quackery? In Conference companion on Human factors in computing systems: common ground (New York, USA, 1996), CHI '96, ACM, pp. 161-162.

[Milgram and Kishino, 1994] Milgram, P., and Kishino, A. F. A Taxonomy of Mixed Reality visual Displays. vol. E77-D(12), IEICE Transactions on Information and Systems, pp. 1321-1329. 1994

[Olsson et. al., 2009] T. Olsson, P. Ihamäki, E. Lagerstam, L. Ventä-Olkkonen, K. Väänänen-Vainio-Mattila. User expectations for mobile mixed reality services. In Proc. of ECCE'09 (Helsinki, Finland, Sep. 30-Oct. 2 2009), pages 177-184, ACM Press, 2009.

[Roto et al., 2009] Roto, V., Rantavuo, H., Väänänen-Vainio-Mattila, K. Evaluating User Experience of Early Product Concepts. In Proceedings of International Conference on Designing Pleasurable Products and Interfaces, DPPI'09 (Compiegne, October 2009)

[Stanney, 1995] Stanney., K. 1995. Realizing the full potential of virtual reality: human factors issues that could stand in the way. In Proceedings of the Virtual Reality Annual International Symposium (VRAIS'95) (VRAIS '95). IEEE Computer Society, Washington, DC, USA, 28-34.

[Turunen et. al., 2009] Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., and Hella, J. Suxes - user experience evaluation method for spoken and multimodal interaction. In INTERSPEECH (2009), pp. 2567-2570

[Tahti et. al., 2004] Tahti, M., Vainamo, S., Vanninen, V., and Isomursu, M. Catching Emotions Elicited by Mobile Services. In Proc. Australian Conference on Computer-Human Interaction, Wollongong, Australia. (CD-ROM) 2004

[Vaittinen et. al., 2010] Vaittinen, T., Karkkainen, T., and Olsson, T. A diary study on annotating locations with mixed reality information. In Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia (New York, USA, 2010), MUM '10, ACM, pp. 21:1-21:10

[Vredenburg et. al., 2002] Vredenburg, K., Mao, J.-Y., Smith, P. W., and Carey, T. A survey of user-centered design practice. In Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves (New York, USA, 2002), CHI '02, ACM, pp. 471-478.

[Weiser, 1993] M. Weiser, Hot Topics: Ubiquitous Computing, *IEEE Computer*, October 1993.

[Weiser, 1991] Weiser, M: The computer for the 21st century, Scientific American 265, 3, 1991, 94-100.

[Zhang et. al., 2000] Zhang, X., Navab, N., and Liou, S.-P. 2000. E-Commerce direct marketing using augmented reality. In Proc. ICME2000 (IEEE Int. Conf. on Multimedia and Exposition), New York, NY.