Tallinn University

School of Digital Technologies

ASSESSING THE INFLUENCE ON USER EXPERIENCE OF WEB INTERFACE INTERACTIONS ACROSS DIFFERENT DEVICES

Master's Thesis

by

Erkki Saarniit

Supervisors:

Mati Mõttus

and

David Lamas, PhD

Tallinn 2016

Confirmations:

Author:

May ___, 2016 ..... /Erkki Saarniit/

Supervisors:

May ___, 2016 ..... /Mati Mõttus/

May ___, 2016 ..... /David Lamas/

Director of the Institute:

May ___, 2016 ..... /Peeter Normak/

I herewith declare that I have written the Master's Thesis independently and it has not submitted for any defence previously. References have been indicated for the all publications, claims, opinions and different sources by other authors in the References section.

May ___, 2016 ..... /Erkki Saarniit/

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Having already received one master's degree five years ago, the second master's studies for me had mainly one goal: to learn more about a topic I come in contact with on a daily basis, by widening the theoretical background of the human computer interaction (HCI) area, and by knowing more practical implementations of the theories.

I am very glad that the studies in Tallinn University have given me exactly what I was expecting, and even more. Small study groups, international diversified background of students and lecturers succeeded in surpassing my expectations towards the experience received from the programme. My special thanks go to the initiation and leadership of Dr. David Lamas who is a professional, smart and motivational person.

The time put into the studies and the master's thesis can be considered as an investment with high returns. The area of the thesis, related to user experience created from web interactions, opened a new area for me that was not widely discussed during the lectures of the programme. The main supervisor of the thesis, Mati Mõttus, inspired by moving substantial focus towards the aesthetics and feelings of user interactions. After all, the user experience created by an interface, has a significant impact on the feelings, which overall tells how much the person likes one system. Therefore, my biggest thanks go to Mati for contributing great amount of his time, and sharing his knowledge and relevant suggestions about the topic.

I also thank my friends and relatives to whom I was less available during the writing of the thesis. Additionally, my gratitude goes to all study participants who took the time to be part of the study.

Significant impact on the experience I have received is from my previous and current employers, Swedbank and Inbank. The inspiring work I have done there has affected

me to focus more on the area, and definitely has its role on the fact that I started human computer interaction studies at all.

For me, learning is a lifelong process. Sometimes, to focus, you just need to push yourself. And sometimes you want to push yourself, to keep your brain fresh and force it to think differently, and come out of the comfort zone.

Everything is possible. Or at least, nothing is impossible…

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AD	AttrakDiff question (e.g. AD12 is AttrakDiff's 12th word pair)
ATT	AttrakDiff group of attractiveness (questions 22-28)
EP	Episode (EP1 is Episode 1)
HCI	Human Computer Interaction
HQ	AttrakDiff groups HQ-I and HQ-S combined
HQ-I	AttrakDiff group of hedonic qualities - identification (questions 8-14)
HQ-S	AttrakDiff group of hedonic qualities - stimulation (questions 15-21)
OS X	The operating system of Apple's computers (Macs)
PQ	AttrakDiff group of pragmatic qualities (questions 1-7)
S	Study participant who was involved in the study (e.g. S131 is a participant code)
UI	User Interface
UEQ	User Experience Questionnaire (Laugwitz et al 2008)
UX	User Experience

1. INTRODUCTION

In the past years the spread of touchscreen mobile devices such as smartphones and tablets has grown rapidly, although traditional desktop and laptop computers are still used a lot (eMarketer Inc. 2015; Smart Insights Ltd. 2016). Consequently, designers and developers of websites are facing challenges, including the choice of platforms and technologies, taking into account the variety of screen sizes (Sukale et al 2014). The reason these challenges exist is that users expect to have a system that is easy to learn and use (Karapanos 2010), and such variety of platforms and technologies make good user experience more difficult to achieve when building a user interface (UI). Despite the users might not recognise or understand why usability is important, perception of a system is often influenced by the way it is designed (Buxton 2010). The research carried out in the thesis investigated one of the challenges that designers and developers face: creation of websites with interactive elements.

Building a website with sophisticated interaction elements can be a matter of choice, preference or necessity, as nowadays using widespread libraries make sophisticated interactions as easy to develop as simpler ones (jQuery Foundation 2016). Interactions can occur between several web pages or within one web page. For example, the way how a simple shift from one web page to another can be designed, can differ from another method significantly. How the link on the initial web page is clicked, does it change colour when mouse cursor is moved on the link, what happens to the link when it is clicked, what occurs before loading the next page and how much does it take time – all of these are examples of small details that can have influence on user experience (e.g. Charland and Leroux 2011; Seipp and Devlin 2014; Kermeka et al 2013).

Even more differences can appear in interactions that happen within one web page. For example, one of the most common interactions, navigating through a long page,

can be done using menus, links, dragging scrollbar, using a touchpad or a finger to scroll up or down, left or right. Elements on a page can appear in an order, they can be layered on top of each other, they might even be movable. While interacting, colours, transparency or other styling properties of related elements can change. There can be animations while loading or leaving the page or while focusing from one element to another.

There are variations in how interactions are built in different devices. For example, touchscreens are mostly interacted with using a finger, whereas desktop computers or laptops are mostly handled by using a mouse (Nielsen 2012). The difference between the size of a finger and the size of a mouse cursor creates limitations on the size of interactive elements (Google Inc.: Material design guidelines 2016). Interactions that are dedicated not only for desktop computers or laptops but also for touchscreen devices must take this into account. Moreover, the development of such interactions differs based on the operating system of the device used (Sukale et al 2014). Standards interpreted by web browsers are similar across devices, whereas mobile applications (e.g. iOS, Android, Windows Phone) use the interactions built into their operating systems. For example, like other operating systems, Android has very specific guidelines for layout, interactions and usability (Google Inc.: Material design guidelines 2016). It is also studied that the input device has an influence on memory retrieval (Senecal et al 2013).

In addition to standard input device like mouse, keyboard or finger, each device can have sensors (Charland and Leroux 2011). They are more common on mobile devices. For example, Android's latest operating system supports 13 sensors to be used as part of any action or interaction (Google Inc.: Android Developers: API Guides. Sensors Overview 2016). These include accelerometer, gyroscope, illumination, temperature and proximity sensors. A specific interaction can happen when a smartphone is tilted at a certain angle, thrown with certain acceleration, is in a pre-defined temperature level, and is is 20-30 cm from a human being. A relation between sensors and user experience has proven to exist (e.g. Froehlich 2007; Chent et al 2013; Rajanna et al 2014).

The multiplicity of potential interaction influencers creates new layers of complexity, but also opportunities for better and more intuitive interactions.

Consequently, there is a variety of interaction possibilities available, and they are partly device, operating system or browser dependent (Heitkötter et al 2012). Developing a set of interactions that has considered all the possibilities the platform offers, adds constraints on and can complicate the development. Therefore, realising all possibilities can raise time to market and the cost of the system.

The author in his career has faced many challenges in the field of web interactions, mostly similar as the ones described in the previous paragraphs. For a developer, a designer, or a person responsible for sales or customer acquisition there are many choices available and questions to be answered regarding web interactions. The author has been in each of the roles and worked together with people in these roles. Many of the end-users of the systems the author has been involved in, are people who work in an office, and use a computer and a smartphone on a daily basis. Even among such group, the differences in the perceptions of the same system seem to differ considerably. Therefore, the author has a personal and professional interest to study such challenges in more detail and to focus the research on the evaluation of web interactions in computers and smartphones.

The problem this thesis addresses is that designers and developers would like to create web-based solutions that would work on different devices, but the user experience may differ across the devices. Designing one solution is easier than designing for all possible devices and ways of interacting. The author has seen that designers and developers are facing the options in their work on a daily basis and they often choose between the options based on their own opinion or experience, but not on scientific analysis. This thesis aims to give designers and developers some additional, but not ultimate input to the decision making process of how to build web interactions.

The main research question is **how users perceive different web interactions.** As the sole completion of interaction cannot be the only aspect that makes users satisfied (Hassenzahl and Tractinsky 2011), this study investigates what users perceive to be

important qualities of interactions that raise users' overall satisfaction with the web application.

Only in the past decades, people have got used to keyboards and compute mice, whereas hands and fingers have been always used to directly interact with objects. Nevertheless, the difference between the efficiency of a mouse and the efficiency of a finger depends on the specific interaction completed by the user. Nielsen has made a comparison (2012) of these two input devices, and has claimed that both have their pros and cons. For example, a human finger wins over a computer mouse in being multi-touch enabled, easy to learn to interact with, it has direct engagement with a screen, and of course, a user does not have to carry anything. A mouse also has advantages compared to the finger. It is much more precise, it has multiple functions (left and right click, scrolling), it supports the measurement of accelerated movements and it is more usable in case of bigger screens. Such differences can provide different user experience and should be considered while designing.

Karapanos has explained (2010) that the difference between clinical psychology studies and user experience evaluation lies in the importance of idiosyncrasy. He discussed that in UX evaluation "the interest is not in the idiosyncratic views of an individual but rather on some more-or-less homogeneous groups of individuals." Hence, the individual evaluations of web interactions need to be somehow grouped, although individual analysis also contributes to understanding the interactions better.

Considering the complexity of the area, the study aims to find out which pragmatic and hedonic qualities differ in user interactions that are similar to each other.

Furthermore, the thesis explores **whether there are any differences in the perceptions of web interactions among computers and smartphones**. Devices can have different goals for which they are used. In addition, they are used and interacted with in a different manner. But if the interaction is the same, does the device itself matter?

The **hypothesis** proposed in this thesis is that **it is possible to design web interactions that provide homogeneous user experience across different devices**.

Although the users may be different, it is expected that a solution where experience does not significantly differ between the experience provided by a computer and a smartphone interaction can exist. As an additional hypothesis, it is also expected that user experience evaluation can provide instructions for designers.

In the following chapters of this thesis, the author reviews the research of other authors, describes the research design, the conducted study and its results. In the analysis chapter the author brings out the specific measurements of user experience, and their conclusions, so that the thesis can give input to the designers and developers.

2. LITERATURE REVIEW

Although common practices, wide-spread code libraries and expert opinions exist, the author has found no absolute answers based on scientific research how user experience is influenced in web interactions across different devices, considering the specifics of the research topic.

The purpose of the review of literature is to give an overview of relevant topics that have been studied previously, and to bring out the results of these studies. It seeks to find out how user experience is evaluated, specifically in web interactions and across different devices.

In this chapter the approach to the review is described, the findings from the literature are presented and a conclusion of the review is made.

2.1. Approach to the Literature Review

A literature review should "provide the reader with what the researcher did during the literature review input" and "what he or she has learned during the literature review processing" (Levy and Ellis 2006). Furthermore, a three-stage approach to the review was proposed by Levy and Ellis: 1) inputs, 2) processing and 3) outputs. Several of the suggestions proposed in the study were taken into account while conducting the review, such as searching, managing and organising the literature. Also, the most common violations of academic writing standards brought out in the study such as falsification or sloppiness were tried to be avoided during the author's literature review.

The focus of the literature review is on finding relevant information on the **user experience evaluation**. As the research is based on the evaluation of UX of web-based interactions across different devices, it should be defined what the **web-based**

interactions are, what properties they have and how the interaction differ **across the devices**. Related studies are reviewed to understand the similarities and differences of the findings compared to the author's research.

The literature review was approached by firstly searching different databases with keywords that are related to the areas that needed to be studied in order to answer the research questions. This also included the search for potential methodologies.

Primarily ACM Digital Library was used for finding the literature. In some cases, also Google Scholar search was used. Filters based on whether the paper was published through a journal, magazine or a conference were not applied.

The initial keyword used in the search was "user experience evaluation", because the research topic is related to UX evaluation. The word "user experience" was also replaced with "UX", "evaluation" replaced by "assessment"; also different wordings were tried out, such as "evaluation of user experience". The search was not limited to titles or other fields: all fields were searched, including the full text. Such search resulted in 118 articles found. All titles and abstracts of the articles were looked through and all relevant articles were saved for later analysis.

Additionally, combinations of "user experience" and other keywords related to the research were used. For example, combinations of "user experience" and "web", "webpage", "interaction", "device", "mobile", "application", "android", "iOS", "desktop", "laptop", "input device", "aesthetic", "pragmatic", "browser", "native", "touchscreen", "HTML", "CSS", "JavaScript", "jQuery" were searched for. Such combinations were used, because they help to find papers relevant to the research question for the part of the web interactions and devices. Some of the keywords focus on different operating systems, thus they help to find relevant devices. Some web programming languages or libraries were used for the keywords to find papers related to different devices.

Usually each combined search produced from tens to hundreds of results. In case there were more than 100 results, additional keywords were combined to reach the most relevant papers. Sorting of results was done mostly by relevancy, in some cases by

citation count. Altogether approximately one thousand titles and 50% of their abstracts were looked through.

In mobile development there are two main development categories: native applications and web-based applications (Charland and Leroux 2011). Native applications are directly saved to the device and are based on the operating system. They are acquired through an application store, and they do not necessarily require internet connection. Web-based applications are usually opened via a mobile web browser and they need internet connection to display the content, with some exceptions.

To narrow down the searched papers even more, the papers that strictly focus on the physical products and industrial design were left out, because the focus of this research is related to web interactions.

After analysing the results of the initial and the combined keyword search 103 relevant papers were found. The papers were saved in Mendeley application for easier management. All of these papers were either read through fully or partly. In case of researches that consisted of different aspects and only some aspects were relevant to the study of this thesis, only the relevant parts were focused on. As a result of the analysis, 33 publications were selected to be used and referred to in this research due to their relevance.

The literature review resulted in no findings of any studies that are truly similar to the study of this thesis. One important finding is that the research questions can be answered using the methods that have already been widely studied and used in practice. Nevertheless, the author sees the need to combine different methods to reach and formulate the conclusions.

2.2. Findings

2.2.1. Designing for Web

One of the challenges of designing for web is to get the right design and also design it right (Buxton 2010). Sukale et al (2014) brought out the main problem in web design: variety. Considering the way today's screens are produced, they have a wide variety of sizes and screen resolutions. Even more, they have different distances they are watched from. Touchscreens are used usually within a hand's reach from a person, but big TV screens are watched meters away. Sukale et al explained that this makes web design complicated not just from the perspective of screen resolution, but also from the perspective of the distance and input method used. One way of solving the problem, brought out by Sukale et al, is responsive design. A designer must take all factors into account, considering not just the pixels, but also the usefulness of the screen space usage.

Another angle of responsive design is design for smaller devices like mobile phones. Although the screens are still small, they have gotten bigger during the past years, making the interfaces more difficult to use with one finger (Seipp and Devlin 2014). Seipp's and Devlin's research also describes that usage of such devices differs based on whether the user is left- or right-handed. Their research gives the following example: right-handed users mostly use their thumb in the right bottom corner, whereas left-handed users use left bottom corner. Taking such details into account raises many questions. Firstly, which level of depth of development should be chosen, and secondly, how the browsers and web standards like HTML and CSS are ready to tackle these problems with similar standards.

One research from Senecal et al (2013) suggested that "need for touch" is different among people, and input devices influence memory retrieval of individuals with higher need for touch. It is brought out that such people can better recognise the stimulus encountered during their navigation.

2.2.2. Design Qualities

As Sanders (1992) has cited, design should be useful, usable and desirable. To put it into today's context of web design, a good design cannot be beautiful, but at the same time not easily usable. Also, a design, although being easy to use, has to have aesthetic enjoyment. As Trevor (2012) has put it, designing for emotion consists of three elements: business goals, technological constraints and user needs. They all have a common share and usually designers have to sacrifice from one in order to focus more on the others. It means that if to provide design for all technological possibilities and take into account user needs, it will be expensive, or if to want to save money, one has to either do less for user needs or focus less on technologies.

Trevor also described, referring to various other studies, how considering emotions in design creation is an important factor in user experience. Any design is emotional and emotions dominate in decision-making. People have a selective attention and are influenced by many characteristics, both consciously and unconsciously. The author of this thesis finds that these factors need to be analysed more thoroughly in the future to understand which of them are relevant for user experience provided by web design.

For developing responsive web, it is important to understand the differences between native and web based applications. It has been widely discussed how native interactions seem better than web interactions. One of the author's goal is to understand the characteristics of the interactions that influence such belief. Charland and Leroux (2011) have taken the topic to the level of code and speed, which is very relevant in the perception of interactions. Yet, the context of interactions cannot be forgotten, as for some products these characteristics might become less important for the user.

Hassenzahl and Tractinsky have stated in their research (2011) that there are three facets of user experience: the experiential, beyond the instrumental, and emotion and affect. They stated the following: "UX is a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g.

organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.)". In addition to highlighting the diversity of UX, it also means that when evaluating user experience, such characteristics should be taken into account. They also state that while at the start of the era of interactions the task just had to be completed, it has taken a decade to get beyond the pragmatics. Therefore, taking the combination of pragmatics and other aspects, such as beauty or usability, into account, has also taken time.

Hassenzahl has also brought out (2013) that satisfaction of instrumental needs is a necessary precondition for valuing non-instrumental needs. For example, in case of web design, a search form can easily do its work by searching and giving the result. But in case the needs of the user or the user's expectations towards the interface are different compared to what the system offers, it can demolish the user experience. For example, if the user who has experienced "instant search results" expects to get some results before pressing enter of the search button, and the system does not provide it, the user experience is already changed in a negative direction.

2.2.3. User Experience Evaluation

User experience evaluations have been conducted using many different methods. Alves et al (2014) have studied how UX has been evaluated, what kinds of methods have been used and what the background of the evaluators has been. That study showed that the most popular methods are observation, think aloud and contextual inquiries, followed by interviews, experience prototyping, task analysis, cognitive walkthrough and questionnaires. The choice of methods depends on the perspective from which the system is studied. Heuristic evaluations have been conducted through interviews, and interviews are often combined with contextual inquiries, observations and experience prototyping. Although it presents the historical view on evaluation, it does not give any ultimate answers what the best method is in which case.

Various methods have been analysed by Kermeka et al (2013) in the context of web. Their analysis studied many methods and standards that had been used so far. These

methods and standards are studied in this thesis in more detail, mostly focusing on the hedonic and pragmatic characteristics.

One of the researches that focuses on measuring hedonic and pragmatic qualities of user experience is written by Hassenzahl et al (2003). The result of this research was a questionnaire called AttrakDiff that initially contained of 21 word pairs that would be assessed by the user. Seven of the word pairs were related to pragmatic qualities, such as assessing whether the object is practical or impractical, simple or complicated. Other 14 questions are related to hedonic qualities, which are grouped into two subgroups: stimulation and identification hedonic qualities. Again, each of the sub-groups contained of seven word pairs, for example, creative or unimaginative, stylish or tacky. Measuring both pragmatic and hedonic qualities is relevant to this research, therefore, it will be analysed further.

AttrakDiff has been used also in different versions. Hassenzahl et al (2000) focused their research in 2000 to appeal, adding 8 word pairs to the questionnaire. As another example, Väätäjä et al (2009) combined the pragmatic and hedonic aspects of the evaluation to appeal, totalling 29 word pairs. Karapanos (2010) has used a shortened AttrakDiff questionnaire. In the current version by the organisation related to the original author, the questionnaire has 28 word pairs, with attractiveness (or appeal) having seven word pairs similarly to other quality groups (User Interface Design GmbH 2013). Overall, it is up to the researcher to find the relevant groups and questions depending on the product that is studied.

After publishing the AttrakDiff questionnaire, it has been used many times to evaluate user experience in scientific research. One of the examples is referred to by Diefenbach et al (2014). In that research it was found that it is very relevant to put the perceptions of usage of a system to a quantitative scale. Stojmenova et al (2013) have used the questionnaire in a study where different web TV interface scenarios were assessed. Also business management software has been tested using the questionnaire (Schrepp et al 2006).

Laugwitz et al (2008) have described a somewhat similar evaluation method UEQ (user experience questionnaire). It has six scales: attractiveness, perspicuity,

efficiency, dependability, stimulation and novelty. Some elements of UEQ have proven to significantly correlate with AttrakDiff.

Evaluation of user experience has to have more quantitative measures than user opinions. As Law (2011) has referred, user actions can be predicted and, therefore, user experience can be predicted. One of the methods to put evaluations to a measurable scale is eye-tracking, which has been used in practice a lot. Bednarik et al (2012) have studied how eye movement is related to user actions and how it can be used to build better systems. In case of different products, such method allows to compare them. It can provide quantitative data about the specific measurement of eye movement and make conclusions about findability, but it does not cover the non-pragmatic part of the user experience evaluation.

System usability scale (SUS) can be used for evaluating usability of a system (Brooke 1996). It is a very simplistic yet widely used scale, because it is easy to measure a variety of systems. The questionnaire contains of ten questions with a five or seven point Likert scale. The advantage of the SUS has shown to be that it measures perceived usability, it is easy to implement and does not require much neither from the researcher nor the study participant.

Lewis (2002) describes the Post Study System Usability Questionnaire that measures factors of system usefulness, information quality, and interface quality. With its 19 questions, it has similarities both to AttrakDiff and SUS, although it is more focusing on the pragmatic qualities of the system.

Comparability of different products is important if the intention is to find out what is better. De Angeli et al (2006) have used an experiment with AB testing to compare web pages with the same information, but with different user interface styles. Similar AB testing has been done with integrating character engagement and user interaction into the I-PEFiC model, which provides an integrative perspective on human-character interaction (Vugt et al 2006). The study used game characters for the comparison.

There are several other UX evaluation methods, such as aesthetics scale, DES, Emocards, hedonic utility scale, UX curve and product attachment scale. The website All About UX (2016) lists a lot of them. Potentially most relevant ones have been analysed in this literature review.

2.3. Conclusion

The literature review can be considered successful. Number of related research was found during the review. In addition to publicly available web pages like All About UX, a great number of scientific researches have been conducted in a similar area.

The approach to the literature review gave many options to search for relevant content. Considering the area of the research topic, other relevant keywords may be used, therefore during further research additional resources may be studied.

None of the study areas was as similar that it could be considered sufficient to answer the research question (and run an additional validation study based on previous one). Consequently, a relevant research has to be defined.

3. RESEARCH DESIGN

Appropriate research design is a cornerstone of achieving reliable results in finding the answer to the research questions. Therefore, the method selection, the principles and requirements for the stimuli selection are described thoroughly in the following section.

In addition, the selection principles of the study participants are introduced. The chapter also explains the selection of data collection method.

The conclusion of the research design is the main pillar for conducting the study.

3.1. Method

Although user experience is widely based on personal opinions and feelings, research provided in the previous chapter refers to the examples of methods how interactions influence user experience.

In assessment of web interactions, the advantage of qualitative research is that with each user the researcher can go very deeply into the topic, revealing the aspects that might not come out using a quantitative research. On the contrary, in case a qualitative research method would be used for the study, the results of participants would be more difficult to compare due to individual differences. An interview would generalise the opinions of the participants. With an observation the perceptions of users cannot be measured.

User experience can be diverse and it is influenced by different factors (Karapanos 2010). Karapanos has emphasised four different sources of diversity in user experience: individual differences, product features, situation's characteristics, and time aspect (Figure 1).

Figure 1. Diversity in UX. Source based on Karapanos 2010.

Considering the diversity, the author saw the need to put individual opinions and feelings to numeric and comparable format. The goal was to understand what kind of design solution provides better user experience in a measurable scale in order to compare different solutions; moreover, in which areas and how much one design is better than the other. Research has proven to find (Diefenbach et al 2014) that it is very relevant to put the perceptions of usage of a system to a quantitative scale. Therefore, a **quantitative research method** was chosen for the study.

The author proposed an **experimental research design** for the study, because it allows to test the relation between dependent and independent variables. Experimental research design has been widely used in user experience research (e.g. Hassenzahl 2013; De Angeli et al 2006).

Individual differences always are in place for all people, and to get a comparison of these differences, more participants were involved in the study than just one. The product the study participants use, should be the same in each episode, and it should not depend on the individual. Situational differences should be tried to be avoided by using very similar settings with all study participants. Time aspect could be considered also invariable, as the product should not be new for the user, and he or she did should not have time to get used to it yet.

Consequently, using very similar situation and time factors in all studies (independent variables) and different design solutions as experimental conditions (independent

variable for each product episode) and multiple individuals (dependent variables) helped to understand the diversity in user experience (Karapanos 2010). By changing the product, the research should be able compare different products (e.g. De Angeli et al 2006; Vugt et al 2006), such as different interactions and different devices. Consequently, **ABCD testing** was chosen, with the four versions being two designs on two devices.

3.2. Stimuli

The author has experience in building information systems, both in strategic and tactical scale. Most of them have been related to the web. The author has observed user behaviour through statistical tools and design sessions. The feelings the users have after using some web-based system, always have seemed to be fluctuating and severely individual. Nevertheless, there also have been observations of specific logics when the user experience is better and worse. Therefore, the author has decided to find out some of the logics that influence user experience.

The world of web interactions is wide, as described in the introduction of the thesis. There are different platforms and input devices. Consequently, it was necessary to set a focus for the research. The author chose for the study web interactions that happen within one web page by concluding an assignment. Web-based interaction was chosen as the platform for the stimuli for one reason: a web interaction can be accessed through all of the devices that have a web browser. The variety of devices that exist, mostly support web browsing and the main standards. Therefore, the web-based interactions can be accessed using different devices in a way that the interaction itself looks similar.

As described in the literature review, the variety of devices that are used for interactions has increased tremendously. In 2016, the main devices used for web-based interactions are mobile smartphones, tablets and desktop computers (including laptops). Tablets in this case are considered as devices that are mainly interacted with via touchscreen, although there are also desktop computers and laptops with touchscreens.

According to eMarketer research conducted in December 2014, the number of tablet computers have surpassed the 1 billion line (eMarketer Inc. 2015). Nevertheless, the number of smartphones was nearly 2 billion at the end of 2015 (Smart Insights Ltd. 2016).

The goal of the research was to also compare the influence of web interactions between different devices. Considering the penetration of different devices, a **laptop** and a **smartphone** were chosen as the devices to complete the study with. Such devices were also used mostly by people who were in the interest area of the author: those who work in an office and use computers and smartphones on a daily basis.

By concluding the same assignment with different designs and different devices, the comparison between interactions can be studied.

3.3. Participants

Taking into account the author's interest area in the research, the population was limited to Estonian people between ages 20-65 who live in bigger cities, and have got an office job where a computer is the main working device. Moreover, the population often uses web browsers, both on a desktop and mobile device.

The number of people who use internet for search, e-mail, forums and communication, was in 2013 around 546 000 among employed people (Statistikaamet: IT37, 2014). It has to be considered that this number includes also the people who do not work with computer, but use it in private life.

There is also available information regarding job profile among employed population. If to sum up lawmakers, higher officials and managers, top specialist, mid-level specialists and technicians, officials, and service and sales personnel, there are 317 000 of them in Estonia (Statistikaamet, 2011). The limitation of this number lies in the fact that not all of the job profiles use computer on a daily basis, and some job profiles who use computer on a daily basis, are excluded from the list.

According to statistics from 2011, the number of people who used computer at work was 326 400 (Statistikaamet: IT33, 2014). This number contains all the people who at that time used computer at work, but it does not define the frequency nor the level of experience.

Most of the relevant numbers from Estonian Statistics Bureau are outdated, as computer usage and experience has grown probably a lot since 2011, and new devices such as smartphones are not covered by those researches at all.

There is no precise data available about the population, still, it is possible to refer to some sources that give information about the potential population. Based on the figures in the statistics, the author assumes that the total number of people in the limited population of the research is around 350 000 to 400 000 today.

Finding a fully representative sample from the population would be very costly and time-consuming for this stage. Therefore, the author decided to run a **pilot study** within the scope of this thesis.

The goal for this study was to find at least 12 participants, 6 males and 6 females. In case the results after 12 studies were very similar for all participants, no more participants were needed to be invited to the study. Otherwise, few more participants could be invited to the study.

For the pilot study, the following criteria were set for the participants:

- The participant has to be between 20 and 65 years old.
- The participant has to live in a bigger city in Estonia: Tallinn, Tartu or their neighbouring suburbs.
- The participant has to be currently employed. In case the participant is a student, he or she must also work at least half-time (20 hours per week).
- The participant must have a work in an office where the main working tool is a computer (desktop, laptop or tablet).

- The participant must have a good experience in using a web browser in a desktop computer or a laptop.
- The participant must have a smartphone and/or some experience in using its web browser.

In order to enhance the quality of the sample, and to make the sample more representative, the following measures were taken into account:

- Not more than 2 people from the same organisation should be invited to the study.
- Approximately 50% of the study participants should be male and the other 50% female.
- Each of the age groups (20 to 29, 30 to 39, 40 to 49, 50 to 59, and 60 to 65) need to have sufficient number of participants, so that there would not be too many people from one age group compared to other groups.
- Both people who regularly use a Windows and a OS X based computer should be involved.

3.4. Data Collection

During analysis of quantitative research methods, several studies were analysed in the literature review. In this chapter the author reasons on the selection of data collection method.

SUS questionnaire by Brooke (1996) was not considered as an option for the quantitative research, as it focuses too much on the usability and measures too little qualities of the product. The Post Study System Usability Questionnaire (Lewis 2002) measures system usefulness, information and interface quality, but it mostly focuses on pragmatic qualities and not so much on the hedonics. UEQ (Laugwitz et al 2008) and AttrakDiff questionnaire (Hassenzahl et al 2000) were considered relevant taking

into account the research questions and the stimuli, as they measure both pragmatic and non-pragmatic qualities with non-pragmatic ones having significant focus.

AttrakDiff has proven several times (e.g. Diefenbach et al 2014; Stojmenova et al 2013; Schrepp et al 2006) to be very useful for measuring user experiences in non-physical products. The author considers the AttrakDiff to be slightly more design oriented, considering different qualitied of a product, i.e. offering suggestions to the designer of the web interaction.

AttrakDiff questionnaire contains of 28 word pairs, each of them in a seven-point scale. The middle value is 0, left-most value as -3 and right-most value as +3. The word pairs are in four groups:

- AttrakDiff answers 1 to 7 are in group PQ (pragmatic qualities);
- AttrakDiff answers 8 to 14 are in group HQ-I (hedonic qualities - identification);
- AttrakDiff answers 15 to 21 are in group HQ-S (hedonic qualities - stimulation);
- AttrakDiff answers 22 to 28 are in group ATT (attractiveness).

All word pairs and groups and shown in detail in Appendix 2. AttrakDiff Questionnaire: AttrakDiff Groups and Word Pairs in Estonian and English. Considering the questionnaire and ABCD testing, each of the designs (interaction episodes) should be assessed using the questionnaire separately four times.

The participants should be asked to assess the word pairs instantly, without thinking too much on the meanings behind the words. Mostly the words are on a negative-positive scale, e.g. “bad – good”, “complicated – simple”, or “rejecting – inviting”. Yet there are some word pairs that might not be directly negative or positive. For example, the goodness of “human – technical” can severely depend on whether the assessed object is for task-oriented technical people or something that should be desirable for the user rather than practical. For word pairs “alienating – integrating”

and “brings me closed – “separates me”, an interaction can be assessed to be “alienating” or “separating”, because the context of the interaction is to separate something. When testing interactions of an online store that sells only cheap items, the word pair “cheap – premium” might get a “cheap” evaluation not because the interaction itself looks cheap, but because the products displayed are cheap.

Consequently, the limitation of the questionnaire is that the right-hand side of the evaluation scale does not always mean the positive side, therefore, it cannot be always said that an interaction with a higher average evaluation is better.

3.5. Conclusion

The main pillar for conducting the study was a quantitative research that involved an experimental design and ABCD testing. This involved altogether four design solutions of the same task that were experimentally designed: two different designs were tested on two different devices, which were a laptop and a smartphone. All interactions were also compared between each other.

Participants involved office workers who use computers and smartphones on a daily basis. The participants were limited to Estonian bigger cities and to ages 20-65.

For data collection, AttrakDiff questionnaire was used. It has 28 questions and allows to measure the interactions based on different perceived qualities: pragmatic, hedonic qualities and attractiveness.

The combination of the aforementioned research methods, participant selection criteria and data collection methods allowed to create a study that could answer the research questions and test the hypothesis.

4. STUDY

The study was based on the research design of this thesis. This chapter explains the selection of a stimuli, including specific interaction episodes and devices. It also describes how study participants were recruited and how their eligibility for the selection criteria was assessed.

The data was collected by using AttrakDiff questionnaire. It is pointed out how the data collection was prepared by developing a technical environment for the study.

The procedure clarifies in detail how the study was carried out.

4.1. Stimuli

4.1.1. Selection of Interactions

For the stimuli, the goal was to assess the interaction related to the research topic. In order to find the relevant interactions, firstly, the author studied all web interactions that are most used by web creators.

From all the interactions, the author chose five, which are used for a large variety of assignments, or which can be found on many web pages nowadays. The reason for choosing these five interactions is based on the author's personal and professional experience regarding what activities users do online and where people have problematic experience. Even if the user achieves the aimed result through the interaction, he or she might do it not in the best or most efficient way, mostly because he or she does not know the way how the designer expected it to work. Examples of the reasoning of the interaction areas are added to each of the interaction.

The following interactions were chosen for the study:

1. **using a search (later used as “search”);**

   Searching based on a keyword or multiple keywords from the desired content.

   Example: Used for searching the web through search engines, but also for searching inside web pages based on keywords.

2. **selecting one option out of many options (later used as “option selecting”);**

   The user has to select one out of many options, or many up to all options out of many options.

   Example: Making a selection of anything on the web, either through dropdowns or other similar type of one or multiple selecting boxes. Selecting product category to search the keyword in, or for choosing the type of accommodation in a booking site.

3. **grouping (later used as “grouping”);**

   Adding desired similar items into one group.

   Example: Used for adding images to galleries, adding friends into groups for group chat, inviting multiple people to do something.

4. **defining a range of sums (later used as “sum range”);**

   Defining the minimum and maximum sum of the product.

   Example: Mostly used on online shopping for finding products within desired price range.

5. **defining a date range (later used as “date range”).**

   Selecting a start and end date.

   Example: Mostly used in booking sites, such as accommodation reservation or ticket purchasing.

The target was to pick one of the interactions with which to continue with the study. The author wanted to choose the interaction which creates the most problems in using, compared to others. In order to find verification to which of them most problematic is, the author concluded a preliminary survey.

4.1.2. Preliminary Survey for Choosing the Interaction

The preliminary survey for choosing the interaction was very simplistic. It was web-based survey done with Google Forms. It contained still pictures of five web interactions together with the assignment description. No interaction at this point was involved, therefore the actual complexity of completing the assignment was not taken into account. Under the picture there was a matrix questionnaire where the rows were the five interactions depicted above and the options were from one to five radio buttons to assess how problematic the interactions could be. "5" was described as "most problematic" and "1" was "least problematic". The reasoning behind the word "problematic" was subjective and decided by the respondent; it was not explained to the user what kind of problems there could occur using such interactions, therefore the results were based on the opinions and experience of the participants. The example of the questionnaire can be seen in Appendix 1. Screenshots of Preliminary Study's Questionnaire.

The pre-study survey was distributed as a convenience sample, by sharing it to friends and acquaintances via e-mail, Facebook post and Facebook message. Also the author asked to share the survey link to others. The survey was active only for 48 hours, on March 4th until March 6th. Altogether 38 people responded to the survey.

The analysis of the results was based on averages and counting of answers. The chosen interaction had to have the highest average score on the survey. Additionally, the author checked the count of each interaction's most problematic appearances.

	Average score	Number of participants who chose score "5" (most problematic)	Number of participants who chose score "4" or "5" (most problematic)
Search	1.553	0	3
Option selection	1.632	1	1
Grouping	2.842	4	13
Sum range	2.079	0	4
Date range	2.605	3	10

Table 1. Results of the preliminary survey for choosing the most problematic interaction.

Interactions of grouping and selecting date range received the highest score, 2.842 and 2.605, respectively. The author counted also the number of participants who selected "5" (the most problematic) for the score. Grouping interaction received 4 and date range 3 occurrences of the score. The number of occurrences of both "4" and "5" as the scores was also higher for grouping interaction, having 13 people assessing it rather problematic.

Based on the results of the survey, the author decided to focus on **grouping interaction** in the subsequent study.

4.1.3. Selection of Stimuli for the Study

There are tens of ways how grouping can be done in web. For example, it can happen by clicking buttons, dragging the elements or pressing keys on the keyboard. Groups can be lists, paragraphs or text, form elements, images etc. The interaction of grouping may be animated or non-animated. When dragging elements, they can appear on top of each other as layers or be forced to a certain grid or to a set of alignment rules. The author came up with a more specific and real-life example in order to choose the characteristics of the interaction that would be studied further.

The assignment was to divide different products – groceries and electronics – into two groups. The initial group would be containing all products and then there would be two additional, initially empty groups with titles "groceries" and "electronics". The

author studied different web-based grouping options and decided to use two of them for the study.

One of the grouping solutions was technically more simple. The elements were in a multi-select-box. There were three of such boxes. The first box contained the initial elements that would have to be grouped, and the second and third box were the empty ones to where to group the elements. It was basically a simple HTML-based form element. Grouping was done using buttons near the 2nd and the 3rd group box. Near each of those there were two buttons – one for adding the element and one for removing the element. Such grouping solution has been used since early times of web pages, and is still used in different variations. This solution is later referred as **"button to group"**.

The second grouping solution was a drag and drop. Each element was a small box that could be dragged to a position inside the same group, or to another group. Again, there were three areas, the first containing the initial elements and the other boxes being empty. The only difference between this and "button to group" solution from the perspective of achieving the goal was that in drag and drop the user could also move elements from the 2nd to the 3rd group, and, vice versa, whereas the "button to group" only allowed to move elements from the 1st group to the 2nd or the 3rd and from there back to the 1st group. The overall assignment could be still accomplished in a similar manner. Dragging and dropping is more widely used in the past 10 years and is covered by the popular jQuery UI library, by combining two of its standard interactions: Draggable (jQuery UI: Draggable 2016) and Droppable (jQuery UI: Droppable 2016). By importing this library's codebase to the web page, creating such solution for the web developer is in a similar difficulty level as the "button to group" option. This solution is later referred as **"drag and drop"**.

4.1.4. Selection of Interaction Devices

In the research design a laptop and a smartphone were chosen for the study. Additionally, the study participants needed to have a device for making the evaluations.

Altogether three devices are used for the study: a laptop and a smartphone where the interactions were done, and a laptop where the questionnaires were filled. One laptop was a MacBook Pro with OSX operating system and the other one was a Samsung laptop with Windows operating system. The phone was iPhone 6. All interactions were done in a Chrome browser where all interactions worked in exact same way, even in the mobile device where the used layout was not changed based on the smaller screen.

4.1.5. Design of Interaction Episodes

As a result of the selection of devices and interactions there were four different episodes that were to be studied:

Episode 1: "button to group" episode in a desktop device (laptop);

Episode 2: "drag and drop" episode in a desktop device (laptop);

Episode 3: "button to group" episode in a mobile device (smartphone);

Episode 4: "drag and drop" episode in a mobile device (smartphone).

Each study participant was expected to conclude all four interaction episodes.

4.2. Participants

For the pilot study the author used his own network to find the study participants. As described in the previous chapters, it was decided to focus on a target group who has experience with both web and mobile usage, and who is an office worker in a bigger city in Estonia. Such target group selection was done because it helped to exclude the people who are not daily web and mobile interaction users or who do not cope well with web interactions in general. Therefore, people who did not have experience in grouping interactions of any kind did not influence the result of measuring the interaction's influence on user experience on already experienced users.

The participants were invited to the study through professional connections using phone calls, e-mails and Facebook Messenger for the initial contact. Some of the participants were acquainted with the author, but the researcher asked each of the participant to involve someone else from the organisation with whom the author had no previous connection with.

The fulfilment of participant criteria was constantly monitored by the researcher. For example, in case there were too many female participants, the researcher asked more men to join the study. In case there were already enough people from one age group, other age groups were approached.

4.3. Data Collection

After completing each interaction episode, the study participant was required to fill in the AttrakDiff questionnaire. It contained of 28 questions, each of them was assessed as a word pair in a seven-point scale. The middle value was set as 0, left-most value as -3 and right-most value as +3. All the values were stored in the database for later analysis.

The following sub-sections describe the environment for the data collection, data model and database, front-end and back-end applications.

4.3.1. Technical Environment for the Data Collection

In order to conduct the study, a technical environment needed to be set up. The author decided to create it on his own. The development had three components:

1. database;
2. front-end application;
3. back-end application.

The system was created in two languages: Estonian and English for enabling also non Estonian speakers to participate, in case such need would occur.

4.3.2. Data Model and Database

In order to store and use the data received from the study, a database needed to be created. Before creating a database, a data model was created. The data model also described the needs for the database.

The data model contained of four elements:

- **Episodes** – all 4 episodes were described here. Each episode had parameters such as the device (laptop or mobile) and type (button to group or drag and drop).

- **Pairs** – these were word pairs of the AttrakDiff questionnaire. Each element of the questionnaire had two words – one on the left and one on the right side of the scale. Also the sequence of the questions was defined.

- **Participants** – all data about study participants, including the participant code, gender, age. Also the start and end date and time of each participant's sessions.

- **Answers** – all data about the answers of participants. Each episode's answer set was saved as one row in the database. This contained the participant's ID, episode's ID, episode's conduction's start and end date and time, questionnaire's start and end date and time, and 28 values of each characteristic that was assessed by the participant.

After creating the data model, the database was created in MySQL. The database contained four tables: episodes, pairs, participants and answers, each of them containing the structure according to the data model.

The next step on creating the study conducting environment was development. Web applications that needed to interact also on the server level, needed to have both front-end and back-end applications. In the following sections the author describes them.

4.3.3. Front-End Application

Front-end application is the part of the application that is either visible or can be felt by the user. The content of it involves the logic how web browsers interpret the results of the code: layout, styling and scripting.

The front-end application built for the study mainly contained the views that were available in the browser when conducting the study.

For the layout part Bootstrap framework was used. It is considered to be the most popular HTML, CSS and JavaScript framework for developing responsive projects on the web (Bootstrap 2016). Bootstrap helps developers to do more with less additional code and gives efficiency in developing common, but difficult to program parts of the interactions.

For scripting of interactions, jQuery and jQuery UI libraries were used. These are sets of JavaScript code that defines a logic how HTML elements are used and interacted with.

The front-end application contained the following views:

- login and logout views;
- user registration view;
- four interaction views for all episodes;
- questionnaire view;
- information / waiting view.

The login view contained just the study participant code, which was given to the study participant by the researcher. The user registration view contained fields like participant's age, gender and questions about web browser and mobile browser usage.

View of each interaction was designed separately. Every one of them included the elements (products) and the logic how they were positioned on the screen. Also, the

interaction logic was built into these views, including how elements could be moved between blocks, which area accepted which elements and which area was not allowed to accept certain elements. The interaction logic was built with jQuery UI's interactions Draggable (jQuery UI: Draggable 2016) and Droppable (jQuery UI: Droppable 2016). The example view of "button to click" interaction is shown on Figure 2 and "drag and drop" interaction on Figure 3.

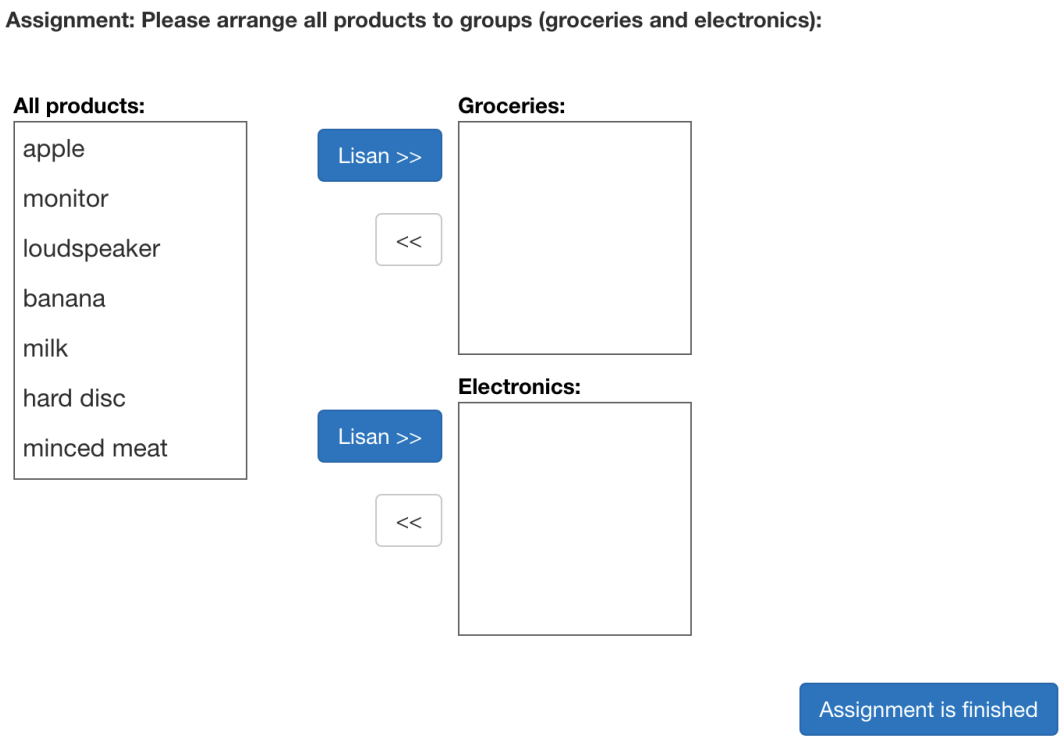


Figure 2. Assignment's "button to group" interaction episode in the study conduction system.

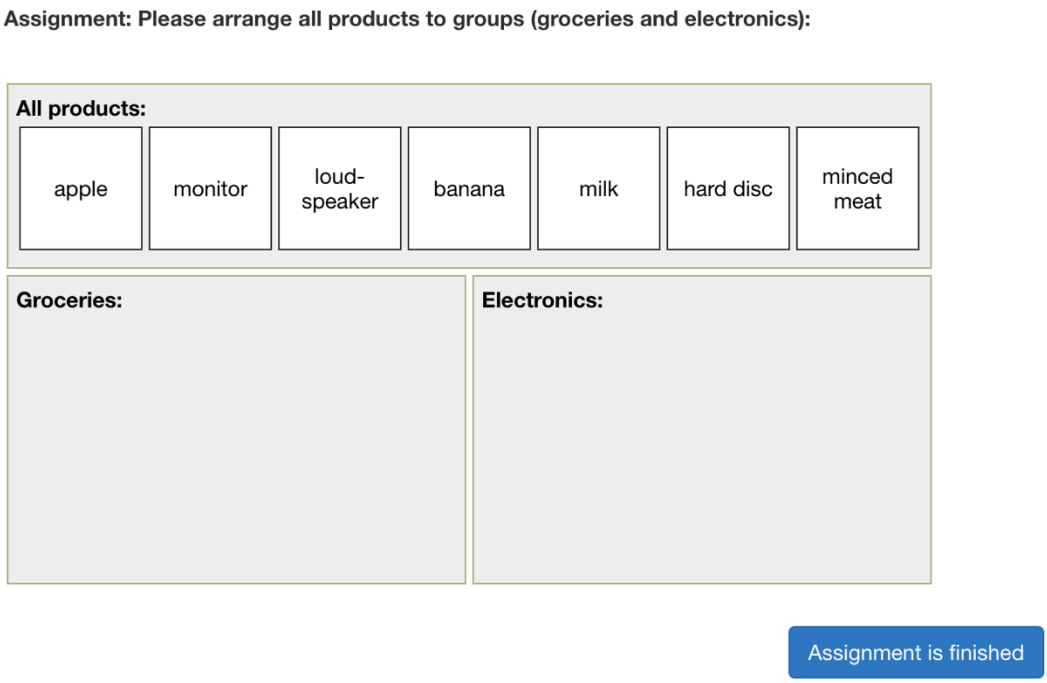


Figure 3. Assignment's "drag and drop" interaction episode in the study conduction system.

Other views are visually demonstrated in Appendix 3. Screenshots of the Study Conduction System.

4.3.4. Back-End Application

While the front-end application instructs the web browsers what to do, what and how to show, then the back-end application tells the server what and how to do. For the study application, the responsibility of the back-end application mostly lied on queries from and to the database and user flow handling throughout the study process.

The back-end application was built using PHP in combination with MySQL database. The application defined the rules when and how the content was displayed from the front-end application.

The back-end application contained the following components:

- database connection;
- session handling;

- login/logout functionality;
- user registration functionality;
- device detection and switching functionality (with help from the front-end application);
- episode finder;
- questionnaire functionality;
- controllers such as was previous episode finished, was questionnaire filled etc.

4.4. Procedure

The procedure defines how the database, front-end and back-end applications work in collaboration to enable the desired functionality to work and enable the study to be conducted. It describes the setting and describes the study sessions in detail.

4.4.1. Setting

Every study session was agreed separately with each participant. The session always took place during work hours, between 11:00 and 15:00, never at the beginning and never at the end of the workday.

Each study session was held in the office where the participant works. The researcher organised together with the participant a meeting room. The requirements for the meeting room included internet connection availability and that it would be separated from other rooms. Also, the room had to be quiet with no significant external noise coming through the walls. Additionally, the room chosen should not have any other disturbing factors such as playing TV or radio.

All participants came to the study from their daily work duties from the same building and without knowing exactly what was going to be done with them.

With such setting rules the researcher tried to eliminate the potential effects of the environment towards the result. Nevertheless, the author admits that there is no setting with perfect conditions. For example, the overall mood or previous conversations the participant had during the day, may have an effect on the user experience. In case the study would have been conducted in a laboratory environment and always in the same place, some other factors could have played a role in the user experience, such as transportation issues (e.g. finding a parking spot or using public transport, or distance from the initial location to the laboratory) or pleasantness of the laboratory environment.

4.4.2. Script

When the study participant arrived to the meeting room, the researcher started by introducing the study. It was said that the study is about assessing web interactions and the participant would conclude four interactions on two devices: a laptop and a smartphone. All four interactions would be assessed using a questionnaire, which would be filled separately. Also, the researcher told that the sessions takes about 15-20 minutes and that the participant should not analyse too deeply the meanings of the words in the questionnaire, but rather should answer based on their first feeling.

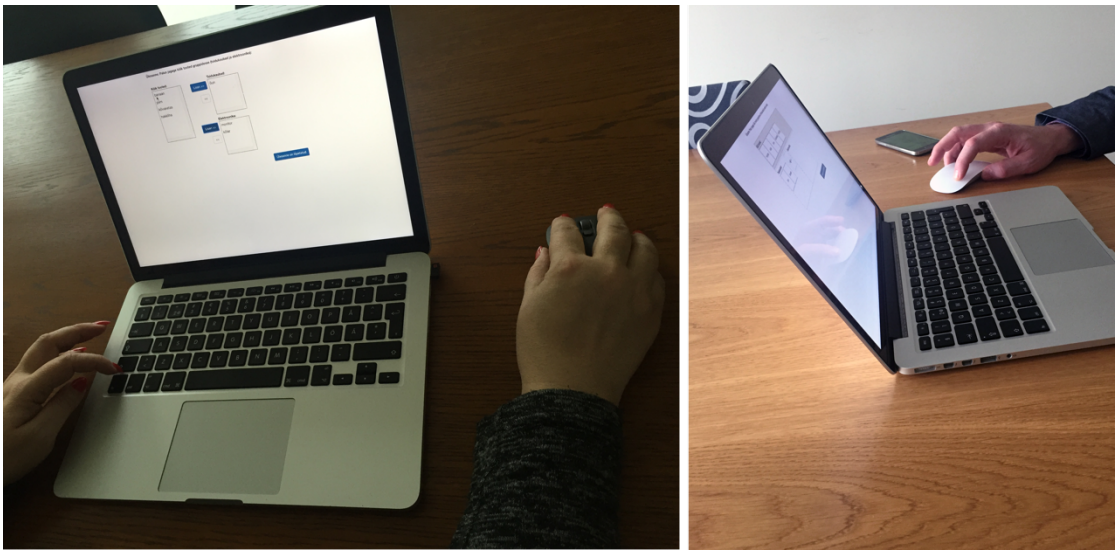


Figure 4. Examples of participants concluding the study.

Firstly, the researcher asked the study participant which device the participant uses on a daily basis. In case the participant used a Mac, the MacBook laptop was given to the participant for completing the interactions. In case the participant used a Windows-based computer, the Samsung laptop was given for the interactions. Figure 4 shows two examples of participants conducting the study.

Random study participant code was given by the researcher who filled the code in the questionnaire filling laptop. After first time login, the user-based questionnaire had to be filled in.

As a next step, the participant was given randomly a device, either the laptop or the smartphone as the first interaction device. This was needed to reduce the overall differences users could experience from starting with one or another device, therefore, it was intended that 50% of participants start with a laptop and 50% with a smartphone.

The researcher again entered the participant code into the interaction device. Then, the application randomly chose the interaction to be done. The randomness was expected to allow 50% of participants to start with "button to group" and 50% with "drag and drop" interaction. The reason for picking a random interaction was the same as for picking the random device.

The interaction contained an instruction which was the same for all four interactions: "Please divide all products to groups (groceries and electronics)". In case of "button to group" interaction, the user had to select the item from all products by clicking or tapping on it. There were two other groups with empty lists: one for groceries and one for electronics. Secondly, buttons "Add >>" and "<<" were shown next to both product groups. Clicking or tapping on "Add >>" added the product from the initial group to the group next to the button. Clicking or tapping "<<" removed the product from the group next to the button and added it back to the initial group.

At the bottom of the interaction there was a button "Assignment is completed". There was no validation on whether all products were grouped correctly; for example, whether "banana" was to be grouped under groceries or electronics. Nevertheless, the

only validation that was used was for checking whether all items were moved to any of the two groups or not. In case they were not and the user clicked on the "Assignment is completed" button, an error appeared: "To complete the assignment, please divide all products to groups." In case all elements were grouped and "Assignment is completed", the completion fact was saved to the database and a notification view appeared asking to fill the questionnaire in the other computer.

The "drag and drop" interaction was otherwise exactly the same as the "button to group" interaction, with only small differences. Firstly, the elements were not presented as list items but as squares. Secondly, there were no buttons for adding or removing the elements, as the elements could be dragged and dropped using the mouse, or in case of mobile device, dragging them with a finger.

After finishing an interaction episode, the other laptop was given to the participant. There the participant filled in the questionnaire about the feedback for the episode.

Assessing the episodes was conducted through the AttrakDiff questionnaire that had four groups: PQ (pragmatic qualities), HQ-I (hedonic qualities - identity), HQ-S (hedonic qualities - stimulation) and ATT (attractiveness). Each group contained seven questions, altogether 28 word pairs in the questionnaire. As a result, each participant had to assess 28 word pairs for each four episodes, altogether 112 assessments. The questionnaire with groups and word pairs in Estonian and English is added to the Appendix 2. AttrakDiff Questionnaire

All 28 word pairs were mandatory to assess. The questions were always in the same sequence and for better readability grouped as five question blocks, which directly did not have any meaning for the participant at that time.

Questionnaire was filled after each episode. If all four interaction episodes were finished and questionnaires filled, the study session was finished for that user.

5. RESULTS

The study was conducted between March 7th and March 31st, 2016. Altogether 17 participants were involved in the study. This section describes how data was collected and how its quality was assessed. Information about the study participants and their study sessions is described. Additionally, the observations made during the study are presented.

5.1. Data Collection and Quality

The research was designed so that all the data to be received was only high quality data. The validation rules in the environment guaranteed that no empty values would appear on the results in the database. Nevertheless, it was validated later by the researcher that all required data was present.

Data about the age, gender and browser usage of each 17 participants was collected and stored. Additionally, start and end time was recorded in the database with one second precision.

For all the AttrakDiff questionnaires, 28 values for four interaction episodes were stored for each participant. This counts altogether 1904 evaluations of word pairs with a value between -3 and +3 with one-point precision stored in the database.

5.2. Participants

Table 2 describes all study participants, including their gender, age, browser usage frequency and time spent on the study. The browser usage was based on the options given to the user. Time spent on the study was the difference between the end of the

last AttrakDiff questionnaire and the start of the user questionnaire. Additionally, it took on average three minutes to make the introduction and thank the participant.

Partici-pant code	Gender	Age	Desktop browser usage	Mobile browser usage	Time spent on the study
S129	Male	47	More than once per day	More than once per day	11 minutes
S130	Female	27	More than once per day	More than once per day	12 minutes
S131	Female	60	More than once per day	At least once per week	17 minutes
S132	Female	20	More than once per day	More than once per day	15 minutes
S133	Male	58	More than once per day	More than once per week	23 minutes
S134	Female	54	More than once per day	At least once per day	15 minutes
S135	Male	53	More than once per day	At least once per day	16 minutes
S136	Female	56	More than once per day	At least once per day	16 minutes
S137	Male	61	More than once per day	More than once per day	18 minutes
S138	Male	34	More than once per day	More than once per week	11 minutes
S139	Female	47	More than once per day	More than once per day	9 minutes
S140	Female	49	More than once per day	At least once per week	11 minutes
S141	Female	29	More than once per day	More than once per day	10 minutes
S142	Male	25	More than once per day	More than once per day	11 minutes
S143	Male	41	More than once per day	More than once per day	12 minutes
S144	Female	33	More than once per day	More than once per day	11 minutes
S145	Male	35	More than once per day	At least once per day	18 minutes

Table 2. Participants of the study: gender, age, browser usage frequency and time spent on the study.

There were eight male (47%) and nine female (53%) participants. The average age of the participants was 43 years. The youngest participant was 20 and the oldest 61 years old.

The average study lasted 14 minutes and 6 seconds, whereas the quickest study took only 9 minutes and 19 seconds, and slowest study took 22 minutes and 48 seconds.

It can be noticed from the table that few participants said that they use mobile web browsers less frequently that on a daily basis. The researcher orally asked explanations from those participants, as the requirement to become the participant was to be a daily computer and smartphone user. The explanation lies in the fact that those users still use their smartphones on a daily basis, but use mostly native applications, not so much mobile web browsers. As they use at least one of them at least once per day, these participants were treated the same way as those who marked to be using mobile web browsers on a daily basis.

5.3. Observations

Despite it was not planned to make observations, many study participants commented orally their experience. Therefore, the author took some additional notes regarding the comments. The notes are generalised and not connected to any specific participant.

Based on visual inspection, those who spent more time on the study, can be grouped into two: the ones who are just a bit slower computer users, and those who tended to analyse more thoroughly the word pairs while filling the questionnaire. There was no significant difference on the time spent on the interactions, rather the difference was only in filling questionnaires.

One general comment received by at least five participants after finishing the interactions was that the “drag and drop” was much easier to use. There were also at least three participants who felt more comfortable with a touchscreen: they considered anything done in the mobile device easier to use.

At the same time some of the comments were controversial. One participant said that he likes using a mouse and a keyboard more than a finger. The same participant also found the “button to group” as better solution.

At least four participants mentioned that they expected the interaction to let them select multiple elements at once, for example, by using a Control or Shift key, although the system was built so that only one item could be selected at a time. This appeared to be a wish in both “button to group” and “drag and drop” interactions.

After finishing a mobile interaction, one of the participants tried to use “drag and drop” feature with a finger on the screen of the MacBook, although the laptops did not have a touchscreen.

Overall, the observations are mentioned just to give a broader picture about how different users are and how much their expectations towards a UI differ. These observations were not taken into account in the quantitative analysis.

5.4. Conclusion

The results of the study were as expected. As data was validated immediately during the study, there were no missing values or other items that could not be used. The study participants were selected carefully to match the criteria. The observations made during the sessions did not affect the quantitative data that was used for the research. Consequently, all results from all 17 study sessions could be used for the analysis.

6. ANALYSIS AND DISCUSSION

In this chapter the author introduces the analysis of the research. To move closer to answering the research questions, firstly the sample is analysed. Each participant is analysed individually and examples of one participant will be brought out.

Group based analysis will be conducted based on AttrakDiff groups and word pairs.

6.1. Sample Analysis

The author saw significant differences in individual opinions based on the results of the study. To understand how distinct participants are, they were firstly studied as a group, without taking into account the episodes or the AttrakDiff groups.

All assessments of all study participants were analysed using multidimensional scaling, which helped to visualise the similarity of participants.

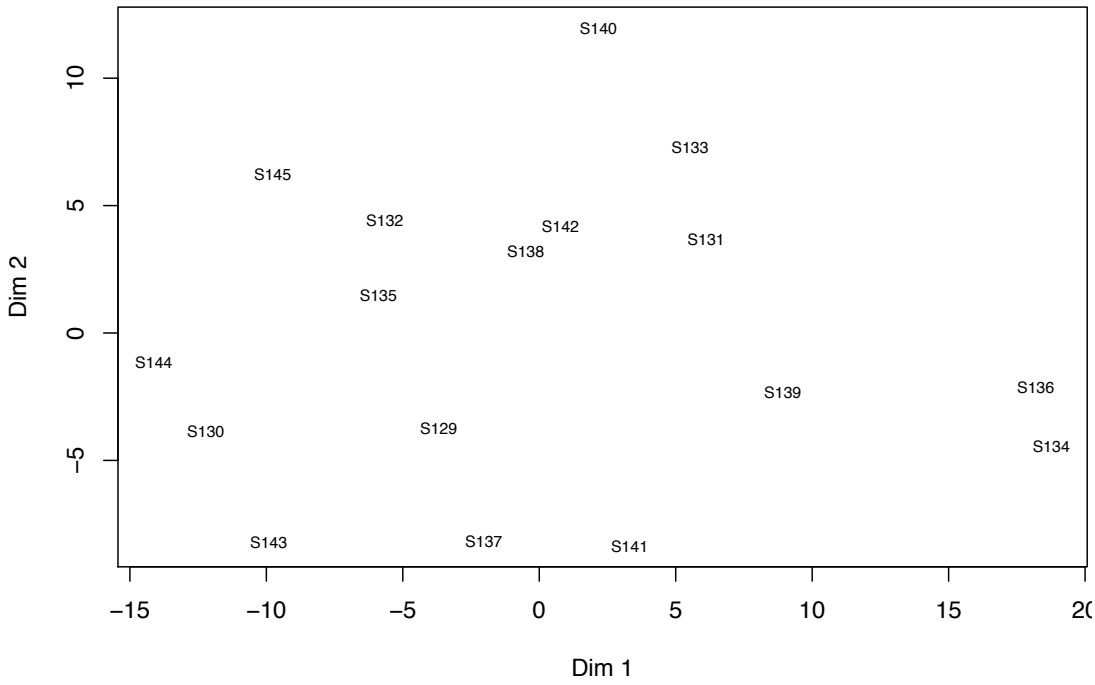


Figure 5. Users similarity assessment using multidimensional scaling.

The figure above firstly shows that the assessments given by all participants vary substantially. The most similar ones are S138 and S142, whereas S134 and S136 are the most different compared to other participants. In Figure 5, the participants on the right side of the scale had on average more positive values, whereas those on the left had more negative values. Those who are positioned on the top of the graph, had assessed interactions more homogeneously, and those on the bottom had more diverse assessments.

6.2. Individual Analysis

For the individual analysis, the author gathered all the individual data from the dataset, including 28 x 4 x 17 = 1904 assessments of word pairs (112 per participant) which was also the basis for the analytical data.

The examples in this section are based on the same study participant who has been chosen randomly.

Initially, a table with all answers for all episodes' answers was created. The table was converted into a visual representation of the word pair values selected by the participant, which helps to instantly understand the assessments of the participant. The chart was equipped with word pairs and the group names where the word pairs belonged to. Each episode is shown as a connected line with different colours. The following chart describes one participant's all individual answers among all interaction episodes:

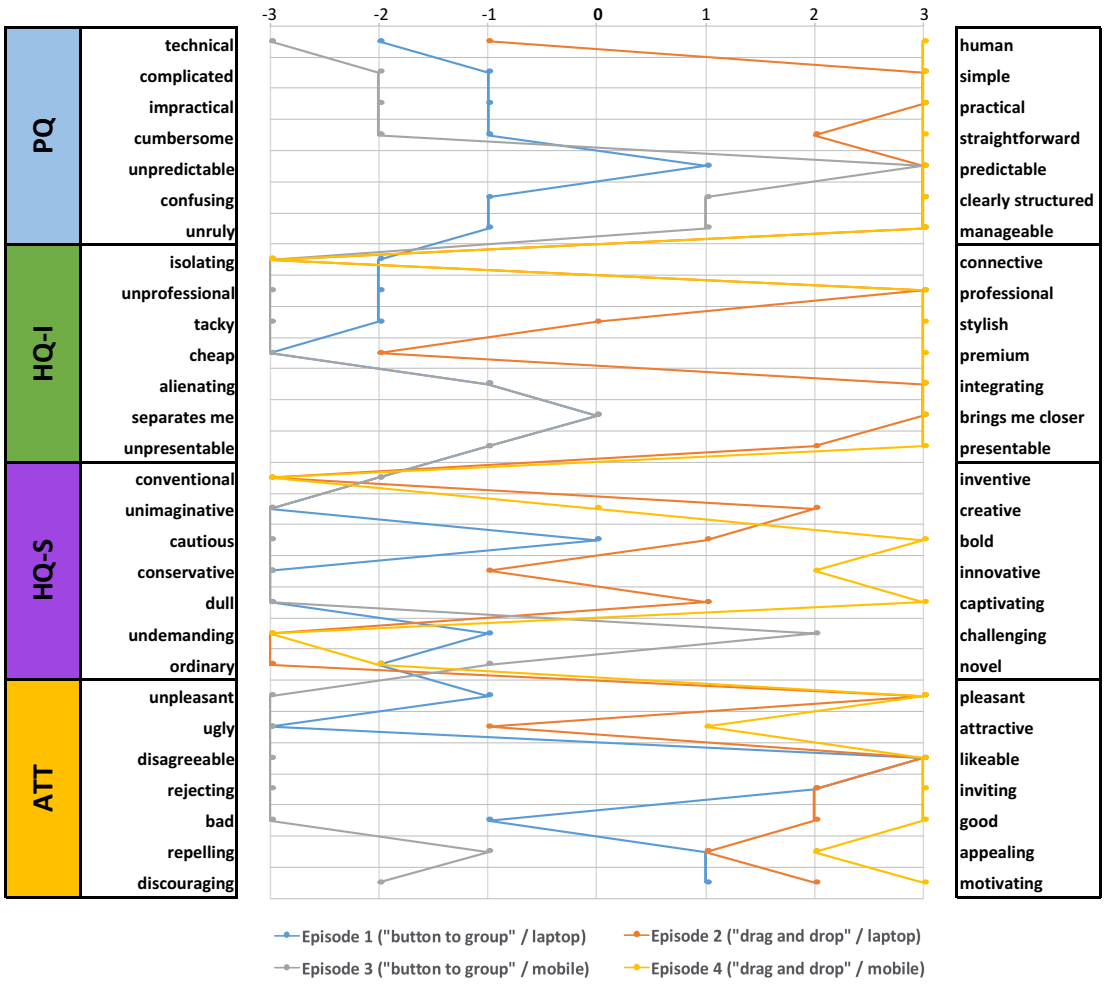


Figure 6. Example of one participant's answers to AttrakDiff questionnaire.

In Figure 6, all AttrakDiff groups and word pairs are in the vertical axis, whereas the assessment is in the horisontal axis. Different lines represent the interaction episodes concluded by the participant. The values in the table represent the chosen option in the word pair, -3 referring to the leftmost and 3 to the rightmost part of the scale.

It can be concluded from the results that for this participant episodes 2 and 4 had higher values than episodes 1 and 3. The most positive values were chosen for episode 4. For example, all assessments for word pairs from the pragmatic qualities received a maximum score. Such chart allows to look at individual word pairs. For example, if the researcher is searching for the most creative and innovative designs, only those word pairs could be assessed. In this study no specific word pair separately was as relevant as the averages of AttrakDiff groups, because the research was not looking for just good *versus* bad, or simple *versus* complicated solutions.

Nevertheless, such analysis allowed to detect any extreme values inside one group. For example, if looking at the individual hedonic qualities group, all values in case of this participant are with the maximum value (3), but one of the seven word pairs – isolating versus connective – has received the minimum value (-3). Such situation may have happened for several reasons. On one hand, although other qualities of the same group seem positive, the user really saw the interaction as isolating. On the other hand, while assessing simplistic web interactions, it can be difficult for the participant to understand the meaning behind the words, or they do not see any connection between the word pair and the interaction. In such cases sometimes a random answer is chosen or the answer is chosen without analysing it too deeply. However, this was also the task – to give assessment based on only the initial feeling.

Each of the participant's given values were also analysed by finding the minimum and maximum value, which gives an overview whether the user places too much on one or the other side of the scale. The average value of the participant's assessments was calculated for each episode. This makes the episodes more comparable in general. Additionally, standard deviation was calculated to understand how big were the differences inside one episode's answers.

	EP1 "button to group" in laptop	**EP2** "drag and drop" in laptop	**EP3** "button to group" in mobile	**EP4** "drag and drop" in mobile
Minimum value	-3	-3	-3	-3
Maximum value	3	3	3	3
Average value	-1,036	1	-1,75	1,929
Standard deviation	1,551	2,211	1,713	2,071

Table 3. Example of calculated values of one participant's answers for each episode.

To get a comparative overview of participants, for each participant the same values as described in the previous table were calculated among all episodes. This specific participant's example in Table 3 shows that the whole scale from -3 to +3 was used by the participant. Based on standard deviation, it also shows that the biggest differences among one episode's answers were in episode 2. On average, the 2nd episode received the highest average score in the word pair assessment.

To better understand the differences between answers inside episodes and among participants, the author created biplots that were generalising episodes and variables to be presented in a two-scale scatterplot. The biplots used principal component analysis and the input for the biplots was calculated by statistical software R.

The horizontal scale of the biplot represented the most-varying direction of the data. The direction of the vectors on the biplot represented the variables with similar answer profile: more similar answers made the vector to point in a more similar direction and the length of the vector shows the strength of the influence.

The first type of biplot created for every participant had each AttrakDiff answer as the first and each episode as the second set of variables. Altogether 28 x 4 = 112 values were included to each participant's biplot. This is an example of the same participant's answers represented as a two-scale biplot:

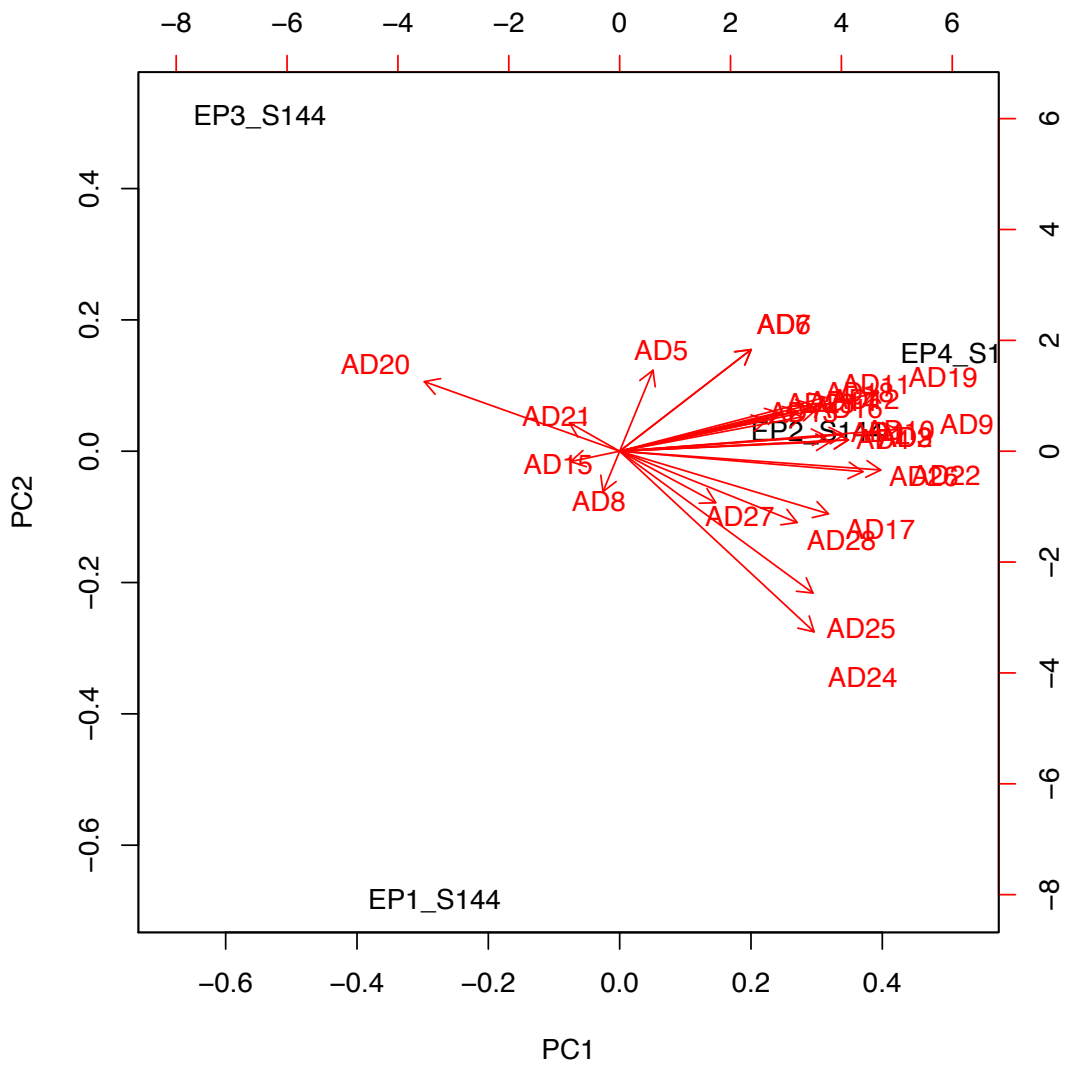


Figure 7. Example biplot of one participant's answers with with red arrows representing AttrakDiff word pairs and black marks representing the interaction episodes.

It can be derived from the Figure 7 that in this participant's case answers tended to differ among questions. Nevertheless, questions related to pragmatic qualities and attractiveness received more similar answers among all interaction episodes. Still, the differences among AttrakDiff groups are difficult to understand from the figure with all answers involved. Episodes 2 and 4 were more similar to each other based on the biplot, whereas episodes 1 and 3 differed among each other and compared to episodes 2 and 4.

Therefore, as a next step, each participant's answers were divided into AttrakDiff groups. Grouping provides more easily comparable data, but loses in precision. There

are four groups according to AttrakDiff: pragmatic qualities (PQ), hedonic qualities – identification (HQ-I), hedonic qualities – stimulation (HQ-S) and attractiveness (ATT).

The calculation for grouping was done via averaging the values of all seven answers inside one group. For example, the average of answers 22 to 28 represented the value for attractiveness.

For each participant the following chart was created representing AttrakDiff groups and their average values for each episode:

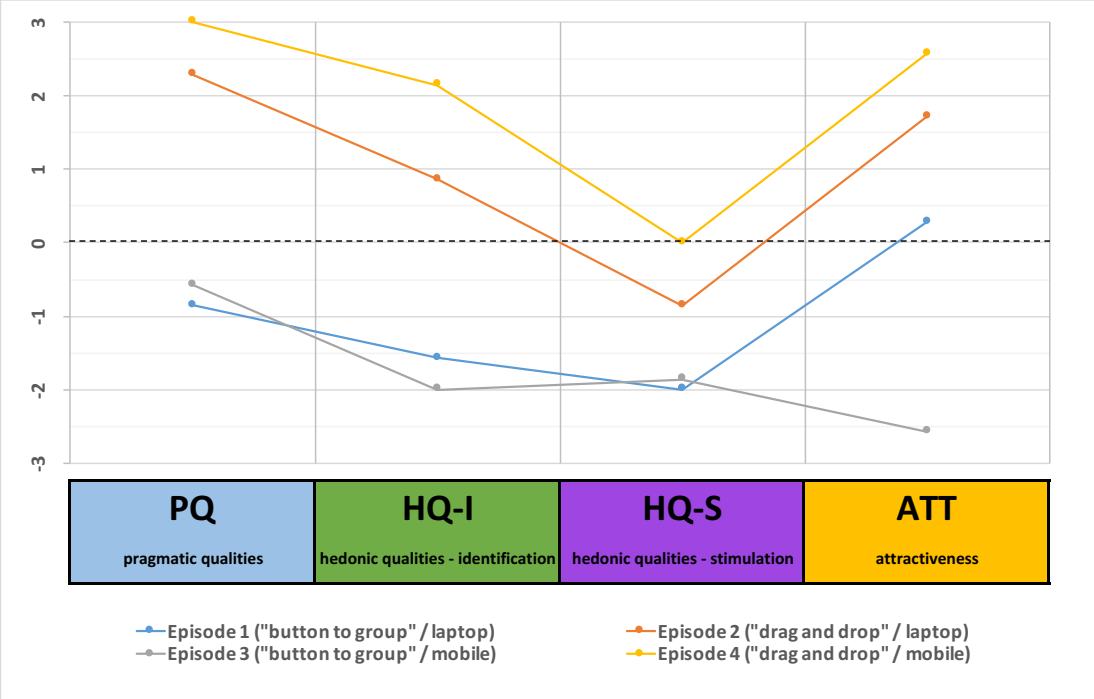


Figure 8. Example of AttrakDiff group values for all episodes of one participant's answers.

The vertical axis represents the average assessment values of the word pairs inside each group, and the horisontal axis shows the four word groups. From Figure 8 it can be seen that on average, episodes 2 and 4 were with higher values than episodes 1 and 3; episodes 2 and 4 were very similar, as the lines are almost parallel. In the first three AttrakDiff groups (PQ, HQ-I, HQ-S), also episodes 1 and 3 were similar, but the attractiveness assessment for the 1st episode was much higher than for the 3rd episode.

The lines of laptop interactions are mildly more similar to each other compared to the lines of mobile interactions.

For the next step, a biplot for grouped qualities was created. The average assessments of AttrakDiff groups were one and interaction episodes the other set of variables.

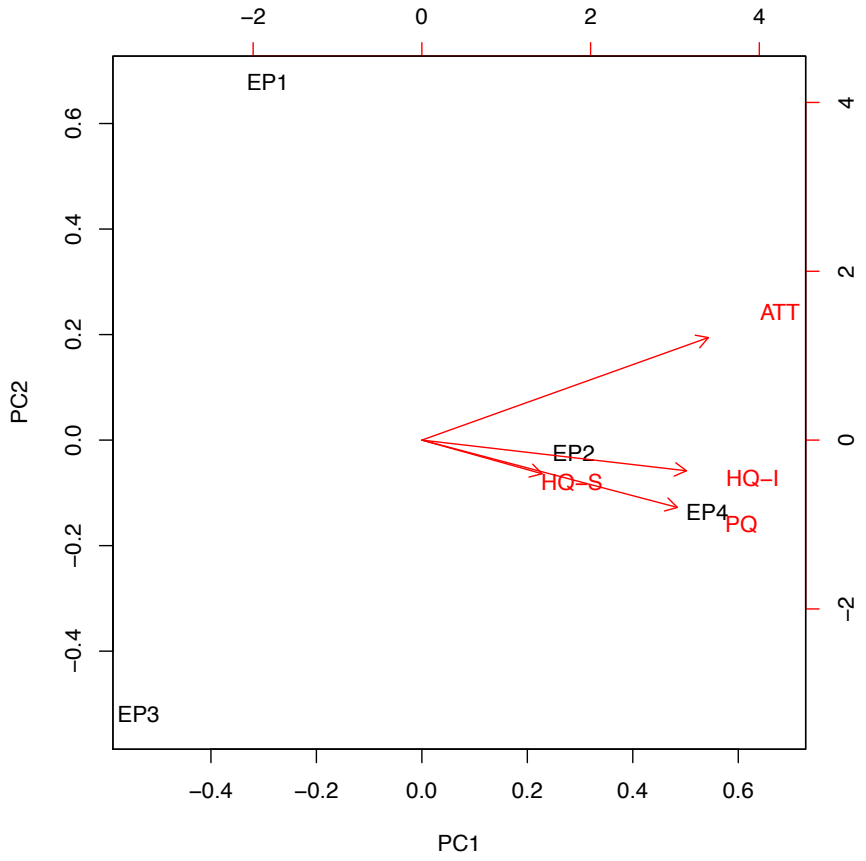


Figure 9. Example biplot of one participant's answers with red arrows representing AttrakDiff groups and black marks representing the interaction episodes.

The biplots, like the one shown on Figure 9, are available for all study participants in Appendix 4. Biplots of Individual Assessments.

In this participant's case, the biplot shows that the pragmatic and hedonic qualities had a similar answer set, whereas attractiveness was comparably different from other groups. Also, it confirms, similarly to the non-grouped biplot, that episodes 2 and 4 were similar to each other, whereas episodes 1 and 3 were different compared to each other and to episodes 2 and 4.

By comparing the results of all individual charts and biplots, the author concludes that all participants are relatively different. However, certain similarities appeared in this investigation. For example, **"drag and drop" episodes in laptop and smartphone tended to be more similar between each other** than other episodes. Additionally, **mobile interaction episodes emerged with higher similarity than laptop episodes**.

While individual analysis helped to understand in detail the differences between elements like participants, word pair assessments, AttrakDiff group assessments and episodes, a generalisation is needed (Karapanos 2010) to visualise and understand, how on average the elements variated.

6.3. Comparison of Pragmatic and Hedonic Qualities

The AttrakDiff model consists of three sets of qualities and attractiveness. The relation of the sets of qualities allows to understand where the users are located in the map of pragmatic and hedonic qualities.

For the analysis, the two hedonic qualities groups – individual and stimulation – were merged into one hedonics group. HQ-I and HQ-S were the averages of seven questions and the merged hedonic qualities group was created by averaging the results of 14 questions.

The comparison of pragmatic and hedonic qualities is important, because it helps to visualise the combination of those qualities. Consequently, each episode can be either neutral or with some tendency to extremes.

A chart with pragmatic qualities on the horizontal and hedonic qualities on the vertical axis was created for the analysis. Every episode was assessed separately. For each of them, the average word pair assessment values for pragmatic and hedonic qualities created a point on the map. Altogether there were four points where the qualities were placed on the chart. The location of the points also showed in which group the point belonged to. In case the point was in the centre of the map (both average values zeros), it would mean that the assessment of the episode is neutral. All other areas

represent some extremes. It is up to the designer and the product whether it is good or not to have the evaluations in the extreme zones.

To assess the confidence of the calculated points, additionally, confidence intervals were calculated (Table 4). Altogether eight datasets were used to calculate them: for each episode's hedonic and pragmatic values. The values were the assessments of the word pairs. Hedonic dataset contained of 238 and pragmatic of 119 values. The significance level α for the study was set to 0,05. A one-sample *t*-test was used to calculate the mean, *p* and confidence levels.

Episode	Qual-ities	t	df	p	Mean	Min confid-ence at 95%	Max confid-ence at 95%	Stand-ard devi-ation
Episode 1	**PQ**	5,768	118	6,564E-08	0,983	0,646	1,321	1,859
Episode 1	**HQ**	2,790	237	0,005704	-0,328	-0,559	-0,096	1,812
Episode 2	**PQ**	18,551	118	2,2E-16	2,017	1,802	2,232	1,186
Episode 2	**HQ**	10,969	237	2,2E-16	1,088	0,893	1,284	1,530
Episode 3	**PQ**	6,158	118	1,052E-08	1,092	0,741	1,444	1,935
Episode 3	**HQ**	-3,829	237	0,0001645	-0,433	-0,655	-0,210	1,743
Episode 4	**PQ**	27,788	118	2,2E-16	2,454	2,279	2,629	0,963
Episode 4	**HQ**	10,194	237	2,2E-16	1,008	0,814	1,203	1,526

Table 4. Values for *t*-test and confidence interval for PQ and HQ values of all episodes.

To illustrate the confidence levels of each combination of pragmatic and hedonic qualities, a rectangle was drawn around each point. The width of the rectangle was the difference between maximum and minimum confidence level of pragmatic quality of that episode, considering α=0,05. The height of the rectangle was the difference between maximum and minimum confidence level of hedonic quality of that episode, considering α=0,05.

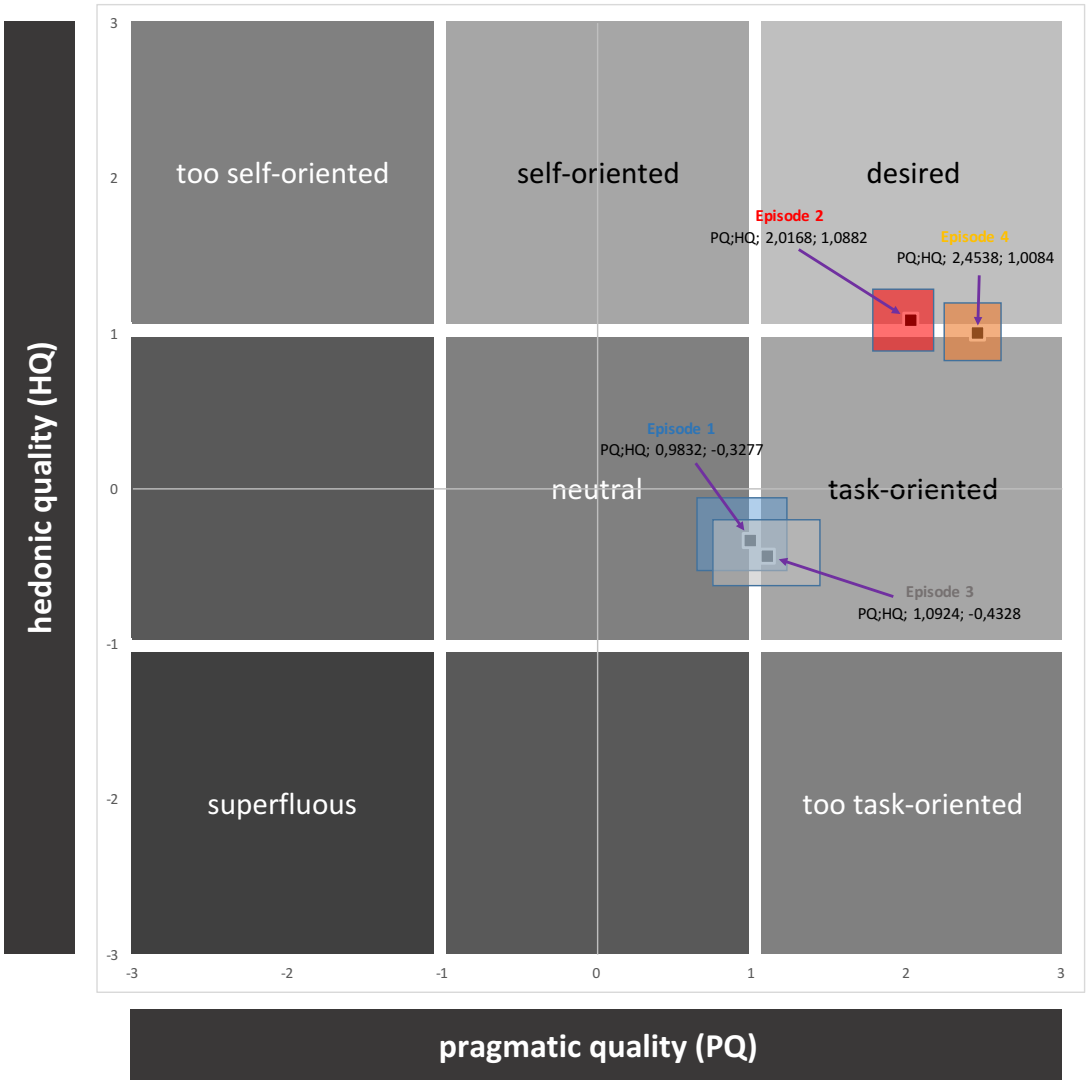


Figure 10. Interaction episodes' location on the map of pragmatic and hedonic qualities together with confidence rectangles.

The positioning of the mean values and confidence levels of episodes in Figure 10 helped to assess the interaction episodes. It can be seen from the map that "drag and drop" episodes in laptop (2) and mobile (4) had both highest values for pragmatic and hedonic qualities. These were the most desired episodes, being in between desired and task-oriented. The episode in laptop was slightly more desired, whereas in mobile more task-oriented.

The location of "button to group" episodes (1 and 3) was nearly the same, being in between neutral and task-oriented. The laptop episode (1) was considered slightly more neutral, whereas mobile episode (3) a bit more towards task-oriented.

6.4. Comparison of Word Pairs

For the next step, the assessment of word pairs was analysed. The average values for each word pair were calculated and are shown in graphical representation:

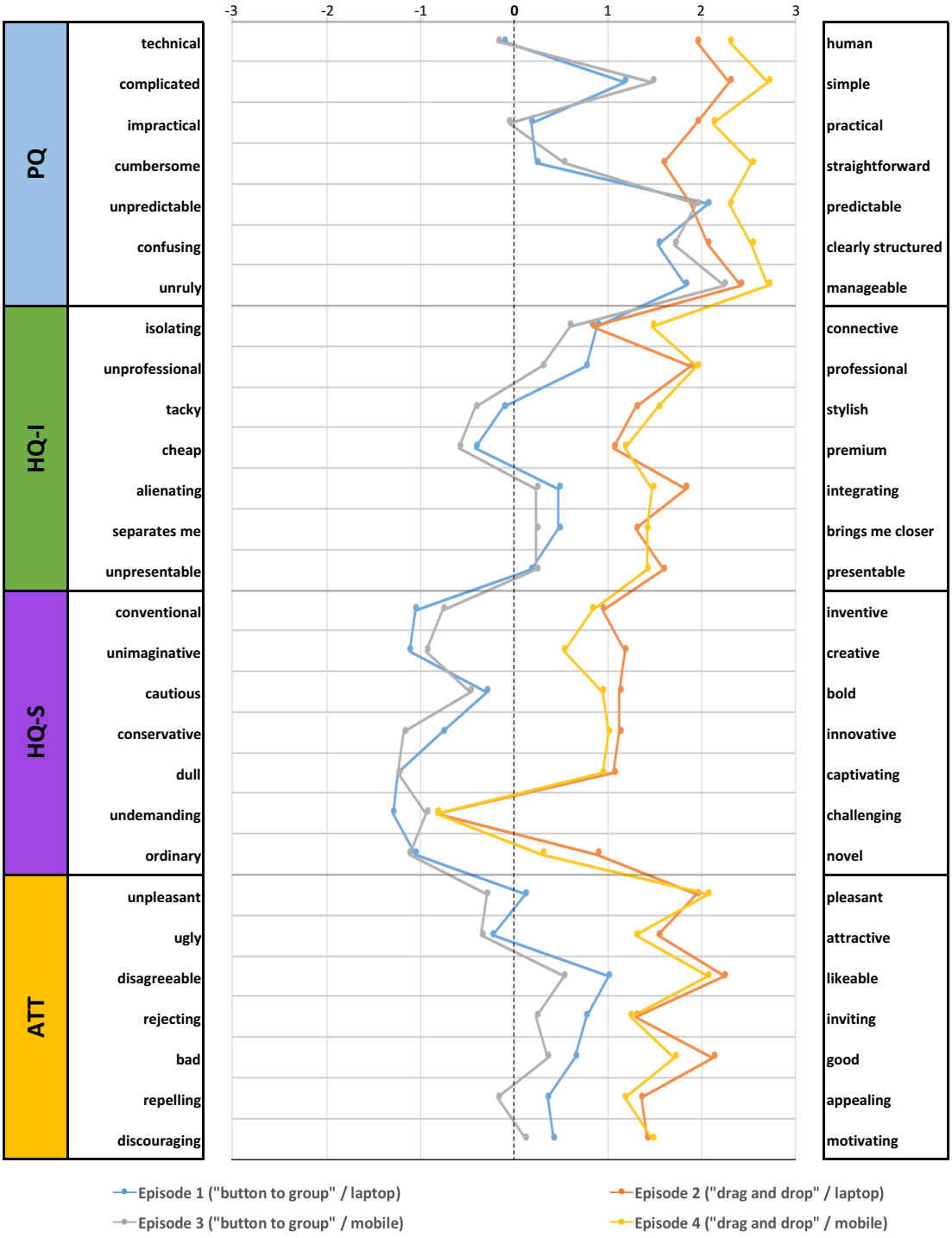


Figure 11. Comparison of assessments of word pairs among different episodes.

The points in Figure 11 show clearly which words had higher and which lower values. For example, the word pairs “unruly – manageable” and “complicated – simple” had the highest scores from all words in case of episode 4. On the contrary, the word pairs “undemanding – challenging” and “dull – captivating” had the lowest scores in case of episodes 1 and 3.

Also similarities between the word pair assessments of different episodes can be seen from the figure. For example, all episodes seemed to be similarly undemanding, but the practicality and pleasantness of episodes differed a lot.

6.5. Comparison of Groups

To compare the AttrakDiff groups, all answers of participants inside each group were averaged. This helps to overall assess the episodes based on all word pairs in the groups.

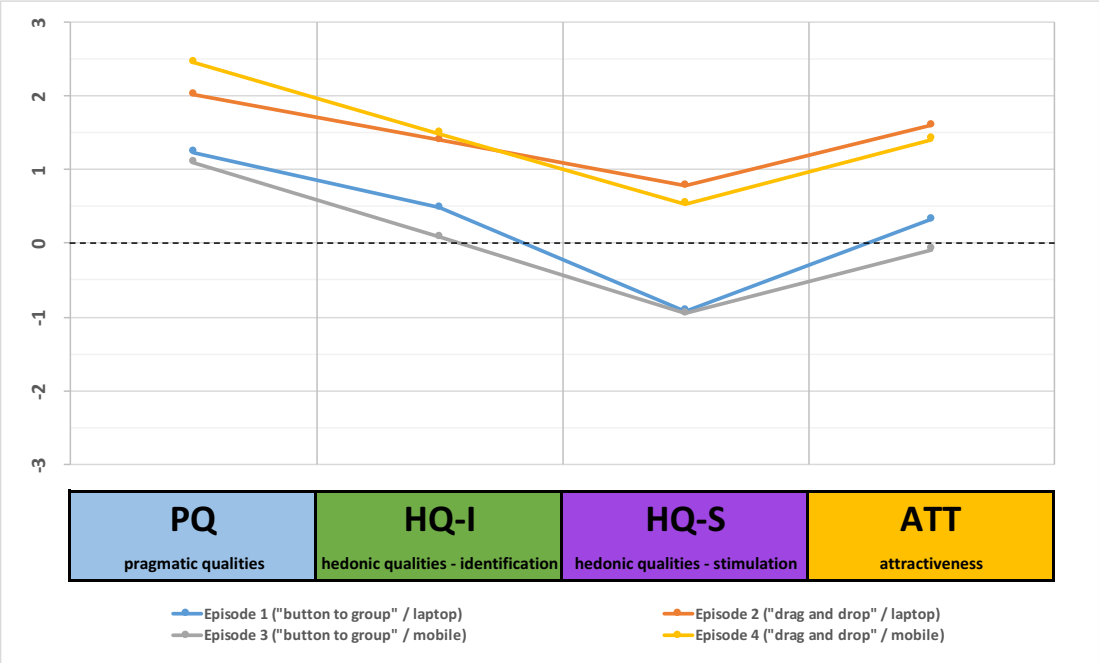


Figure 12. Comparison of assessments of AttrakDiff groups among different episodes.

It can be derived from Figure 12 that “drag and drop” episodes surpassed “button to group” ones in all AttrakDiff groups. The highest scores were present in the pragmatic qualities and the lowest ones in the stimulation hedonic qualities.

While the laptop episode of “button to group” had higher values than the mobile session in all groups, the “drag and drop” had differences among groups. In pragmatic and identification hedonic qualities the mobile interaction had slightly higher scores, whereas in stimulation hedonic qualities and attractiveness the laptop received better average values.

To assess the differences among episodes, multidimensional scaling was used. Each dataset was combined from all word pair assessment values from each episode.

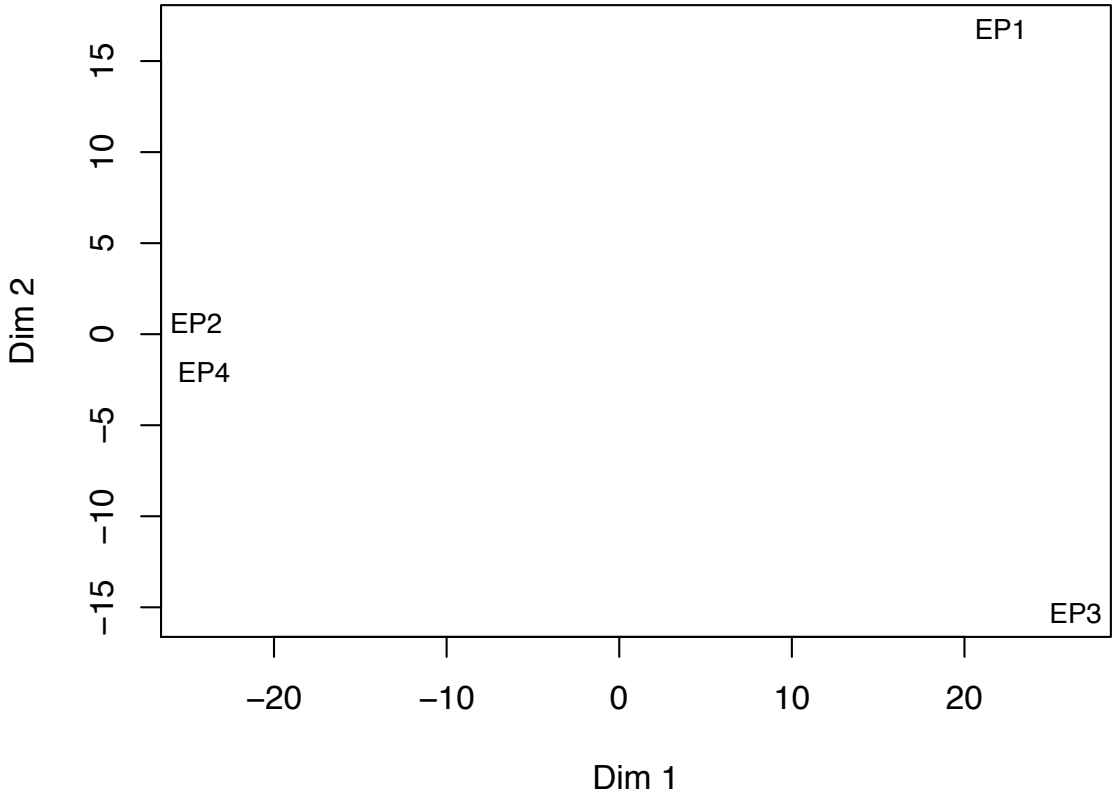


Figure 13. Differences between episodes based on multidimensional scaling.

The figure above shows that “drag and drop” episodes (2 and 4) were very similar to each other. On the contrary, “button to group” episodes (1 and 3) were relatively different from “drag and drop”. At the same time, episodes 1 and 3 were more similar between themselves compared to their similarity to episodes 2 and 4.

7. CONCLUSION

7.1. Answers to the Research Questions and Hypothesis Validation

The diversity in user experience has proven to exist also in this study. Most importantly, the research showed a variety of opinions: all participants were different from each other (Figure 5). Therefore, as part of the answer to how users perceive web interactions, it can be concluded that **design of web interactions influenced user experience and the experience was fairly individual**, no matter which device is used. The significance of the understanding lies in the fact that designers and developers face exactly the same challenge: users are different. Probably nothing can be designed to be perceived exactly the same way by all people.

The analysis proved that **different interfaces with the same goals provided different user experience** (e.g. Figure 11, Figure 12). Within this pilot study significant differences among stimuli existed: the "drag and drop" interaction episodes received vastly higher scores than "button to group", even though the task in the interactions was exactly the same. Nevertheless, **the assessment of different qualities of user experience such as pragmatic and hedonic qualities, and attractiveness, were proven to be homogeneous among different designs**.

The interaction episodes had some, but not major differences in user experience among each other if compared between devices: laptop and smartphone (e.g. Figure 12). While the laptop episode "button to group" in both episodes had higher values than the mobile session in all groups, the "drag and drop" had differences in values among AttrakDiff groups. In pragmatic and identification hedonic qualities the mobile interaction had slightly higher scores, whereas in stimulation hedonic qualities and attractiveness the laptop received better average values.

Similarities and differences between different designs were described by multidimensional scaling of the data (Figure 13), which clearly showed that design solutions were similar on one device, although different on the other device. The "drag and drop" episode's user experience was very similar for both laptop and smartphone. Therefore, a suggestion based on this study can be made for designers that **a "drag and drop" interaction can provide more homogeneous user experience compared to a "button to group" interaction**, considering the limitations of the study.

The study has proven to be suitable for assessing user experience of web interactions across devices. Although differences among designs and devices must be considered, based on this research it can be concluded that **user experience evaluation can provide instructions for designers**.

7.2. Limitations

The research was limited to a specific target group, devices and designs. This also creates restrictions on interpreting the results of the research.

Firstly, the participants were limited to be only from Estonian bigger cities (Tallinn and Tartu), to ages 20-65, they had to be working at an office job, and they needed to be using both desktop/laptop computers and smartphones on a daily basis. The user experience of those who are from a different population, might have been different, especially for those who do not have an office job or who on a daily basis do not use computers or smartphones.

The selection of devices was limited to a laptop and a smartphone. More specifically, either OS X based MacBook Pro or Windows-based Samsung laptop, and iPhone6 were used for the study. First of all, there are several other devices that can display web content, such as tablets, smart TVs and smartwatches. Secondly, even choosing a specific laptop or smartphone can have an influence, e.g. Apple product lovers might like the interactions on an iPhone more than interactions on an Android based phone. Overall, the user experience with similar content may be different among the devices.

From different web pages and interactions available, a specific grouping interaction was used in this study, designed and developed by the author specifically for this research. The author claims that the results of the study are very much based on the specific interaction and the task related to it. The limitation is important, because probably by choosing a different web interaction, results can vary significantly. Moreover, even if grouping interaction is used, there are many different ways to design it, which again, can affect the user experience.

7.3. Further Studies

This research focuses on the interactions and does not seek to answer the practical questions why one visual design was better than the other, or why one device received higher scores based on the interaction's visual appearance. For example, the study does not provide conclusions whether in case of a grouping interaction the elements should be in a list or as a square, how much space should be between interaction elements, or should the elements be place from left to right, or from top to bottom. This is something that should be further studied, as it would give additional and valuable input to the community of designers and developers.

Additional research could be conducted to study the differences among AttrakDiff groups. For example, it can be investigated which interaction elements provide better scores for pragmatic or hedonic qualities.

Considering the conducted research, a wider audience could be addressed. Even when limiting the study to one country, all areas, not only bigger cities can be included. As younger ages are becoming the most frequent internet users, it would be important to involve them also to the further studies, including children. People from all professional spectres can be involved, including blue-collar and non-working segments of the population. Also people with different experience in web usage should be involved. For example, a new web site that teaches how to use internet should be assessed by people who yet do not have any experience with web interactions. Even throughout the population of skilled web users, some are less and some more advanced. Within the same population, more people than 17 can be

involved in the study to get a more trustworthy result, as this researched proved the user experience to be very individual.

In case the goal is to assess web interactions as a whole, a large scope should be set, considering all possible ways of interacting, either different interactions or different devices. If a specific product is interesting, that product can be used as the basis of studying the interactions.

Even though there are many possibilities to study the topic further, the methods and procedures as described in this thesis can be applied for a similar research.

REFERENCES

Adams, Edie; van Gorp, T. (2012). *Design for Emotion*.

All About UX. (2016). All UX Evaluation Methods. http://www.allaboutux.org/all-methods. Accessed 01.02.2016.

Alves, R., Valente, P., & Nunes, N. J. (2014). The state of user experience evaluation practice. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI '14*, 93–102. doi:10.1145/2639189.2641208.

Bednarik, R., Vrzakova, H., & Hradis, M. (2012). What do you want to do next: A novel approach for intent prediction in gaze-based interaction. *ETRA '12 Proceedings of the Symposium on Eye Tracking Research and Applications*, *1*(212), 83–90. doi:10.1145/2168556.2168569.

Bootstrap. (2016). http://getbootstrap.com/. Accessed 10.02.2016.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. Usability Evaluation in Industry. London: Taylor & Francis

Buxton, B. (2010). Sketching user experiences: getting the design right and the right design. FocalPress.

Charland, A., & Leroux, B. (2011). Mobile application Development: Web vs. native. *Communications of the ACM*, *54*, 0–5. doi:10.1145/1941487.

Chent, Z., Lint, M., Chent, F. et al. (2013). Unobtrusive Sleep Monitoring using Smartphones. 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops.

De Angeli, A., Sutcliffe, A., Hartmann, J. (2006). Interaction, Usability and Aesthetics: What Influences Users' Preferences? Proceedings of the 6th conference on Designing Interactive systems. doi: 10.1145/1142405.1142446.

Diefenbach, S., Kolb, N., & Hassenzahl, M. (2014). The "Hedonic" in Human-Computer Interaction – History, Contributions, and Future Research Directions. *Proc. DIS 2014*, 305–314. doi:10.1145/2598510.2598549.

eMarketer, Inc. (2015). Tablet Users to Surpass 1 Billion Worldwide in 2015. http://www.emarketer.com/Article/Tablet-Users-Surpass-1-Billion-Worldwide-2015/1011806#sthash.IXdzaww5.dpuf. Accessed 15.03.2016.

Froehlich, J., Chen, M., Consolvo, S. et al. (2007). MyExperience: A System for In situ Tracing and Capturing of User Feedback on Mobile Phones. Proceedings of the 5th international conference on Mobile systems, applications and services. doi: 10.1145/1247660.1247670.

Google Inc. (2016). Android Developers: API Guides. Sensors Overview. http://developer.android.com/guide/topics/sensors/sensors_overview.html. Accessed 10.03.2016.

Google Inc. (2016). Material design guidelines. Usability. https://www.google.com/design/spec/usability/accessibility.html#accessibility-layout. Accessed 10.03.2016.

Hassenzahl, M, Tractinsky, N. (2011). User experience - a research agenda. Taylor & Francis: Behaviour & Information Technology. doi: 10.1080/01449290500330331.

Hassenzahl, M. (2013). Hedonic, emotional, and experimental perspectives on product quality. Encyclopedia of Human Computer Interaction.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. Mensch & Computer 2003. Stuttgart, Leipzig: B.G. Teubner.

Hassenzahl, M., Platz, A., Burmester, M., Lehner, K. (2000). Hedonic and Ergonomic Quality Aspects Determine a Software' s Appeal. Proceedings of the SIGCHI conference on Human Factors in Computing Systems.

Heitkötter, H., Hanschke, S., Majchrzak, T. (2012). Evaluating Cross-Platform Development Approaches for Mobile Applications. 8th International Conference, WEBIST 2012, Porto, Portugal, April 18-21, 2012, Revised Selected Papers.

Karapanos, E. (2010). Quantifying Diversity in User Experience. Eindhoven: Technische Universiteit Eindhoven.

Kermeka, T. O. A. G. D. (2013). Evaluating the perceived and estimated quality in use of Web 2.0 applications. *The Journal of Systems and Software*.

Laugwitz, B., Held, T., Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. HCI and Usability for Education and Work. doi: 10.1007/978-3-540-89350-9_6.

Law, E. L.-C. (2011). The measurability and predictability of user experience. *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems EICS 11*, *29*, 1–9. doi:10.1145/1996461.1996485.

Levy, Y., Ellis, T. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. Informing Science Journal. September 2006.

Lewis, J. R. (2002). Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. International Journal of Human-Computer Interaction · September 2002. doi: 10.1080/10447318.2002.9669130.

Nielsen, J. (2012). Mouse vs. Fingers as Input Device https://www.nngroup.com/articles/mouse-vs-fingers-input-device/. Accessed 09.04.2016.

Rajanna, V., Behera, D., Goldberg, D., Hammond, T. (2014). Step Up Life : A Context Aware Health Assistant. Proceedings of the Third ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health. doi: 10.1145/2676629.2676636.

Sanders, Elizabeth B. N. (1992). Converging perspectives: Product development research for the 1990s. Design Management Journal, 3(4).

Schrepp, M., Held, T., Laugwitz, B. (2006). The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers 18. doi: 10.1016/j.intcom.2006.01.002.

Seipp, K., & Devlin, K. (2013). Enhancing one-handed website operation on touchscreen mobile phones. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 3123. doi:10.1145/2468356.2479626.

Senecal, S., Léger, P., Fredette, M. et al. (2013). Mouse vs. Touch Screen as Input Device: Does it Influence Memory Retrieval? Thirty Fourth International Conference on Information Systems.

Smart Insights Ltd. (2016). Mobile Marketing Statistics compilation. http://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/. Accessed 15.03.2016.

Statistikaamet. (2011). TT259: Palgatöötajad soo ja ametiala pearühma järgi. http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=TT259&ti=PALGAT%D6%D6TAJAD+SOO+JA+AMETIALA+PEAR%DCHMA+J%C4RGI+%281989%2D2010%29&path=../Database/Sotsiaalelu/15Tooturg/06Palgatootajad/02Aastastatistika/&lang=2. Accessed 02.02.2016.

Statistikaamet. (2014). IT33: 16-74-aastased arvutikasutajad kasutajate rühma ja kasutamiskoha järgi. http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=IT33&ti=16%2D74%2DAASTASED+ARVUTIKASUTAJAD+KASUTAJATE+R%DCHMA+JA+KASUTAMISKOHA+J%C4RGI&path=../Database/Majandus/05Infotehnoloogia/04Infotehnoloogia_leibkonnas/&lang=2. Accessed 02.02.2016.

Statistikaamet. (2014). IT37: 16-74-aastased internet kasutusoskuse ja isikute rühma järgi. http://pub.stat.ee/px-web.2001/Dialog/varval.asp?ma=IT37&ti=16%2D74%2DAASTASED+INTERNETI

+KASUTUSOSKUSE+JA+ISIKUTE+R%DCHMA+J%C4RGI&path=../Database/Majandus/05Infotehnoloogia/04Infotehnoloogia_leibkonnas/&lang=2. Accessed 02.02.2016.

Stojmenova, E., Hauswirth, D., Winkler, R. et al. (2013). WEB on TV: Cross-National User Study for Designing Better User Experience. AcademicMindTrek '13: Proceedings of International Conference on Making Sense of Converging Media.

Sukale, R., Voida, S., & Koval, O. (2014). The Proxemic Web: Designing for Proxemic Interactions with Responsive Web Design. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp ’14 Adjunct, 171–174. doi:10.1145/2638728.2638768.

The jQuery Foundation. (2016). jQuery UI: Draggable. https://jqueryui.com/draggable/. Accessed 28.02.2016.

The jQuery Foundation. (2016). jQuery UI: Droppable. https://jqueryui.com/droppable/. Accessed 28.02.2016.

User Interface Design GmbH. (2013). AttrakDiff: Measurement. http://attrakdiff.de/sience-en.html#messen. Accessed 10.01.2016.

Vugt, H., Hoorn, J., Konijn, E. (2006). Affective Affordances: Improving Interface Character Engagement through Interaction. International Journal of Human-Computer Studies.

Väätäjä, H., Koponen, T., Roto, V. (2009). Developing Practical Tools for User Experience Evaluation. VTT Symposium (Valtion Teknillinen Tutkimuskeskus).

APPENDICES

Appendix 1. Screenshots of Preliminary Study's Questionnaire

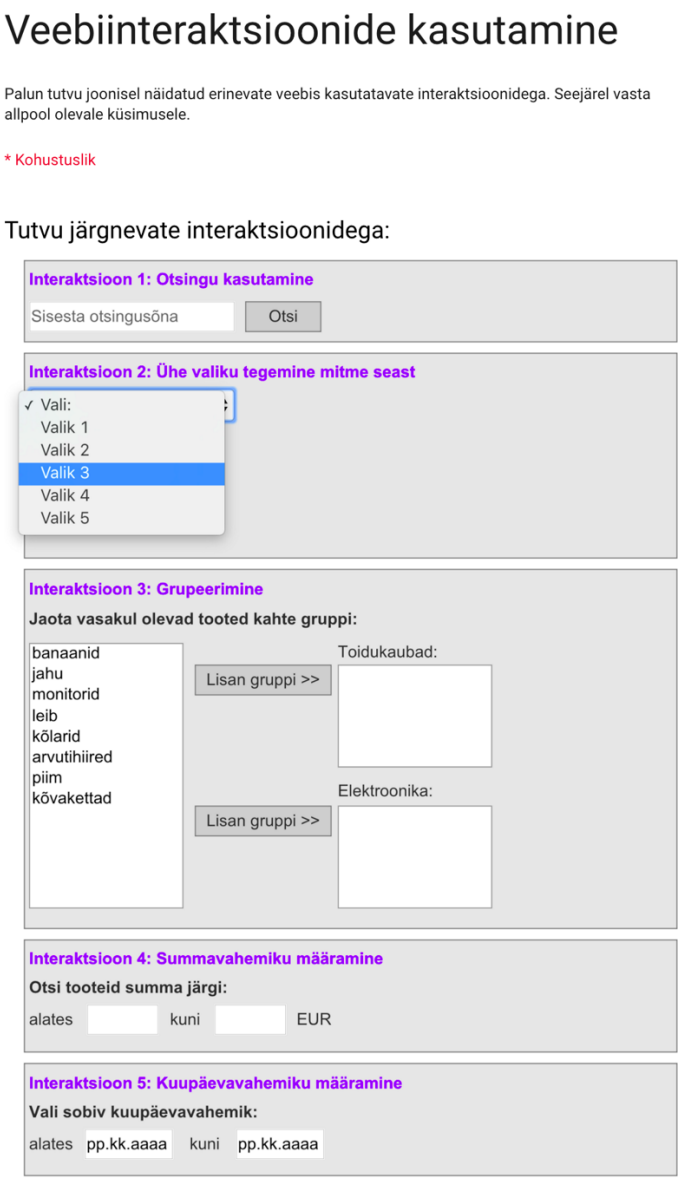


Figure 14. Introduction to the Questionnaire for the Preliminary Study.


Hinda igat interaktsiooni eraldi ning märgi ära, milline neist põhjustab vähem ja milline rohkem raskusi. *

	1 (põhjustab vähe raskusi)	2.	3.	4.	5 (põhjustab palju raskusi)
Interaktsioon 1: Otsingu kasutamine.	○	○	○	○	○
Interaktsioon 2: Ühe valiku tegemine mitme seast.	○	○	○	○	○
Interaktsioon 3: Grupeerimine.	○	○	○	○	○
Interaktsioon 4: Summavahemiku määramine.	○	○	○	○	○
Interaktsioon 5: Kuupäevavahemiku määramine.	○	○	○	○	○

SAADA ÄRA

Figure 15. Questionnaire for the Preliminary Study.

Appendix 2. AttrakDiff Questionnaire: AttrakDiff Groups and Word Pairs in Estonian and English

The following table explains the AttrakDiff questionnaire with groups and words used in Estonian and English:

AD ID	Group	Group code	Left word (Estonian)	Right word (Estonian)	Left word (English)	Right word (English)
AD1	Pragmatic qualities	PQ	tehniline	inimlik	technical	human
AD2			keeruline	lihtne	complicated	simple
AD3			ebapraktiline	praktiline	impractical	practical
AD4			kohmakas	kindla-jooneline	cumbersome	straightforward
AD5			etteaimamatu	etteaimatav	unpredictable	predictable
AD6			segadust tekitav	selgelt struktureeritud	confusing	clearly structured
AD7			korratu	kontrollitav	unruly	manageable
AD8	Hedonic qualities – identi-fication	HQ-I	isoleeriv	ühendav	isolating	connective
AD9			eba-professionaalne	professio-naalne	unprofessional	professional
AD10			maitsetu	stiilne	tacky	stylish
AD11			odav	esmaklassiline	cheap	premium
AD12			võõrastav	kaasav	alienating	integrating
AD13			eraldab mind	toob mind lähemale	separates me	brings me closer
AD14			esitamis-kõlbmatu	esinduslik	unpresentable	presentable
AD15	Hedonic qualities – stimulation	HQ-S	traditsiooniline	leidlik	conventional	inventive
AD16			fantaasiavaene	loominguline	unimaginative	creative
AD17			ettevaatlik	julge	cautious	bold
AD18			konservatiivne	innovatiivne	conservative	innovative
AD19			igav	kütkestav	dull	captivating
AD20			vähenõudlik	väljakutset pakkuv	undemanding	challenging
AD21			tavapärane	uudne	ordinary	novel
AD22	Attract-iveness	ATT	ebameeldiv	meeldiv	unpleasant	pleasant
AD23			inetu	atraktiivne	ugly	attractive
AD24			vastuvõetamatu	vastuvõetav	disagreeable	likeable
AD25			hülgav	kutsuv	rejecting	inviting
AD26			halb	hea	bad	good
AD27			tõrjuv	veetlev	repelling	appealing
AD28			heidutav	motiveeriv	discouraging	motivating

Table 5. AttrakDiff questionnaire - groups and word pairs.

Appendix 3. Screenshots of the Study Conduction System.

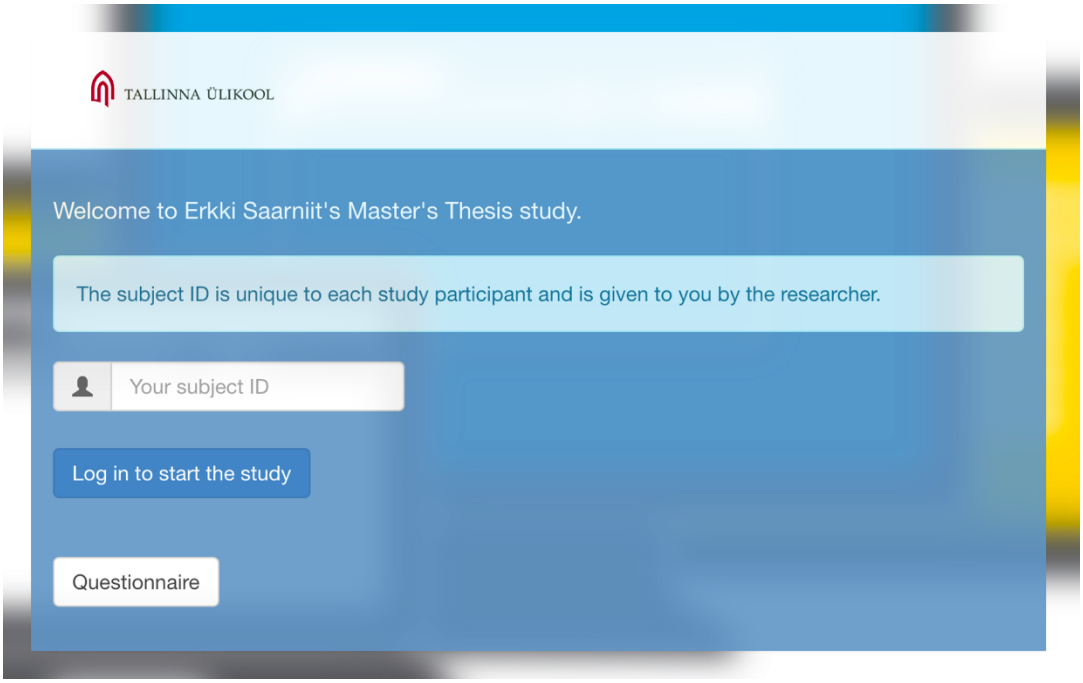


Figure 16. Login window of the study conduction system.

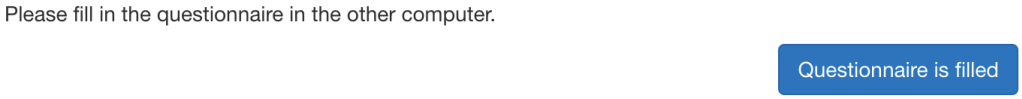


Figure 17. Information window showing instructions what to do next in the study conduction system.

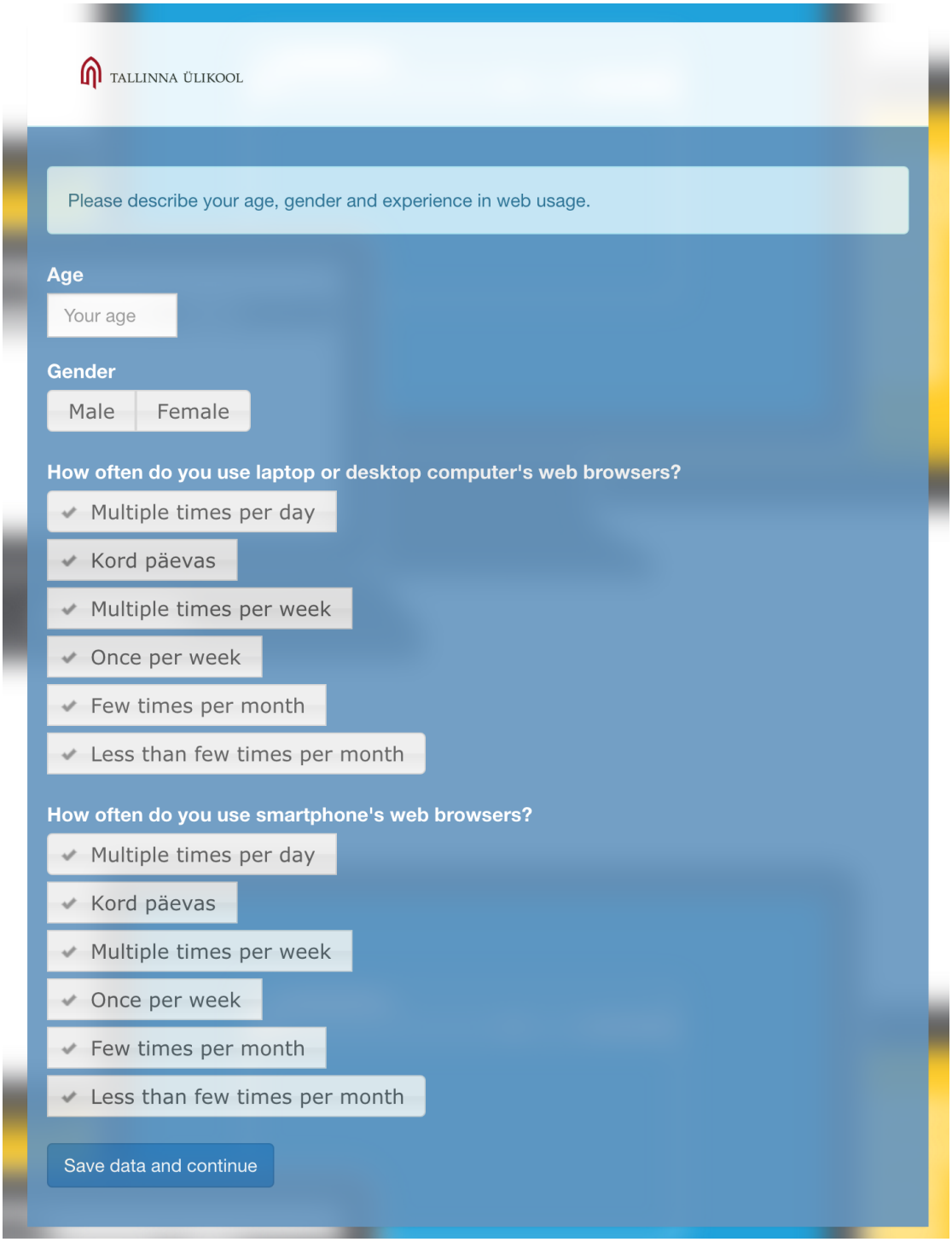
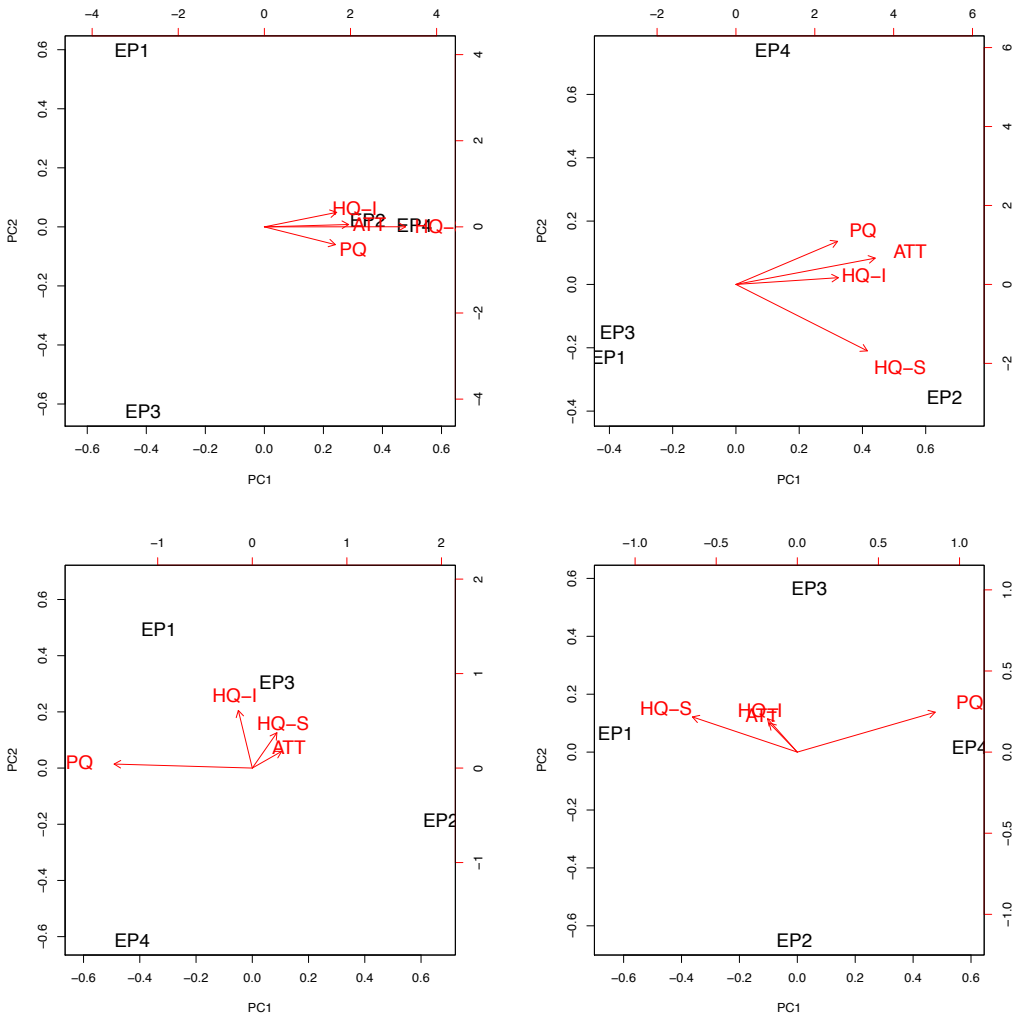


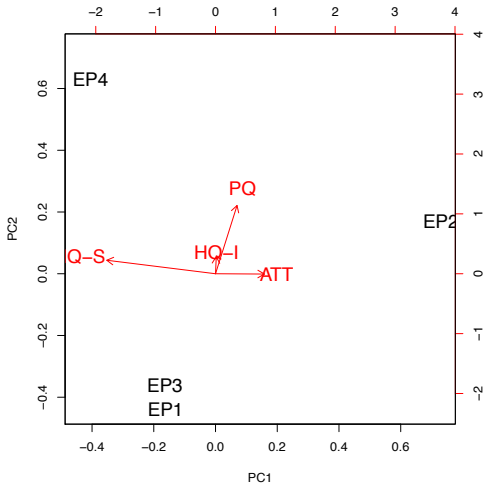
Figure 18. User registration form in the study conduction system.

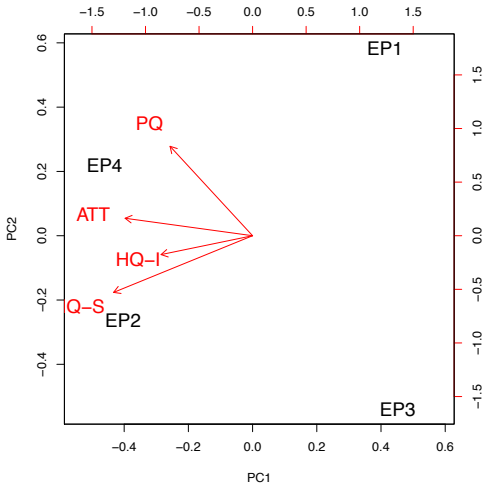
Figure 19. AttrakDiff questionnaire filling in the study conduction system.

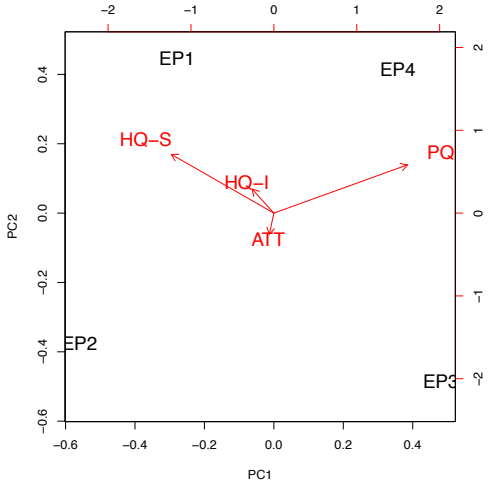
Appendix 4. Biplots of Individual Assessments of the AttrakDiff Groups and Interaction Episodes

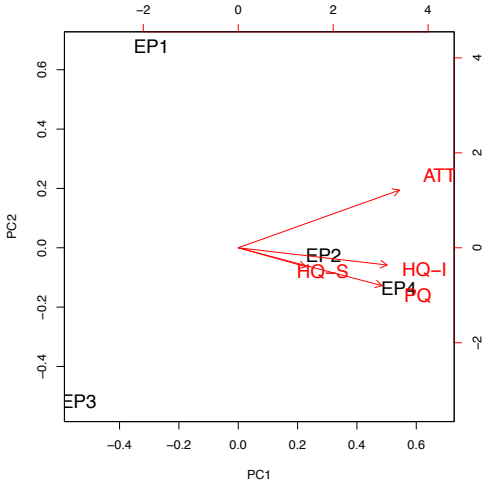
The following biplots describe the average assessments of AttrakDiff groups as one and interaction episodes as the other set of variables, for all participants in a random order. The red arrows represent AttrakDiff groups, whereas the black marks represent the interaction episodes (EP1 to EP4).

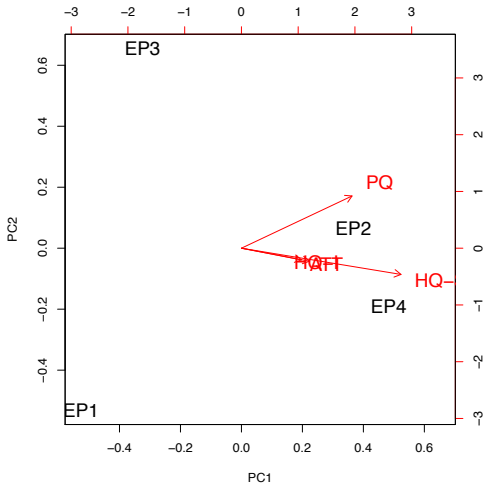


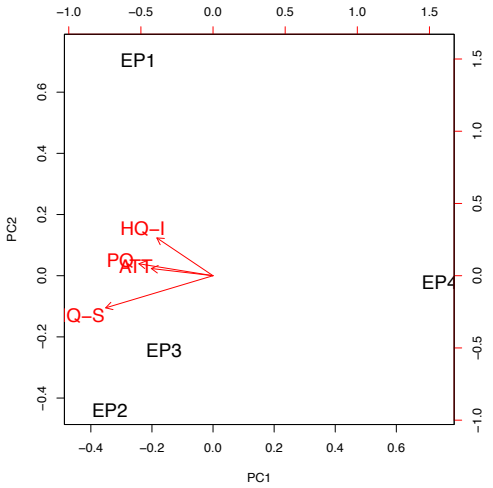

EP4
PQ
EP2
Q–S
HQ–I
ATT
EP3
EP1
PC1
PC2

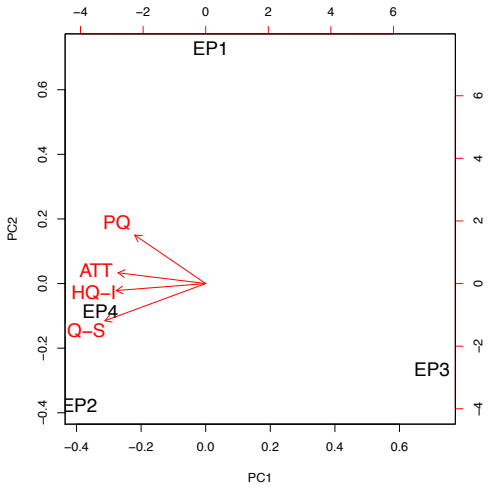

EP1
PQ
EP4
ATT
HQ–I
Q–S
EP2
EP3
PC1
PC2

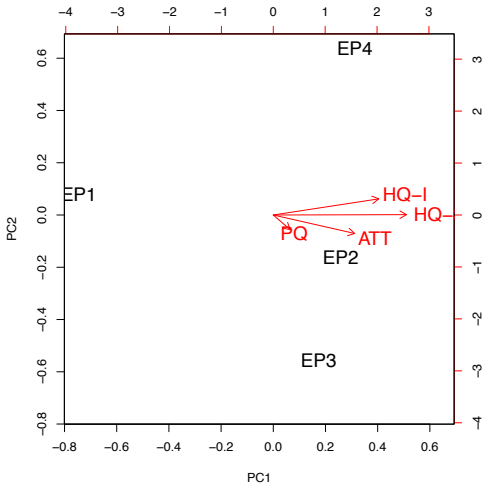

EP1
EP4
HQ–S
PQ
HQ–I
ATT
EP2
EP3
PC1
PC2

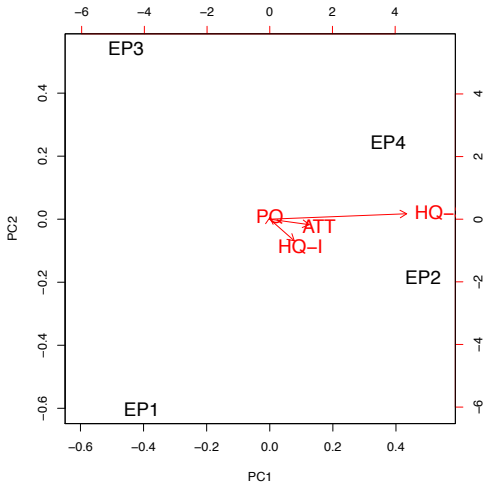

EP1
ATT
EP2
HQ–S
HQ–I
EP4
PQ
EP3
PC1
PC2

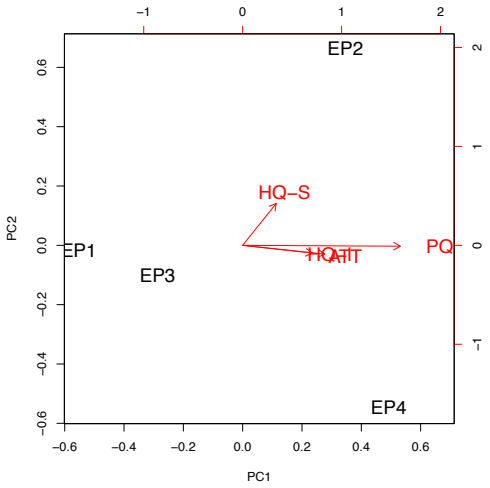

EP3
PQ
EP2
HQ–
EP4
EP1
PC1
PC2

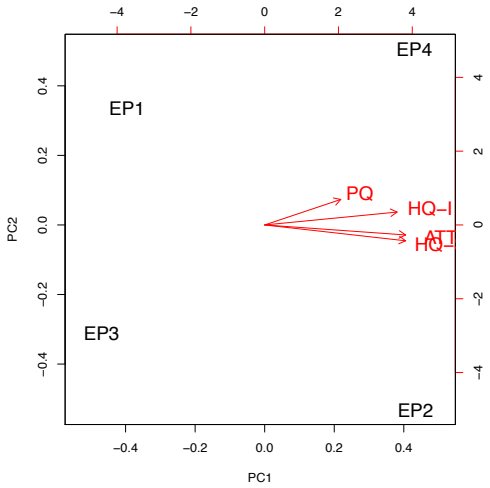

EP1
HQ–I
Q–S
EP3
EP4
EP2
PC1
PC2

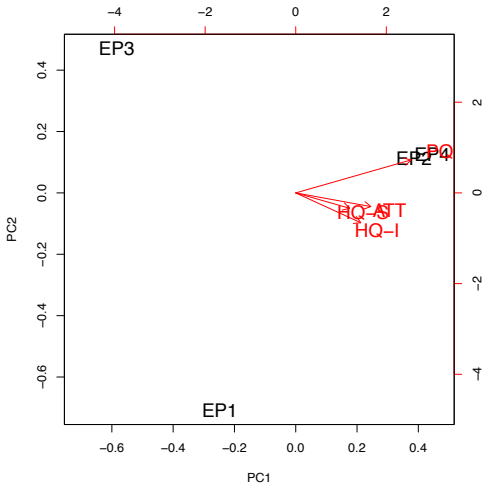

EP1
PQ
ATT
HQ–I
EP4
Q–S
EP3
EP2
PC1
PC2

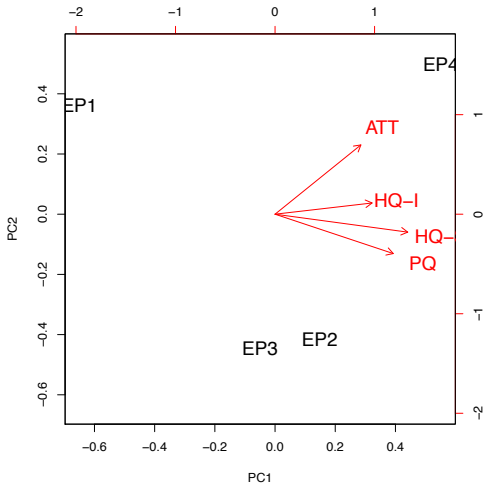

EP4
EP1
HQ–I
HQ–
PQ
ATT
EP2
EP3
PC1
PC2


EP3
EP4
PQ
ATT
HQ–
HQ–I
EP2
EP1
PC1
PC2


EP2
HQ–S
EP1
HQ–
ATT
PQ
EP3
EP4
PC1
PC2


EP4
EP1
PQ
HQ–I
ATT
HQ–
EP3
EP2
PC1
PC2


EP3
EP2
PQ
HQ–S
ATT
HQ–I
EP1
PC1
PC2


EP1
EP4
EP3
EP2
ATT
HQ–I
HQ–
PQ
PC1
PC2

ABSTRACT

Saarniit, E. Assessing the Influence on User Experience of Web Interface Interactions Across Different Devices. Master's Thesis. Tallinn University, Tallinn 2016, 88 pages, 19 figures, 5 tables, 43 references, in English.

USER EXPERIENCE, USER EXPERIENCE EVALUTION, INTERFACE, WEB INTERACTION, PRAGMATIC QUALITIES, HEDONIC QUALITIES, CROSS-DEVICE, MOBILE, AB TESTING, ATTRAKDIFF.

The problem this thesis addressed is that designers would like to create web-based solutions that would work on different devices, but the user experience may differ across the devices, which may complicate the work of designers. The research sought answers to how users perceive different web interactions and whether there are any differences in the perceptions of interactions among laptops and smartphones. It was expected that it is possible to design web interactions that provide homogeneous user experience across different devices, and that user experience evaluation can provide instructions for designers.

A pilot study with experimental design and testing of two designs on two devices was conducted. AttrakDiff questionnaire was used for data collection. 17 participants were involved from a limited population of Estonian office workers.

It can be concluded that different interfaces with the same goals provided different user experience, and the experience was fairly individual. The assessment of different qualities of user experience such as pragmatic and hedonic qualities, and attractiveness, were proven to be homogeneous among different designs. The interaction episodes were found to have some differences in user experience among each other on different devices – a laptop and a smartphone. A "drag and drop" interaction was found to provide more homogeneous user experience compared to a "button to group" interaction. The research found that user experience evaluation can provide instructions for designers.

ERINEVATE SEADMETE VEEBILIIDESE INTERAKTSIOONIDE MÕJU HINDAMINE KASUTAJAKOGEMUSELE

Erkki Saarniit

Resümee

Magistritöö eesmärk ja uurimisküsimused

Disainerid ja arendajad sooviksid luua veebipõhiseid lahendusi, mis töötaksid erinevate seadmete peal, kuid kasutajakogemus võib seadmeti erineda.

Magistritöö uurib, kuidas kasutajad tajuvad erinevaid veebiinteraktsioone ning kas veebiinteraktsioonide tajumises on erinevusi arvuti ja nutitelefoni vahel.

Hüpotees

Peamiseks hüpoteesiks on, et on võimalik disainida veebiinteraktsioone, mis loovad homogeense kasutajakogemuse üle erinevate seadmete. Lisaks arvas autor, et kasutajakogemuse hindamise kaudu on võimalik disaineritele juhiseid anda.

Kirjanduse ülevaade

Kirjanduse ülevaates keskenduti kasutajakogemuse hindamisega seotud teemade uurimisele. Lisaks uuriti, kuidas veebidisaini luuakse, milliseid omadusi on veebidisainil ning milliste takistustega tuleb tegeleda veebi disainimisel.

Uuringu disain

Pilootuuringu jaoks valiti eksperimentaalne uuringudisain. Erinevate disainilahenduste võrdlemiseks otsustati kasutada testida kahte disainilahendust kahel seadmel: sülearvutis ja nutitelefonis. Stiimuliks valiti veebipõhine interaktsioon, mis toimub ühe veebilehe piires. Stiimuli kasutajakogemuse võrdlemiseks erinevate seadmete vahel valiti välja kaks seadet: sülearvuti ja nutitelefon. Üldkogumit piirati Eesti suuremates linnades elavate 20- kuni 65-aastaste kontoritöötajatega, kes kasutavad igapäevaselt nii arvutit kui nutitelefoni. Andmete kogumiseks valiti AttrakDiffi küsitlus, mis hindab pragmaatilisi (PQ) ja hedoonilisi (HQ-S, HQ-I) omadusi ning atraktiivsust (ATT) 28 sõnapaarile hinnangu andmise kaudu.

Uuringu läbiviimine

Hinnatavaks veebipõhiseks interaktsiooniks valiti eeluuringu põhjal grupeerimine. Grupeerimiseks loodi kaks sama eesmärki täitvat, kuid erineva disainiga lahendust. Ühes neist tuli elementide ühest grupist teise liigutamiseks vastavaid nuppe vajutada ("grupeerimiseks vajuta nuppu"), teisel juhul sai elemente liigutada gruppide vahel "tirides". Andmete kogumiseks loodi autori poolt süsteem, kus sai läbida interaktsioone ja seejärel neid hinnata AttrakDiffi küsimustiku põhjal. Uuring viidi läbi osalejate töökohas asuvas eraldatud koosolekuruumis ning selleks kasutati kolme erinevat seadet: kahte interaktsiooni läbimiseks ja ühte hindamiseks.

Tulemused

Uuringus osales 17 inimest, kellest igaüks läbis kaks interaktsiooni sülearvutis ja kaks nutitelefonis. Kokku koguti 1904 sõnapaaridele antud hinnangut ning kogutud andmed olid kvaliteetsed. Autor tegi ka algselt mitteplaanitud vaatlusi.

Analüüs ja diskussioon

Esmalt uuriti uuringus osalejate vahelisi sarnasusi ja erinevusi ning leiti, et kasutajate hinnangud olid väga erinevad. Seejärel uuriti iga kasutaja individuaalseid hinnanguid kolmel viisil: tutvuti kõikidele sõnapaaridele antud hinnangutega, sõnapaaride gruppidele keskmiselt antud hinnangutega ning koostati *biplot*id interaktsiooniepisoodide ja AttrakDiffi sõnapaaride hindamiseks.

Järgmisena võrreldi erinevate interaktsioonide pragmaatiliste (PQ) ja hedooniliste (HQ-I ja HQ-S) omaduste seoseid üle kõikide uuringus osalejate. Leiti, et "tirimise" interaktsioon oli ihaldusväärsem kui "grupeerimiseks vajuta nuppu" interaktsioon, seda nii sülearvutis kui nutitelefonis.

Viimasena vaadeldi sõnapaaride ja sõnapaaride gruppide hinnanguid tervikuna uuringus osalejate üleselt ning leiti märgatavaid erinevusi. Kõigi interaktsioonide pragmaatilisi omadusi hinnati oluliselt kõrgemalt kui hedoonilisi omadusi. "Tirimise" lahendus sai võrreldes "grupeerimiseks vajuta nuppu" lahendusega märkimisväärselt kõrgemaid hinnanguid. 75% episoodidest olid nutitelefonis läbitud interaktsioonid kõrgemate hinnangutega sülearvutis läbitud interaktsioonidest, kuid erinevused olid väikesed.

Kokkuvõte

Uuringu tulemusena võib öelda, et veebiinteraktsiooni disainil on mõju kasutajakogemusele ning kasutajakogemus on küllaltki individuaalne. Erinev, kuid sama eesmärgiga kasutajaliides loob erineva kasutajakogemuse. Kasutajakogemuse erinevad omadused (pragmaatilised ja hedoonilised omadused ning atraktiivsus) olid homogeensed üle erinevate seadmete. Interaktsiooniepisoodidel on erinevate seadmete võrdluses erinev mõju kasutajakogemusele, kuid see ei ole märkimisväärne.

Hüpotees, et on võimalik disainida veebiinteraktsioone, mis loovad homogeense kasutajakogemuse üle erinevate seadmete, sai kinnitust. Leiti, et "tirimisega"

interaktsioon pakub erinevate seadmete lõikes ühtlasemat kasutajakogemust kui "grupeerimiseks vajuta nuppu" interaktsioon. Võib järeldada, et kasutajakogemuse hindamise kaudu on võimalik disaineritele juhiseid anda.

Magistritöö pilootuuring oli piiratud nii üldkogumi, seadmete kui interaktsioonide lõikes. Uuring keskendub interaktsioonide hindamisele ning ei anna praktilist vastust küsimusele, miks mingi visuaalne disain lähtuvalt interaktsiooni väljanägemisest ühes või teises seadmes parem või halvem. Järgnevates uuringutes soovitab autor laiendada sihtrühma laiema geograafilise asukoha, vanusevahemiku, töölaadi või erineva senise kasutuskogemusega populatsioonile. Samuti on võimalik uuringut laiendada teistele seadmetele, näiteks tahvelarvutitele, nutiteleritele ja nutikelladele. Uuringus rakendatud interaktsiooniepisoodid põhinesid grupeerimisel, kuid laiema kasutajakogemuse hindamiseks peaksid olema kaasatud mitmesugused episoodid. Ka grupeerimist on võimalik lahendada mitmel viisil ning neid viise on võimalik tulevikus tehtavates uuringutes käsitleda.

Autor leiab, et magistritöö raames tehtud uuringut on võimalik edukalt rakendada teiste sarnaste interaktsioonide uurimiseks.