UNIVERSITETET I OSLO

Department of informatics

Prototypes in usability testing: the implications of richness in interaction fidelity

Master thesis

60 credits

Jo Christian Magnussen

May 3rd 2010

Abstract

Current research on prototypes in usability testing revolves mainly around the comparison of traditional high- and low-fidelity prototypes. Rapid development in technology presents interaction designers with new and more advanced tools for building prototypes, allowing for different levels of fidelity for different dimensions.

Aiming to uncover how the *fidelity of interaction in prototypes* influences the results from usability testing, this thesis can be a valuable addition to the field of interaction design. As such, this study's purpose is to investigate the effects of the level of richness of interaction in prototypes used for usability testing.

16 participants took part in the usability testing of two mixed fidelity prototypes. The two prototypes were identical, both visually and in the types of interaction mechanisms shown, but were of different level in terms of interaction. By analyzing and comparing the results from the tests it was investigated how these differences affected the outcome of the usability tests.

The analysis points towards a lesser need for high-fidelity richness of interaction in functionality that is familiar to users. Furthermore, analysis suggests that a higher level of fidelity is needed in unfamiliar features for participants to be able to provide relevant feedback.

Consequently, these results have an impact on the practice of prototyping and usability testing, as they will allow for a wiser distribution of resources when producing prototypes and require stricter criteria for recruitment of participants for usability testing.

Acknowledgements

Working with this thesis has been instructive, interesting and rewarding. I would like to take this opportunity to thank the people who helped make it possible, and those who assisted me in seeing it through to the finish.

First and foremost, I would like to thank my two advisors, Kari Hamnes and Jo Herstad. Kari took the time to offer guidance and feedback, and provided valuable assistance during the usability testing. Jo always had suggestions for improvements and ideas for future work.

Secondly, I would like to thank Handicare Norge, and Line Gry Lombnes, for volunteering their project for this thesis. NetLife Research, for providing the necessary resources, and everyone there for welcoming me with open arms.

Thanks also to all of my friends, who have supported me throughout this master's. Especially Synve Røine, who put me in touch with NetLife Research, Lars Rørstad Fossum, for the countless hours at Lasse's, and Ingrid Elise Løvlund Rekaa, for proofreading this thesis.

I would also like to thank my family, Mamma, Pappa, and my brother, Sondre. Your support, guidance, and encouragement has brought me to where I am today.

Last, but not least, my girlfriend, Silje Ramsvatn. Thank you for all the encouragement, feedback and support.

Table of Contents

1 Introduction

Usability testing is a major part of user-centered design. The use of prototypes for testing during development is common practice. Acting as physical materializations of a design idea at any stage in the development, prototypes enable early testing of features in new designs.

As such, development and testing of prototypes is an important part of the iterative process of interaction design. Besides providing a model of a design idea, they can act as a means of communication between users, developers and designers, ensuring they all have the same view.

Although prototypes can be made quickly and with a low fidelity, it is thought that a higher fidelity of interaction is needed to test the flow of and how easily users can navigate the design (Badre, 2002). However, more advanced prototypes demand more resources. Gulliksen et al. (2006) stated that usability is one of the first things abandoned when developers run out of time and money. As such, it would potentially benefit the interaction designers' community to investigate how the level of interaction fidelity in prototypes affects the results gathered through usability testing.

Little research has been performed to find out what effects different levels of interaction fidelity in prototypes will have on the results from usability testing. Perhaps a reduction in the level of fidelity is possible without affecting the participants' ability to provide feedback. If this research should show that the level of interaction in prototypes has such implications, it could have an impact on procedures for both prototyping and usability testing.

Two mixed fidelity prototypes were built for the purpose of this study. The subject of these was a website for a company specializing in equipment for people with disabilities. In appearance, the two prototypes were identical but differed when it came to the level of interaction. Multiple usability tests were performed to make it possible to provide input on the proposed research questions. A number of different books exist on the subject of usability testing, all of which give explanations on how to plan, perform and analyze the outcome (Barnum, 2002, Dumas and Redish, 1999, Rubin and Chisnell, 2008). Using the think-aloud technique during testing allowed for a better understanding of how participants were thinking when performing tasks. It also provided valuable insight to their expectations and thoughts on functionality.

In total, 16 participants from three different user groups engaged in usability testing of the prototypes. They were recruited based on their age, experience with the Internet and familiarity with disabilities of mobility. One group had no connection to anyone with this kind of disabilities, a second group lived with people who suffered from this kind of disability whilst the third group consisted of people who themselves were disabled.

The participants were divided into two groups and asked to perform a series of tasks using one of the two prototypes. One half performed the tasks using the prototype with high-fidelity interaction, whilst the other half tested the prototype with low-fidelity interaction.

A qualitative analysis of the gathered data was performed. Partial transcriptions of the recordings made during testing provided an important overview of trends and important findings, and allowed for comparison of the two groups.

Participants' expectations and impressions were compared to uncover possible differences in results from the two prototypes. Results gathered were used to answer the proposed research questions.

1.1 Research field

It was not until the early 1980s that the term *human-computer interaction* (HCI) became common. With roots in the emphasis of manual tasks in the factories of last century, it transformed as the increasing number of machines led to more and more people specializing on the interaction between humans and machines. Originally referred to as *man-machine interaction*, this field of research is now called *human-computer interaction* (Dix et al., 2004). The reason for this is a desire to emphasize that it is computers that are being interacted with, and that not only the male half of the population engage in this interaction.

The first users of computers were engineers. They were also the ones who developed new hardware. This lead to hardware systems that were easy to operate for other engineers, but virtually impossible for someone without the knowledge that the engineers possessed. As more and more people began using computers, this became a problem that had to be dealt with.

The solution came through combining the field of computer science with other research fields. As psychologists teamed up with computer specialists, the graphical user interfaces (GUI) were now built based on knowledge about both the human mind as well as the technology. This mixture of research fields, with a basis in computer science, gave rise to the field of *interaction design*.

The development of computers continued, and although some thought the Internet to be nothing but a hype in the mid 1990s (Osvold, 1996), it has grown to become a part of everyday life for millions of people all over the globe. When only professionals used computers, it made sense to require more learning and expertise. Today, when almost everyone owns and uses a computer, the demands put upon users cannot be that challenging (Nielsen, 1993). Usability engineering is about putting the user in the center of development. In the online society we inhabit, usability has grown to be of vital importance. Playing a major part in deciding whether or not a website will be successful or not (Nielsen, 2000), it is not to be taken lightly.

As more and more people use and depend on computers in their everyday life, the need for additional resources into interaction design presents itself. Graphical designers, technical writers, sociologists; they are all invaluable during interaction design, working together to ensure products that are usable for the majority of the population.

The interaction design process for making a GUI can be divided into four parts (Preece et al., 2002): *to understand users*, *develop new designs*, *prototype* and *evaluate*. The first part is where the gathering of requirements is performed. The second part is the development of design specifications to ensure that the product will cover the needs identified through the first step. Next, interactive models of the new designs are built, and the fourth step is to test these to ensure they fulfill the requirements. The data gathered in one step will result in information that can be used in the next step, giving a basis for the iterative process of interaction design. Repeating the cycle over and over again, findings from evaluation in one round are brought into the knowledge designers have of users, new solutions are built into the prototype and it is tested again. The results here gives rise to a new round of specification, prototyping and testing. Rubin and Chisnell (2008) identify gathering of data to support the next round of iteration as the overall goal of usability testing.

One of the central questions of interaction design is how it is possible to design products that are truly usable (Preece et al., 2002). A key factor in achieving this is to involve the user in the development. This is possible through the first step in designing a GUI.

In the second part of the process, the requirements and needs that have been identified form a basis for the new design. Developing alternate solutions to aid these needs and support the requirements allow designers to come up with plans that can be modeled in the next part of the process.

Gaining knowledge about what users need, want and their limitations is invaluable as a basis for requirements to ensure that a product supports usability. This design philosophy, where the software is tailored to the user and not the other way around, where the user has to adapt to the software, is termed *user-centered design* (Norman, 2002, Rubin and Chisnell, 2008). Involving users in every step of the development can help ensure a design that actually meets the users needs. As such, the methodology is important to ensure usability.

"Getting inside the head" of a typical user gives the ability to take into account what users are good at, and what they are not so good at. It allows for discovery of what might enhance their achievements and provide them with a more enjoyable and effective user experience. And last, but not least, find out what they actually want. This is possible through the third and fourth step of the design process.

Prototyping allows designers to explore their ideas and test out designs on actual users. It is an essential element of an iterative design approach, where designs are created, evaluated, and refined until the desired performance or usability is achieved. These models of the design can also help ensure that every member of the design team conceives things the same way. As interaction design teams are often made up of people from different research fields, their backgrounds and hence perspectives on things will differ. This can make a design hard to convey amongst team members, hindering the progress of design (Preece et al., 2002). Miscommunication amongst team members is one of the main reasons for products that are difficult to use (Rubin and Chisnell, 2008). In making prototypes, one creates a universal language, identical for all members of the team. This emphasizes the need for and the importance of building prototypes.

As prototypes are built, the design features in them have to be tested to ensure user satisfaction. The fourth step in designing a new GUI is to evaluate the design. To be able to perform such an evaluation, data on how the prototype performs has to be collected. This can be done through a number of various methods. Although they can all provide data on how good the functionality of a product is, usability testing is the only evaluation method that provides data from real users solving real tasks using the design.

1.2 Research questions

The focus of this thesis is on the usability testing of prototypes, and in what way different levels of interaction in these affect the results. Much research has been performed in order to study the differences between traditional high- and low-fidelity prototypes in usability testing, but the effects on participants' ability to provide feedback caused by different levels of interaction in mixed fidelity prototypes has not been investigated.

Focusing on issues such as willingness to criticize and the quantity and type of issues detected, several studies conclude that the differences are minor, and find that both high- and low-fidelity prototypes return approximately the same results, as shown through reviews of available literature in chapter 2.3.

As advances in the tools available for prototyping are made, interaction designers are faced with functionality that allows them to build far more advanced prototypes. Although having the possibility to do so, building high-fidelity prototypes still require more resources than low-fidelity ones.

Searching to uncover if different levels of interaction fidelity in prototypes influences the outcome of usability testing, this thesis approaches the issue from a new angle. Will differences in the interaction fidelity result in contrasting outcome from usability testing? Or will the level of interaction have little effect on the performance?

The following research questions are addressed throughout this thesis:

1. In what way does the level of interaction in prototypes influence the participants' ability to provide feedback in usability testing?
2. How do the findings from this research influence the process of prototyping and usability testing?

The answers to the proposed research questions could provide interaction designers with valuable knowledge to help them decide on how much interaction is actually needed to ensure quality feedback from users.

A brief evaluation of the tool used for the development of both prototypes will also be performed.

1.3 Theory

Usability

The term usability has previously been mentioned. Defining it is a difficult task. True usability is something that we do not notice, something that we do not pay any attention to. In other words, it is invisible (Rubin and Chisnell, 2008).

It is much easier to explain why something is *not* usable. Websites where it is impossible to find what you are looking for, online stores with checkout procedures so difficult users give up and check in kiosks at the airport with no possibility for backwards navigation to correct errors. When something does not work they way we anticipate it becomes difficult to use.

Supporting users in the tasks they are trying to achieve is important in usability. It is about having a product that allow users to reach their goals, and is supporting them in doing so.

> *”Usability means that the people who use the product can do so quickly and easily to accomplish their own tasks”*
>
> *(Dumas and Redish, 1999)*

To ensure products that actually fulfill this, it is important to set usability goals that will help guide the development (Mayhew, 1999). Although all projects should have some specific usability goals based on user profiles and tasks, some universal goals should apply to all designs (Preece et al., 2002). First of all, the design should be *effective*. It should help the user perform and complete the task they are doing quickly and without requiring much work. Secondly, it should be *efficient*. By this is meant that the users are actually supported in their tasks, and provided with the tools they need when they need them. *Safety* is also something all designers should strive for. For interaction designers working with online designs and products, safety is not concerned with protecting the user from physical injury. It is rather about helping them avoid situations they cannot recover from, and conditions that are potentially harmful. It should also provide the users with the tools and functionality they need to perform the desired actions, it should have a *good utility*. The last two universal goals designers should aim to fulfill are that their products functionality is *easy to learn* and *easy to remember*. Users should be able to quickly and intuitively use the product, and also remember how it functions after having learned its functionality.

Ultimately, usability is about designers being able to convey their conceptual model, the *design model*, to the user through interaction with the design (Norman, 2002). A conceptual model is a mental simulation of how something will function. The users mental model, the *users model*, is formed through interpretation of a device's perceived actions and visible structure, referred to as the *system image*. This process is shown in the figure below.

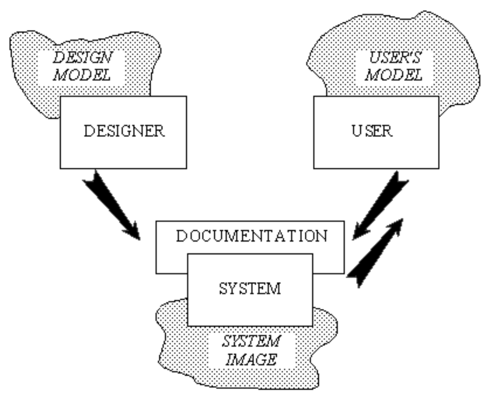

Figure 1 - The process of transferring mental models (Norman, 2002, pg. 16).

When designing a product, the designer has a conceptual model of how it is going to look and function. But for the user to develop an identical model of the designs functionality, it has to be conveyed clearly and consistently through the design. As the designer cannot guide or instruct the user directly as to how the product is meant to function, it is important that the design itself gives as many clues as possible on how it is used. If the system image is not conveyed clearly, the user will end up with the wrong mental model.

Mental models allow for a prediction of the effects caused by an action, and people build these by combining two different sources of information; *knowledge in the head* and *knowledge in the world* (Norman, 2002). Knowledge in the head is the information gathered through experiences, such as using a keyboard and eventually remembering where the keys are placed without having to look. As the user becomes experienced in using the keyboard, they become less dependent on knowledge in the world to understand the effects of an action. When people encounter unfamiliar features, they make use of this knowledge, their existing mental models, to understand (Jonassen and Henning, 1996). Knowledge in the world is the information that is available through interaction with a feature or design, such as the characters shown on the keyboard. Without any knowledge in the head, users are dependent on this to form an understanding.

Norman (2002) has come up with a few design principles, meant to guide designers towards creating a usable product and ensuring that users can understand the knowledge available to them in the world.

- *Visibility*: Functionality should be visualized clearly to the user. Hiding features or functionality will make users insecure as to what will be the next step towards reaching their goal. Hiding functionality, e.g. by limiting the number of controls for a large number of features, is destined to cause problems for users. Making functions visible increases the possibility that people will know what to do next.

- *Feedback*: Providing users with feedback on their actions is essential. Imagine sending an email to someone. If no feedback is given, how can you be sure that it is actually sent? Or when registering your contact information on a webpage. If no feedback is given on the actions you perform, it is impossible to know whether the goal has been accomplished or not.

- *Affordance*: The affordance of a design refers to how well it conveys its own functionality. A button affords clicking on, whilst a text field affords writing. Suggesting the range of possibilities, affordances give the user clues as to how and what it is possible to achieve.

- *Constraints*: This principle is closely linked to the usability goal *security*. Limiting users alternatives as to what actions are possible at any given moment, they can help ensure that users do not end up in undesired situations or cause damage due to performing the wrong actions. Constraints limit the possible actions, and are closely linked to *affordances*.

- *Mapping*: It should be possible to understand the relationship between controls and the effects caused by using these. One example of how mapping should not be done is the method Apple used to eject CDs and DVDs in earlier versions of their operative system. To eject a disc, users had to drag this from the desktop to the trash bin. As the trash bin symbolized throwing something away or deleting it rather than ejecting, many users found this confusing. The action did not match the effect caused by it, and Apple have since corrected this error by changing the icon from being a thrash can to depicting an eject symbol when users highlight discs on the desktop.

- *Consistency*: Making functionality perform the same outcome throughout the design can help make it easier to learn. Changing the placement of buttons, or their functionality and effects throughout a design, will cause it to be more difficult to use.

Although the previously described usability goals and design principles apply to all development within the field of HCI, there are some fundamental differences when comparing the web to software GUI's (Badre, 2002). Perhaps the most important of these is the lowered switching costs users experience in using the web compared to applications installed on the users machine. When users buy software for their computers, they are much more likely to adapt to small usability issues in the design and learn how to use this rather than changing it for something else. As they have spent money to acquire it, and time to install and learn this new software, it is harder for them to change it.

On the web however, users often have several sites offering the same functionality and services easily available. Choosing to switch from one site to another is easy and free, hence users are much more liable to search for other alternatives if they run into usability issues in the design. As such, usability is key to whether a website will be successful or not (Nielsen, 2000). Also, designers have to take into account that users can enter the site from any page in the hierarchy, through bookmarks or search engines. When designing for the web, several other factors such as variations in web browsers, monitor sizes and what kind of device the user is utilizing also has to be taken into account when working to ensure designs that support usability.

Interaction

Also the term *interaction* is difficult to define. It is made up of the two words *inter* and *action* (Harper, 2010), *inter* meaning *among* or *between* whilst the verb *action* stems from the Latin *agere,* meaning *to do* or *a doing*. In psychology, interaction is defined as a *"relationship between two or more systems, persons or groups resulting in reciprocal influences"* (Corsini, 2002). In other words, every action results in a reaction.

In HCI, the focus is on the relationship between humans and machines, and the structure of communication between these (Hewett et al., 2009). Here, interaction is about the joint performance of tasks between humans and machines.

For the purpose of this thesis we define interaction as *the changes in appearance that occur when the user interacts with the prototypes features*. By features is meant the elements offering possible manipulation.

For a prototype to have a high level of interaction, the features it inherits have to be functional and provide feedback when the user manipulates them. Such features include hover effects causing hyperlinks to change color when the cursor is placed over them, drop-down menus that function, image browsing functionality and working tabs. The closer the features used in a prototype are to that of a finished design, the higher the level of interaction. Consequently, a prototype with a low level of interaction will have few features that are fully working and no visible changes to the contents.

Prototype Fidelity

The term fidelity refers to the *level of detail* in a prototype, and there exists several definitions for this. Rudd et al. (1996) identify high-fidelity prototypes as fully interactive, providing users with the ability to interact as with the real product. As an opposite, low-fidelity prototypes do not provide navigation or interaction and, although they include colors, icons and controls, they serve only as indications as to the direction the design is heading. They argue that the fidelity of the prototype is not based on its closeness to a finished product, but rather on how the user experiences it through interaction. Bonner and Van Schaik (2004), although using *level* as a substitute for *fidelity*, also define a high-level prototype as one "where all or most of the functionality and often the form of the interface is fully represented". Sauer et al. (2008) has the following definition of prototype fidelity:

> *"The degree to which a model of the system resembles the target system refers to the fidelity of the model. The fidelity of the model (or prototype fidelity) may vary considerably, ranging from a low-fidelity simulation of the system (e.g., paper prototype) to a fully operational prototype, which is (almost) identical to the real system."*

As such, prototype fidelity refers to the level of detail compared to the finished product. Consequently, the more detailed and closer to the finished product a prototype is crafted, the higher the fidelity.

Throughout this thesis, both *fidelity* and *level* are used to describe the degree of resemblance to the finished system.

The concept of high- and low-fidelity prototypes is further explained in chapter 2.1.1.

1.4 Overview

The remainder of this thesis is made up of six chapters. The following chapter gives a presentation of prototypes, the different groups of these, and reviews previous research involving comparison of traditional high- and low-fidelity prototypes in usability testing. Starting with a brief look at qualitative and quantitative methods, chapter three gives a thorough description of usability testing as an evaluation method. It describes the process in detail, from planning and preparing, through testing and data gathering, to the evaluation of results. It also offers some insight into possible pitfalls of this method. Chapter four describes the practical work performed throughout this study. It presents the tool used to build the prototypes, the prototypes themselves and the usability testing of these. The fifth chapter presents the results gathered through usability testing. In chapter six, these findings are thoroughly discussed in relation to the proposed research questions, along with a brief review of the prototyping tool. Chapter seven presents the conclusions of this study, together with limitations and suggestions for further research.

For readers with an additional interest in the study, further information on the research can be found in the appendix. In appendix A, the complete criteria for recruitment of participants are presented. Appendix B lists the participants who took part in the usability testing. The tasks used for testing of the two prototypes are presented in appendix C, whilst the informed consent form participants signed prior to testing can be found in appendix D.

2 Literature review

This chapter offers a description of prototypes, the categorization of these, and presents some earlier research on comparative testing of high- and low-fidelity prototypes.

2.1 Prototypes

Although usability testing can be performed at any stage of development, it is often wise to do some testing early in the process to ensure that the design works as intended. As such, a prototype is an invaluable tool for designers, allowing them to explore their ideas without having to spend much resources on development (Berkun, 2000).

The use of prototypes began in the development of hardware (Dumas and Redish, 1999). Basically, a prototype is a representation of a design idea at any stage in the development process. They are used to convey designers' thoughts to others, test features of a design or as a means of comparing different design solutions (Preece et al., 2002). A prototype can range from sketches on paper to fully functional software, from simple cardboard models to detailed constructions in metal or plastic. Houde and Hill (1997) state that a prototype is, regardless of what medium is used, any representation of a design idea. A pizza box with a piece of wood can simulate a laptops weight and size, whilst a sketch on a piece of paper can be used to visualize the GUI of a new website. It all depends on what you want to find out and the context in which the prototype is built.

Prototypes can be used as a basis for graphical design, or to form design specifications and communicate these to developers (Nielsen, 1993). The whole idea behind them is that they are both quick and cheap to produce, provided you keep the fidelity at a low level. Real users can test the functionality of the design, whilst it is still possible to make changes without it being too expensive.

There are several reasons for building prototypes. Having a physical object to show often enables a "common" language for developers and clients, making discussions easier. A prototype is also a way to show developers thoughts and intentions for a product. The reasons for making prototypes can be divided into three categories according to Berkun (2000), namely *proof of concept*, *design exploration* and *technical exploration*.

P*roof of concept* simply refers to using a prototype to convey a new idea to others. Having a physical design that shows how it is intended to work often makes it easier to make others share your point of view. As such, the use of prototypes can be vital if disagreements arise within a design team.

Design exploration means using a prototype to explore solutions for specific problems in a design. This reason for prototyping also includes communicating the functionality and appearance of a design to developers.

The third reason for building prototypes, *technical exploration*, refers to when different technical solutions for a design are tested. This is valuable as different technologies come with different tradeoffs.

Managing and allocating resources becomes a lot easier if the purpose for building a prototype is established (Berkun, 2000), along with precise goals for the development. Making the correct decisions about the prototypes fidelity is essential to ensure good results in an iterative design process (Olsen, 2005). This decision is often influenced by factors such as available time and resources. Low-fidelity prototypes are cheap to produce, but can result in invalid data. High-fidelity prototypes provide valid data, but require a lot more resources. Choosing between them is not always easy.

There are several ways of categorizing prototypes. Houde and Hill (1997) propose that this should be done using the artifact itself rather than its attributes. They can be divided into horizontal, having a wide range of features but none fully implemented, or vertical, with only a small set of features that are modeled completely (Dumas and Redish, 1999). Nevertheless, the most frequent way of categorizing prototypes has been to divide them into two main groups: high-fidelity and low-fidelity.

2.1.1 Traditional divide

Traditionally, prototypes have been divided into two main categories. Either as simple depictions of the design, called *low-fidelity prototypes*, or more advanced and finished looking representations, *high-fidelity prototypes*. The two categories are further explained below.

Low-fidelity prototypes

Low-fidelity prototypes, sometimes also referred to as *mockups*, are characterized by that they are simple and rough representations of a design. For example, a drawing on a napkin can be considered a low-fidelity prototype. Other examples are screenshots and simple wireframe layouts of the design, sometimes with a limited level of interaction.

Sefelin et al. (2003) describes the low-fidelity prototypes as physical manifestations of ideas designers have at an early stage of development. As they are rough depictions of the design, similarity to how the finished product will eventually look and function is often very limited. Intended to give a basic understanding of how the completed design will be (Preece et al., 2002), they will seldom be implemented into the final design.

That they are both cheap and quick to produce is perhaps the most noteworthy features for low-fidelity prototypes. Often utilized in the early stages of the design process, they allow for testing of the established concepts prior to having spent the entire budget on development (Eie, 2010). This enables fast changes, and makes it less difficult to discard them if they fail to perform as expected.

One major drawback in using low-fidelity prototypes for usability testing is that users tend to mentally compensate for lack of fidelity when faced with these (Sauer and Sonderegger, 2009). This is a potential hazard, as it could result in invalid results.

High-fidelity prototypes

High-fidelity prototypes are characterized by that they look much more like the finished design than low-fidelity prototypes (Preece et al., 2002). Often, the gap between a high-fidelity prototype and the finished product is relatively small. Colors used, level of interaction and functionality tend to be more or less the same as in the finished design.

A major strength in the high-fidelity prototypes is that they are excellent tools to convey specifications and system behavior to the ones developing it (Virzi et al., 1996). Often, when developers do not have clear and precise specifications to build on, they tend to think up their own solution to a problem. As they are seldom the typical user of the product being designed, this often results in unwanted usability issues (Dix et al., 2004). Being able to show developers how the design should look and function decreases the chance of this happening.

Another advantage in the high-fidelity prototypes is that they are often close enough both in functionality and design to provide strong conclusions about the finished product when being evaluated.

As these types of prototypes are more advanced and detailed, they require more resources to produce. Rettig (1994) identifies this, amongst other issues, as a problem with this kind of prototypes. The increased need for resources such as time and money in development can also result in a hesitant attitude to make changes to, or even discard the prototype. Having spent a lot of time developing a prototype creates "ownership" that makes it harder to accept that things might not work as well as they are supposed. Also, an advanced prototype can create false expectations as to how far along development has come, as they often give the impression of a fully operational system. Rettig (1994) also points to that a visually perfected prototype tends to generate more feedback on visual elements, such as selection of font, colors and pictures, rather than operational ones.

Finally, an advanced prototype needs to be fully operational when testing begins, as one small bug can make the whole test come to an abrupt stop. This stresses the importance of running pilot tests prior to testing the prototype with real participants.

2.1.2 New perspectives

The development of computers is happening at a staggering pace. This also has an effect on the tools available for prototyping. These changes are forcing us to view prototypes in a different way.

The separation into high- and low-fidelity prototypes used to be based on the overall similarity to the finished product. However, as new tools become available, dividing them into two groups is no longer sufficient. It is now possible to build prototypes where the different dimensions have different levels of fidelity. A prototype can in other words have low visual fidelity, but at the same time have a high level of interaction. In this new way of looking at prototypes, they are referred to as *mixed fidelity* (McCurdy et al., 2006).

McCurdy identifies five different dimensions that can be used to define a prototype. These dimensions are found in all prototypes, and each one can either be low-fidelity or high-fidelity. The first dimension is *the level of visual refinement*, also referred to as the *look* of the prototype (Snyder, 2003) or its *aesthetic refinement* (Virzi et al., 1996). This refers to how finished the prototype looks, in regards to its visual attributes. At one end of the scale, we find simple wireframes and hand drawings. These are often very simple, without color and generally look unfinished. At the far end of the scale, prototypes look more visually finished. Often, these prototypes will look like the finished website, with pictures and colors. As mentioned by McCurdy (2006) and Rettig (1994), a high level of visual refinement is not always something to strive for, as this has a tendency to make the people using the prototype in testing provide more feedback on visual attributes than on the functionality.

The next dimension is *the breadth of functionality*. Virzi et al. (1996) refers to this dimension as *breadth of features*. This can also be viewed in relation to what is called a horizontal prototype. It refers to the amount of functions that are present in the prototype. Take an online library as an example. Is it possible for users to create user profiles, reserve books, renew loans or make inquiries about available books? The wider the representation of functions, the easier it is to understand what capabilities will be found in the finished product, and the higher the fidelity of breadth of functionality.

The depth of functionality is the next dimension used to define a prototype. Virzi et al. (1996) also comments on this being one of the dimensions where prototypes can vary in their fidelity. This dimension is also referred to as a vertical prototype and describes how close to the finished system the different features are modeled. Consider the online library again. Is it possible to carry out a renewal of loans in the prototype, all the way from logging in to getting a confirmation that the loan is renewed? A prototype with a high-fidelity depth of functionality will have functionality that allows users to complete an action, whilst one with low-fidelity depth will not. The closer functions get to their conclusion, the higher the level of depth the prototype.

The fourth dimension is *the richness of interactivity*, and this simply deals with how much interaction has been implemented in the prototype compared to the finished system. Does the system respond to users actions, or is it flat and lifeless. Are text fields and drop-down menus possible to manipulate? Will text from one input appear on other pages? Historically, richness in interactivity was only found in high-fidelity prototypes, but advances in prototyping tools have made it possible to create rich interactive behavior in prototypes with e.g. low level of visual refinement. Usually, a high-fidelity richness of interactivity will entail higher production costs for the prototype. One of the main reasons for building prototypes with a high-fidelity richness of interactivity is to be able to test both the flow of and how easily users can navigate the design (Badre, 2002).

Virzi et al. (1996) has a slightly different approach to this, through what they call *similarity of interaction*. Whilst McCurdy refers to responses from the system and the representation of interactive elements, Virzi et al. describe this dimension as how the user can physically interact with the prototype. Their examples of interaction methods are pressing buttons, clicking a mouse and touching a screen.

The last dimension McCurdy indentifies, is *the richness of data model*. This refers to whether or not the prototype uses the actual data model intended for use in the finished product. Using the example of the online library again, does lists of books available reflect the actual list being used by this library? Is the list complete, or incomplete? The closer the data model is to the actual one that will be used, the less likely it is that problems will occur later in the process.

This new classification of prototypes enable interaction designers to allocate resources to prototyping in accordance with what they want to find out.

2.2 Comparative research using prototypes in usability testing

There has been much research on the comparison of high- and low-fidelity prototypes in usability testing, and whether or not they produce different data. Catani & Biers (1998) compared three different prototypes to investigate the effects of variations in fidelity. In their study they used one interactive computer based prototype, one non-interactive slide show displayed on screen and one non-interactive paper based version. They concluded that there was no significant difference neither in frequency nor significance of usability issues observed for the three prototypes. Also, the different issues identified were fairly similar for all prototypes, further supporting this claim.

Virzi et al. (1996) also compared the type of usability problems detected using low- and high-fidelity prototypes. Trough testing of one high-fidelity computer based prototype and one paper based low-fidelity prototype, they found that there was little difference in the results between the two. Participants for both prototypes uncovered largely the same usability issues. Also, they found that a lot of the problems were found by a similar number of participants for both prototypes, suggesting that the difference between high- and low-fidelity prototypes is minimal.

Walker et al. (2002) used one low-fidelity paper prototype and one high-fidelity hypertext markup language (HTML) prototype in their research, following up on previous studies of whether differences in fidelity would result in identification of different usability problems. They found that there were few differences in both number and type of issues identified, as well as the severity of these, further supporting the view that there are no major differences to be found between the two.

Wiklund et al. (1992) researched whether the aesthetic quality of prototypes would influence the participants' subjective perception of usability. Through testing of four prototypes with different level of aesthetic refinement, they compared several factors through ratings and performance data. Measuring how easy the design was to learn, how easy it was to use and how forgiving it was along with the aesthetics for each version, they found that the aesthetics had little impact on the perceived usability.

This was also researched by Sauer and Sonderegger (2009) whose findings contrasted those of Wiklund et al. They found that users rated highly appealing designs as having better usability than those with moderate appeal. As such, they concluded that perceived usability is more related to attractiveness than objectively measured usability. They also speculate that users compensate for missing aesthetical refinement in low fidelity prototypes, what they call *deficiency compensation*, through creating mental models of the product being far more aesthetically pleasing than the actual finished design. In their research, the low fidelity prototype received a higher score for aesthetics than the finished product did. Through a cognitive process, users imagined how the finished design would look, and gave their ratings based on this.

Sefelin et al. (Sefelin et al., 2003) approached the subject from a different angle researching whether a difference could be found in participants willingness to criticize and comment on their thoughts for improvement when testing two low-fidelity prototypes. They found that a paper- and a computer-based low-fidelity prototype lead to almost the same quantity and quality of critical user statements. They also conclude that subjects prefer computer based prototypes as they feel they have more freedom to move around and explore using these, than when a moderator has to actively manipulate the low fidelity prototype for them. Using a computer based low-fidelity prototype also made participants feel less observed. Although the two prototypes seemed to generate the same amount of feedback, the researchers observed that the feedback provided by the computer-generated version tended to revolve more around the graphical details of the prototype.

However, Lim et al. (2006) also compared prototypes of different fidelity, finding that the medium used had an effect on the results produced from testing. They used a finished product as the high-fidelity prototype whilst two low-fidelity versions were constructed, one paper- and one computer-based. Although some of the problems participants identified were found in all prototypes, the computer-based low-fidelity version and the finished product revealed far more. They concluded that the choice of representation for interface elements is vital in prototyping.

The literature available on this subject tends to focus on the number and type of issues detected in their comparison of high- and low-fidelity prototypes. As such, it does not shed light on how differences in only one dimension, i.e. the level of interaction in prototypes, will influence the outcome of usability testing. However, as most of this research points to there being little difference in the results the two produce, there is a possibility that this will apply also to this study. Also, the deficiency compensation mentioned by Sauer and Sonderegger (2009) may be applicable to this research. Approaching the subject of comparative usability testing of prototypes from a different angle than earlier research, this thesis aims to fill the gap on this subject.

As mentioned, the effect caused by differences in the level of interaction in prototypes is an area relatively untouched by research. Consequently there is not a lot of previous literature to describe this. However, Blackler (2009) and Blackler et al. (Blackler et al., 2003b, Blackler et al., 2003a, Blackler et al., 2004, Blackler et al., 2005a, Blackler et al., 2005b, Blackler et al., 2007b, Blackler et al., 2007a) have researched what they call *intuitive interaction*. Through their research, they discovered that participants were able to use features faster and more intuitively if they had prior experience with products with similar appearance and functionality.

Defining intuitive interaction as a cognitive process, they claim knowledge from previous experience is used when people interact with new products. This enables them to understand the functionality more quickly than the ones without this knowledge. In other words, as they are already familiar with the functionality found in this feature, they have an existing mental model of how it works. As such, their mental models are already close to that of the design image, through being familiar with a similar system image (Norman, 2002). Therefore, they are not reliant on having to construct these from scratch when encountering features that are similar to ones they have previous experience with. This is also supported by a recent study (Beelders et al., 2008), concluding that users who are familiar with an interface can easily adapt to small changes in this.

3 Method

This chapter is a description of the methods used to gather and analyze data during this project. First, a brief description of the two groups of research methods is given. In chapter 3.2 usability testing as a method is described in detail.

3.1 Research methods

Research methods can be divided into two main groupings, namely *quantitative* and *qualitative*. Where quantitative research aims to collect statistical data, often in large numbers to get a broader perspective, the qualitative generally is more focused and in depth (Neill, 2007). As qualitative research is subjective, and deals with interpretations of events rather than precise measurement and analysis, it provides data that is less suitable for generalization.

Qualitative research entails many different forms of data gathering (Denzin and Lincoln, 2005). Observation of a subject or practice, either passive or as a participant taking part in the activity is one way to collect data. Often, observation is used in combination with techniques such as interviews, focus groups and researching text and documents to gather data. Audio and video can also be used, giving the researcher the possibility to go through and study the material more than once.

Several different research techniques for evaluating graphical GUIs are available, their common denominator being that they are all tools for gathering qualitative data. Examples of such methods are focus groups (Dumas and Redish, 1999), heuristic evaluations (Nielsen and Mack, 1994) and usability testing (Rubin and Chisnell, 2008). The fact that users perform actual tasks using a model of the design is what separates usability testing from the other user-centered evaluation methods (Salvendy, 2006).

3.2 Usability testing

Usability testing can be applied to a number of different consumer products. Anything from toothbrushes to dishwashers, car stereos and shopping carts can be the subject of these tests. Tests can be run at any point in the design cycle, either with a rough sketch of a design or a functional prototype (Eie, 2010).

Dumas and Redish (1999) identify the following five common features found in all forms of usability testing. First, the overall goal of testing is to improve the usability of the product. Secondly, the participants should represent real users of the product being tested. The third common feature is that the tasks participants are to perform during testing are similar to those they are expected to perform with the finished product. Fourth, all the observations that are made of the participants' reactions and statements are recorded for evaluation. And finally, the recordings are studied to uncover usability problems with the product. These are in turn investigated further to form a suggestion for changes in the design, eliminating the uncovered usability problems.

However, as this thesis revolves around the redesign of a website, it is natural to have a closer look at usability testing in the context of the web. Although the practice of usability testing is more or less the same regardless of the type of product, there are some differences. In the case of usability testing for the web, giving the user a satisfactory experience is a bigger focal point than when testing other products (Badre, 2002).

What provides the user with an enjoyable experience when visiting a website depends on the users intention. Imagine that you are looking for a book on usability testing at an online bookstore. The purpose of your visit would then be to find, and perhaps purchase, a book on the subject. When making purchases online, we want information about the product we are looking at. Pictures and in-depth information that can help us make a decision. But what if the site does not provide you with any of this? No pictures, no list of contents to help you decide whether or not this is the book you want, just a list of available titles. What happens? You try another bookstore in search of more information.

Now imagine that you, in search of another bookstore, use an online search engine. Upon hitting the search button, the engine provides a huge mass of cluttered text and images. Each result is presented with vast amounts of information, making it difficult to sort out the relevant bits. Many users would give up and try a different search engine.

These examples illustrate the need for an understanding of which tasks users find difficult to perform when visiting a website. Usability testing is an excellent way to uncover just that. It offers an opportunity to get "inside the head of the user", contributing to ensuring a satisfactory and effective website that users actually enjoy using and that fulfills their goal.

One of the main reasons for doing usability testing is to experience how actual users perform on a website, making it easier to focus on the real usability problems rather than the imagined ones (Badre, 2002, Dumas and Redish, 1999, Hoekman Jr., 2009). Having actual users testing the product is of great value to the developers, and can help uncover unforeseen problems. Allowing for first hand experience of how users actually use a website is also highlighted by Tognazzini (2000) as a main reason for doing usability testing, along with four other good incentives to perform such tests.

As usability testing allows developers to spend their time developing instead of discussing possible solutions, it can help save resources and reduce the time to market for a new product or design. Also, it makes it possible to deal with usability problems prior to launching a product, rather than having to refine it and launch a second or third update to have a fully functional product.

Used to uncover flaws and functionality that is missing from a design, it contributes to the next round of iteration by generating results and input that can be incorporated. Preece et al. (2002) present the DECIDE framework to help in the planning of any evaluation process. The framework consists of six steps:

- Determine the overall goals for the evaluation process. This should be done as the first step of planning, as they are the overall guides for the process.

- Explore the questions. It is important to identify questions that can provide answers to the overall goals. These can also be broken into sub-questions.
- Choose the evaluation paradigm and techniques. Usability testing is just one of these paradigms. The selection of paradigm also influences the evaluation technique used.
- Identify the practical issues. Several practical issues have to be identified and considered, such as the available resources, facilities, equipment and the people involved.
- Decide how to deal with ethical issues. When involving people in the evaluation process it is important to consider the ethical implications. People's privacy and confidentiality should always be protected.
- Evaluate, interpret and present the data. After the planning is completed, one should check the evaluation design for aspects such as possible biases.

Usability testing can be divided into three separate steps: Planning and preparing, doing the actual testing and evaluating the outcome.

3.2.1 Getting ready - planning and preparing

Planning a usability test can take from a few days to a few months. There are many things influencing this, such as the complexity of the product, the number of product elements that are being tested and how much work that has to be done in preparation for the tests (Dumas and Redish, 1999).

The Location

Usability testing can either be carried out in permanently set up usability labs, or on-site at the customer's location. Important factors to consider when deciding upon this is the design and measures of the testing, the logistics of the location, the number of observers, what will give them the best experience and also the availability of participants (Rubin and Chisnell, 2008).

Performing the usability testing at the customers' location entails a lot of work, but is preferable if the test requires observation of users in their actual context, participants are unable to take part in testing unless it happens at their workplace, or simply if a permanent lab is unavailable. Performing usability tests on site allows for observations of users in their natural context, but planning and logistics usually becomes more complex and the testing itself more time consuming (Rubin and Chisnell, 2008). Not using a permanent lab can also entail having to settle for whatever space is available, often reducing the efficiency of the tests.

When specific equipment is needed, or it is important that observers are located in a separate room, a permanent lab is preferred. A permanent lab can be set up in many different ways, all with different advantages and disadvantages (Rubin and Chisnell, 2008). Although there are some good reasons for testing in a lab, there are also some negative aspects of using these. First, the sterile setting of a usability test lab might seem unnatural for many participants. This could in turn affect their performance during testing. Secondly, testing in a permanent lab will in most cases involve some travel for participants to be able to attend.

The People

One can separate the people who take part in usability testing into three groups: the people who plan and carry out the tests, *the test team*, people observing the usability testing, *the observers*, and the people doing the assigned tasks and interacting with the product, *the participants*.

The test team

Depending on the skill level of the persons involved, everything from two to five persons is adequate to make up a test team (Dumas and Redish, 1999). It is also possible to run usability testing with only one person on the team, but the amount of work needed to collect data and moderate the participant makes this a difficult task. Also, having a team of only one person, observations tend to become "biased" and one-sided.

Several roles have to be filled in order to conduct a usability test. As personal differences such as level of experience, skills and mindset may affect the data each team member should remain in the same role throughout testing.

At the top of the hierarchy, the *administrator* is found. This is the person in charge of the testing and the team (Dumas and Redish, 1999). Often this person is also the project manager for the specific project. The administrators' tasks consist of, but are not limited to, delegating roles and responsibilities to the other team members, speaking on behalf of the team and making sure that the test produces valuable data.

The *briefer* is the member of the test team that interacts with the participant (Dumas and Redish, 1999). This role is often also referred to as the *moderator* (Rubin and Chisnell, 2008). This involves greeting the participant when he/she arrives, explaining the test procedures, informing them of their rights and collecting informed consent. The moderator is also the person who accompanies the participant throughout the test, and answers questions and gives guidance if the user has problems when carrying out the tasks.

Other roles defined by Dumas and Redish (Dumas and Redish, 1999) include camera operator, data recorder and product expert. Although some of these roles are outdated due to advances in technology, some are still highly relevant to usability testing. Often, these roles are combined so that one person on the team inhabits two or more roles. It is for example possible for one person to occupy both the roles of the administrator and the data collector. Combining the role of the moderator with others might present some problems, as this person is situated in the same room as the participant and should make an effort not to influence the outcome of the test.

Members of the test team may very well be usability experts. These are people who are familiar with testing procedures, who know what to look for and who are trained in analyzing the collected data. However, these persons are seldom experts on every product being tested. Therefore, it is often useful to have product developers as members of the team. As they get to see real users solving tasks with their product, they are forced to think in terms of usability. As these know the product better than anyone, they are also very useful if problems occur during testing. However, developers are seldom experienced in observation and using them on the team means they may have to postpone other tasks resulting in higher costs for the company.

The observers

Observation allows developers to experience how users interact with their product, hereby increasing their awareness of user-centered design. Having clients participate in tests as observers is something that should be encouraged. Often, designers will picture themselves as the typical user and model the design after their wishes and needs. This can be seen as a relation to how things were done back when engineers were the ones designing for engineers. More often than not, it will lead to usability issues for the rest of the users. Therefore, observations can help rid the one sided mindset developers tend to have when building new products (Rubin and Chisnell, 2008).

Having observers partake in usability testing can also help usability practitioners convince clients that their product suffers from usability issues. Having concrete examples to show, of how actual users struggle, makes it more difficult to contend than if a heuristic evaluation was performed.

Nielsen (1993) states that there should be as few observers as possible present during testing. Still, this depends on the set up of the test facility. If the observers are located in the same room as the participant, one should try to keep the number of observers low. The reason for this is that the mere presence of spectators will influence the results (Cottrell et al., 1968). However, as technology now allows for observers to be situated in a separate room, this effect can be avoided. Video and sound can be transmitted to adjacent rooms, removing the observers from the participant. Hence, utilizing a separate room for observation, a larger number of people observing is possible without interfering with the results.

The participants

Recruiting participants that represent users of the product for usability testing is of vital importance. Running the test on participants that are not representative of a typical user might result in wrong conclusions about the usability of the product (Rubin and Chisnell, 2008). As such, developing a user profile of the people using the product or design can help ensure that the correct participants are recruited. This profile should include relevant behavior, skills and knowledge of the typical users.

Finding participants for usability testing can be done in several different ways. One can use client lists, the web, or hire a recruiting company to search for suitable participants. Unless the design to be tested is for internal use, co-workers should never be used as participants in usability testing (Eie, 2010). Although it might ensure a sense of security if a product is not yet released, or if it is vital to keeping it a secret, this can just as easily be accomplished by getting participants to sign documents ensuring confidentiality.

The number of participants to recruit is also an ongoing discussion in the field of usability testing. This is influenced by factors such as available resources, the number of subgroups needed to satisfy goals and concerns, and whether or not it is important to be able to make statistics from the test results (Dumas and Redish, 1999).

The general rule is that the more people you run the test with, the more problems you will uncover. Research has shown that 4-5 participants will uncover approximately 80% of the usability issues in a design, and that utilizing 10 people will reveal 90% of the problems (Virzi, 1992). However, Landauer (1995) concluded that after 5-6 participants the benefit of using additional people diminishes. In other words, the cost of running the test on more participants begins to grow compared to the expected outcome. Having reached this point, it is better to run a new usability test using new tasks than to keep going. This is shown in the figure below.

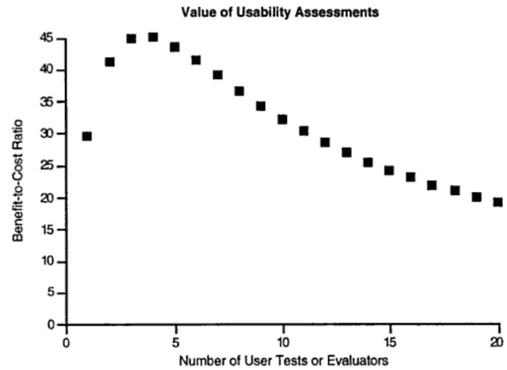

Figure 2 - The highest cost/benefit ratio is obtained when using three to five participants (Landauer, 1995)

Others find that using six participants is far from the number needed to identify 80% of the usability problems in a design (Molich, 2010), but state that this number is high enough to ensure that enough data is gathered to continue the iterative process.

Ethical issues

Participants in usability tests are often nervous and unsure as to what they have signed up for. It is therefore important that the moderator explains the proceedings of the test. This may help to calm them down, so that their focus lies on the product being tested. It is also important to stress that it is not the participant that is the subject of testing, but the product they are using.

Dumas and Redish (1999) state that it is vital, both for the participants but also for the company conducting the tests, to obtain informed consent. This is done through an informed consent form, used to secure the rights of both parties. It gives the participant information about testing procedures, the purpose of testing, that it is possible to ask questions, and that they are free to abort the test when ever and if they feel like it (Rubin and Chisnell, 2008).

Participants should also be informed that their confidentiality is ensured through that no data that can reveal their identity will be distributed or used for other purposes. It is also important to provide the participants with information as to what the results of these tests will be used for.

Data gathering

There are several ways of gathering information during usability testing. Observation is a natural part of this. Having one or more members of the test team observing from an adjacent room allows for capturing of valuable data from the tests. Several techniques for collecting data are used, either alone or in combination with others.

Interviews

Interviews can be separated into four groups; *Unstructured*, *semi-structured*, *structured* and *group interviews* (Preece et al., 2002). Structured interviews are rigid and do not allow deviations from the planned questions, whilst the unstructured interviews allows for deviations from the interview guide. The semi-structured interviews place themselves somewhere in between the two. Group interviews, or focus groups, can be productive as participants can discuss their view and offer a consensus.

Interviews are a natural part of the welcome/screening and debriefing during usability testing.

Note taking

Taking notes is the simplest form of data gathering used in usability testing. Writing down important statements or actions may help during evaluation. Although simple and cheap, this technique does not capture emotions or facial expressions. Notes also have to be transcribed and organized when testing is completed. Another disadvantage to this technique is the observers' limited speed of writing, making it difficult to capture all the data.

Audio/Video recordings

Recordings of video, audio or a combination of the two, can be a valuable source of data when analysis begins. As everything that happens during the testing is captured, it allows for multiple reviews that enable the researchers to notice details that might not have been spotted during the actual tests.

One disadvantage in using recordings of audio/video to gather data is that the focus lies on what is seen through the lens, or heard on the recording. Unless multiple cameras are used, it is difficult to get a good understanding of the situation. Seeing what the participant does on screen is one thing, but if one can't see the frustration in their faces the collected data might be misinterpreted.

A second disadvantage to using video or audio to capture data from usability testing is that it is a very laborious task to go through once testing has been completed. Transcription of one hour of recorded material may take as much as four hours to go through (Boston, 2009), influenced by a number of different factors such as the quality of the recording, how fast people talk, and the level of background noise.

The "think aloud" technique

Another way to collect data during usability testing is to make use of the "think aloud" technique. The technique itself is simple. Having participants commenting on their trains of thought whilst performing tasks allows the moderator to "read their minds". Using this technique in the early stages of the design process is especially valuable, as it can expose what participants expect from a product or feature (Rubin and Chisnell, 2008). Giving participants a brief demonstration of how this is done prior to starting the test can help them to understand how it works, and also ensures them that it is not "dangerous" to perform. Nevertheless, it is important not to force this technique upon participants.

There are several advantages to utilizing the "think aloud" technique. First, it allows for the facilitator to simultaneously gather data on both the preferences and the performance of the participant (Rubin and Chisnell, 2008). Secondly, it allows for an easier tracing of the source of mistakes made by users, as they are constantly supplying information about their understanding of the design. Allowing the moderator to understand the reason for a problem even before the participant makes a mistake, it is easier to retrace the steps and locate the culprit. A third advantage is that it generates large quantities of qualitative data using a small number of users.

Nielsen (1993) identifies this form of data gathering as *"the single most valuable usability engineering method"*. However, it is not without disadvantages. One of them being the silence that often occurs due to it being unnatural, and often quite demanding, for some participants to constantly verbalize their thoughts (Rubin and Chisnell, 2008). These participants require encouragement to keep explaining what they are thinking throughout the duration of the test, forcing the facilitator to interact with the participant. On the other hand, if a participant suddenly stops talking, it can also be a signal that they are struggling to complete a task due to issues in the GUI, or that the task is simply to difficult.

However, it is dangerous to base too much on the data gathered through this method alone, as there are big differences in how well people manage to think aloud, and the fact that the users theories as to why something is difficult may not always be the real reason. The technique can also result in invalid results as participants, due to a slower thought process, become more aware of their actions. Having extra time to think through what they are doing, mistakes that would be made in an actual use situation might be prevented.

Eye tracking

A more advanced way of gathering data during testing is through eye tracking. Using special equipment to capture the participants' point of focus, this enables the gathering of quantitative data such as how noticeable certain features or elements in the design are. It can also be a very useful asset when transcribing recorded video, as it allows for observations of where participants are looking when they perform tasks. Combining this with the think aloud technique provides evaluators with an opportunity to gain insight to the users thoughts and understanding.

Using this technique on multiple participants also enables generation of heat maps, superimposed screenshots that show where participants focused their attention when using the page.

3.2.2 Testing

When all the planning and preparation is done, it is time to commence with testing. The actual testing consists of three steps: *welcoming/screening*, *testing* and *debriefing*.

Prior to running the test on actual participants it is a good idea to perform a pilot test to make sure everything works the way it is supposed to and make sure the actual testing will run smoothly. Some of the problems, identified by Nielsen (1993), that may be uncovered through pilot testing include difficult instructions to the tasks and that there is not enough time to perform them. The most common problem however, is that the tasks participants are to perform are too difficult.

Welcome/Screening

Most participants are somewhat uncertain as to what they are partaking in when arriving the test location. It is the job of the moderator to explain the proceedings, provide information about the participant's rights, and inform them that they are not the subject of testing.

This part of the test is also used to gather some background information about the participant, in order to confirm that they are actually a representative user of the product or design.

Testing

During testing, the moderator presents the participant with the tasks they are to solve. It is vital that the moderator stays impartial throughout the test, so as to not affect the results in any way. This entails being aware of body language and reactions to the participants making mistakes. Moderators should allow participants to struggle a bit with the tasks and not rush to save them at the slightest sign of difficulties.

One or more persons from the test team will also be observing the test, gathering data for later analysis.

Debriefing

The debriefing session of the testing allows the moderator to explore the users experience further. Questioning the participants about their overall experience using the product allows moderators to shed light on the reasons for problems that occurred during testing. It offers a final chance of getting “inside” the head of the user, and seeing things from their perspective.

3.2.3 Analysis of gathered data

Analysis of the gathered data allows for identification of observed usability issues, and provides background to form suggestions for improvements. This part of testing is made up of four major steps (Rubin and Chisnell, 2008):

1. Compile and summarize data
2. Analyze data
3. Develop recommendations
4. Produce the final report

Compiling and summarizing the gathered data enables evaluators to identify patterns. Performing this activity throughout the course of testing saves time, but also enables capturing important observations that may be forgotten if they are not recorded immediately. Combining data provides evaluators with a snapshot of the events that occurred during testing. Through summarizing the data, it is also possible to identify differences between test groups or designs.

Once all the compilation and analysis of data is performed, development of recommendations and a final report can begin.

3.2.4 Limitations in usability testing

There are several reasons why usability testing can provide faulty results, such as if there are no clear goals as to what the tests are supposed to uncover, testing is performed using the wrong users, or if the results are misinterpreted.

Still, one of the biggest limitations in usability testing is the *evaluator effect*. Several studies have found that although being presented with the same data from usability testing, the identification and classification of usability issues is strongly influenced by the individual performing the analysis (Jacobsen et al., 1998, Hertzum and Jacobsen, 2003). This effect is also present in usability teams, who have been found to detect and classify usability issues differently although being presented with the same material (Molich et al., 2004, Molich and Dumas, 2008). The reasons for this is that teams carry out testing differently, often using participants that are not representative or tasks that are not suitable (Hoekman Jr., 2009). Not identifying clear and precise user goals or success metrics prior to testing can be catastrophic. Also, not having prepared a good script for the moderator to follow throughout testing can have a major influence on the results (Mayhew, 1999).

Usability testing is no guarantee for good design and usability. In some cases it is actually a hindrance for good design (Wickens, 2004). Performing test on prototypes that are not built from an understanding of the users and their needs, or without having considered basic theories and principles, will result in the iteration of a prototype that was not that usable to begin with. As such, the understanding of users and their needs is vital to ensure usability.

The location of the tests can also be a limitation as it takes place in a lab where all sources of disturbance are removed (Wickens, 2004). As such, potential problems unidentified during testing may arise when the product is used in its real surroundings.

4 Case

This chapter presents the case. Chapter 4.1 outlines the project itself and gives a brief introduction of the two companies involved in carrying out this research. Chapter 4.2 gives a description of the tool selected for building the prototypes utilized in this study, and then explains how the prototypes were constructed. Chapter 4.3 is a thorough walkthrough of how the usability testing of the prototypes was planned and carried out.

4.1 The project

For the practical part of this study, a project suitable for running comparative usability testing of the two prototypes had to be identified. This thesis is based upon a project carried out in collaboration between NetLife Research and Handicare Norge.

NetLife Research

NetLife Research is a consultancy firm based in Oslo, Norway, specializing in user experience (UX). The company was established in 2000, and has since then performed usability testing with more than 3500 users.

Offering a wide variety of services ranging from usability testing and prototyping, through graphical design and search engine optimization, NetLife Research delivers a complete UX package.

Handicare Norge

Handicare AS offers products and services for persons with disabilities, such as modification of vehicles, technical aids and medical equipment and materials. Handicare has until now acted as a wholesaling business, but wants to make the transition to a more end-user friendly supplier.

The project this thesis is based upon follows up on previous work done by NetLife Research. Some work related to the usability testing was already carried out. Sketches of the pages to be included in the prototype were finished and personas[1] were already created. The prototypes that were built and the tasks presented to participants during usability testing were based upon the material gathered and produced by NetLife Research.

The goal for this project was primarily to gather data in order to answer the research questions presented in 1.2. In addition, usability testing the prototypes would also supply data about possible usability issues with the new website design that Handicare could use in their development process.

4.2 Prototyping

Two prototypes were built for this study. As to identify how the differences in richness of interaction affected the results from usability testing, the two were kept identical apart from the richness of interactivity.

4.2.1 The tool

Upon selecting a tool for building prototypes in this project, a brief evaluation of different possibilities was carried out. In total, 13 different tools were evaluated. Looking at aspects such as the ability to build interaction, functionality that allowed collaboration and feedback as well as suitability for design of web site prototypes, the thirteen were narrowed down to three possible candidates. These three were Axure RP[2], Adobe Flash Catalyst[3] and Protoshare. Axure RP looked promising, and most likely would have been a suitable tool for this project. The fact that it could only be run on computers using Windows unfortunately made it unsuitable (Axure RP has since become available in an Mac compatible version). Adobe Flash Catalyst also looked promising, but had characteristics that made it seem much more difficult to use.

[1] *Personas* are fictitious representations of archetypical users (Calabria, 2004). Creating personas based on knowledge of real users can help focus the design effort on user goals.
[2] More information about the Axure RP software can be found here: http://www.axure.com/.
[3] The interaction design tool Adobe Flash Catalyst is further described here: http://labs.adobe.com/technologies/flashcatalyst/.

In the end, *Protoshare*[4] was selected for creating the prototypes used in this thesis. As some employees at NetLife Research were already familiar with this application, the decision to utilize this became easier. For this study, a Professional account, $49 per user per month, was created.

Protoshare is an online application. This means that no installation of software is required, and once a user account is created one has access to projects from any computer connected to the Internet. Project specifications can be exported to Microsoft Word format, and the contents of the export can be tailored to the specific recipient. The professional and enterprise accounts also allows for more advanced content giving the user the possibility to add components made using HTML, Javascript and cascading style sheets (CSS). These accounts also allow for export of projects to HTML. It is compatible with both PC and Mac, and runs on three major web browsers, namely Firefox, Internet Explorer (IE) and Safari. However, later trials showed that IE was not a compatible browser, as it was not able to correctly depict alterations to element designs done using CSS. The figure below shows the GUI in Protoshare.

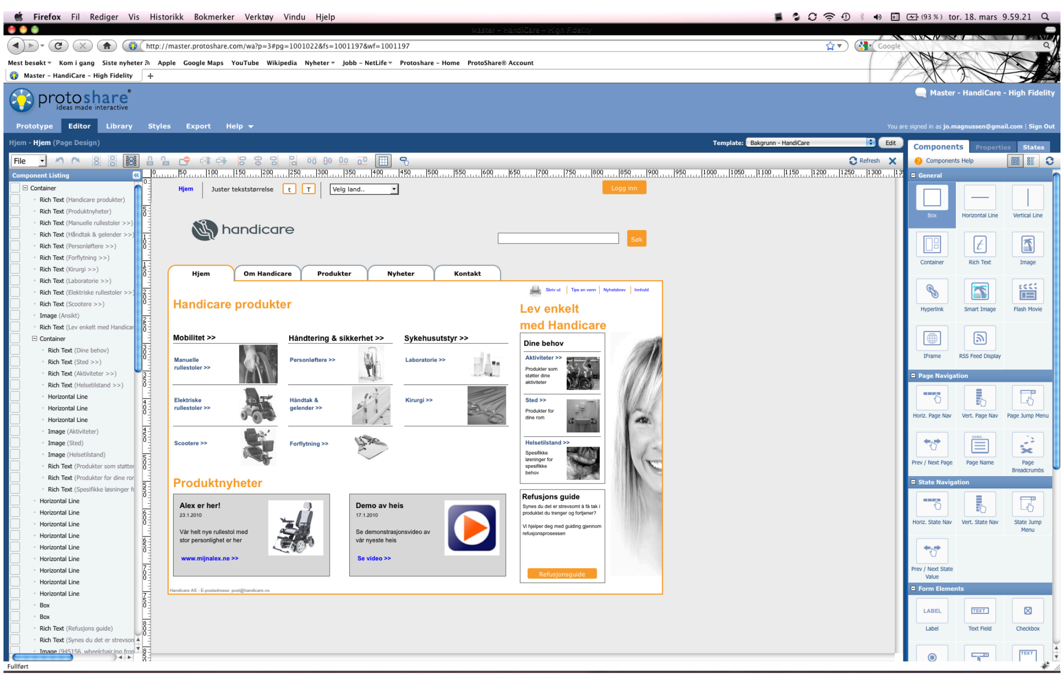

Figure 3 - The graphical user interface in Protoshare. Component listing on the left side, the page view located in the center of the page and component-, properties- and states tabs on the right side.

[4] Protoshare can be found at the following url: http://www.protoshare.com/

The service allows users to create clickable and dynamic wireframes using drag and drop, "*what you see is what you get"* (WYSIWYG) functionality. Several commonly used web elements are available in the components list, all with the functionality one would expect. Using these, building a wireframe with a moderate fidelity of interaction is quick and easy. Using the properties tab, the components can be modified in a number of ways, e.g. altering size and placement, and properties for mouseover, click and double clicks. It is also possible to disable components functionality using this tab.

Using the *states* tab, Protoshare allows for building more advanced interaction. Elements on the page can be set as invisible as long as certain states are turned off, and become visible when the state is activated. Each state can be activated or deactivated through interaction with elements in the prototype, such as buttons, drop-down menus, hyperlinks and tabs.

Another exciting feature of Protoshare, is the way it allows for more people to work together on projects, and the possibility of granting people outside the development team access to review the design. This is done through email invites, where it is possible to select what kind of access the people invited should have. These can be assigned as reviewers, developers, project-, company- or billing-managers. Reviewers will only be granted access to view and comment on projects, whilst the other user groups will receive editing rights as well. Once invited to review, it is possible to make comments on design features. This allows for focused communication between developers and reviewers, making the design process easier and saving valuable time that would otherwise be spent discussing small issues rather than reviewing more important aspects.

4.2.2 The prototypes

Both prototypes were built using Protoshare. Using the concept of mixed fidelity (chapter 2.1.2), the two prototypes were built with the same level of fidelity across the different dimensions, except for the level of richness of interactivity. However, as an effect of these differences, small variations in the two can also be identified in the breadth of functionality.

Richness of interactivity

This dimension is where the main differences between the two prototypes are found. The prototype with a high-fidelity richness of interactivity provided the user with a more complete and detailed system image, offering more visual clues on hyperlinks and more functional features. In this prototype, every action performed by the participants resulted in feedback from the system. In the low-fidelity prototype, several elements were disabled and there were no visual clues apart from the change in cursor visualization.

The first differences between the two are found on the *Home page*. This was the first page participants encountered, and also the starting point for many of the tasks they were presented with. The page is presented in the figure below.

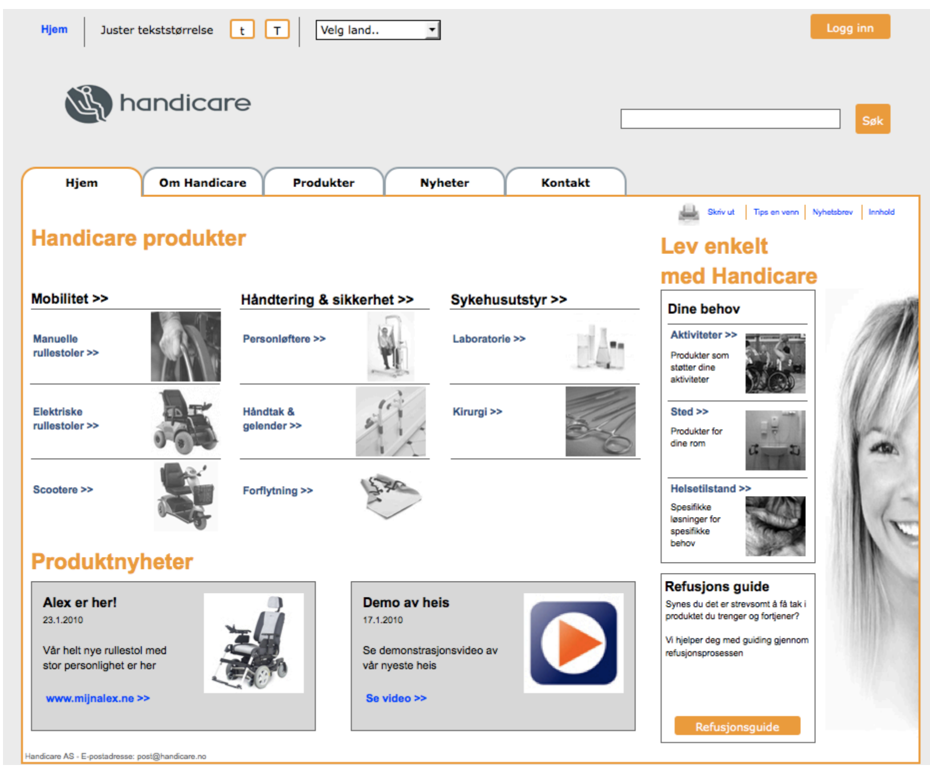

Figure 4 - The Home page.

In the prototype with high-fidelity interaction, figure 5, the main category hyperlinks, *Mobilitet*, *Håndtering & sikkerhet* and *Sykehusutstyr*, were shown with an underline and a change in color, from the default black to a bright red, when the cursor was placed over these. In addition, brackets were used to visualize that they were hyperlinks.

Figure 5 & 6 – The visualization of main category hyperlinks in the prototypes.

In the low-fidelity interaction prototype, figure 6, the only visual clues as to that these were in fact hyperlinks, were the brackets placed after the text and the change in visualization of cursor on mouseover.

The high-fidelity version also offered more visual clues as to where the participant was in fact pointing with the cursor in other features. When the cursor was placed above one of the category links on the home page, the entire area was highlighted using a shade of grey, shown in figure 7.

Figure 7 & 8 – The difference in feedback for the category hyperlinks.

Although the same size area acted as a hyperlink in both prototypes, the low-fidelity version did not provide participants with these visual clues. Figure 8 shows the low-fidelity version.

Other text hyperlinks in the two prototypes were presented using a blue color. In the high-fidelity version, these were underlined on mouseover whilst no change in visualization occurred in the low-fidelity prototype.

Differences between the two prototypes are also found on the *virtual bathroom* page. The idea behind this page was to provide users with an alternate form of navigation. Figure 9 shows the virtual bathroom.

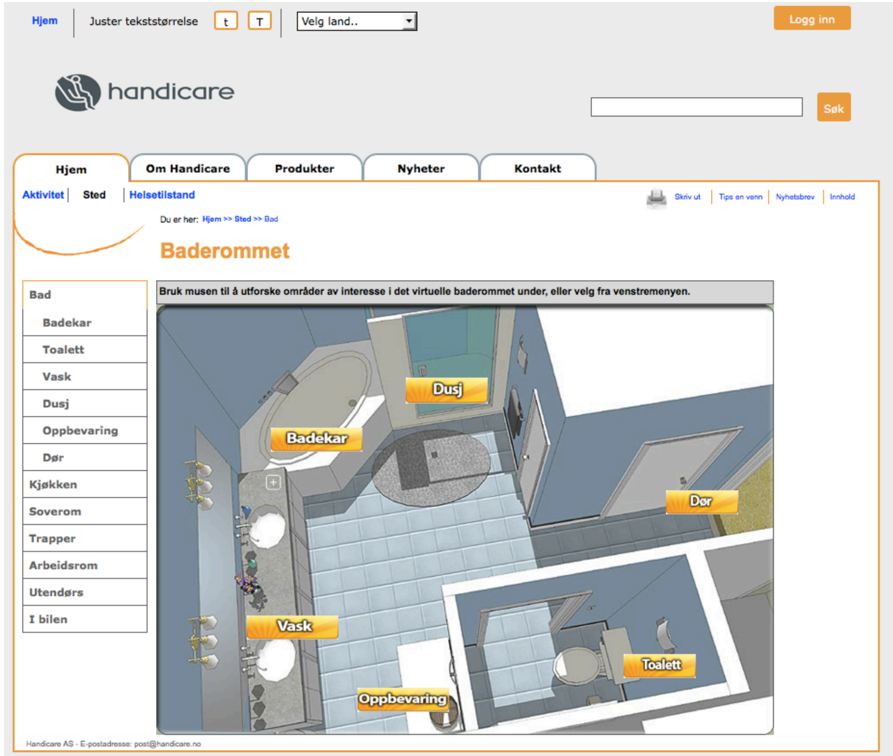

Figure 9 - The virtual bathroom.

Interaction with the buttons in the virtual bathroom would present users with an info box, as shown in figure 10. These boxes contained links to the different product categories related to the specific area in the room. Clicking these links had the same effect as using the menu on the left. In the high-fidelity prototype, figure 10, these boxes appeared without clicking, as a hover effect, when the cursor was placed over the buttons. The low-fidelity prototype required participants to actively click the buttons to reveal the info boxes, figure 11. This functionality was made using the states feature in Protoshare, described in 4.2.1.

Figure 10 & 11 – The level of interaction in buttons on the virtual bathroom page.

Although the buttons required active manipulation by participants in the prototype with a low level of interaction, they contained the same affordance and visual refinement in both prototypes.

Similar to the category links on the home page, a difference in feedback on placement of the cursor was also found other places in the prototypes. Shown in the figures below is the menu in the virtual bathroom (this menu was also present on other pages in the prototypes).

Figure 12 & 13 – The difference in feedback on placement of the cursor.

When the cursor was placed over the menu in the high-fidelity prototype, shown in figure 12, the menu item was highlighted. These visual clues were lacking in the low-fidelity version, as seen in figure 13.

More differences in the level of interaction for the high- and low-fidelity prototypes are found on the *product pages*. These provided users with detailed information and images of the different products. Figure 14 shows the layout of the product page.

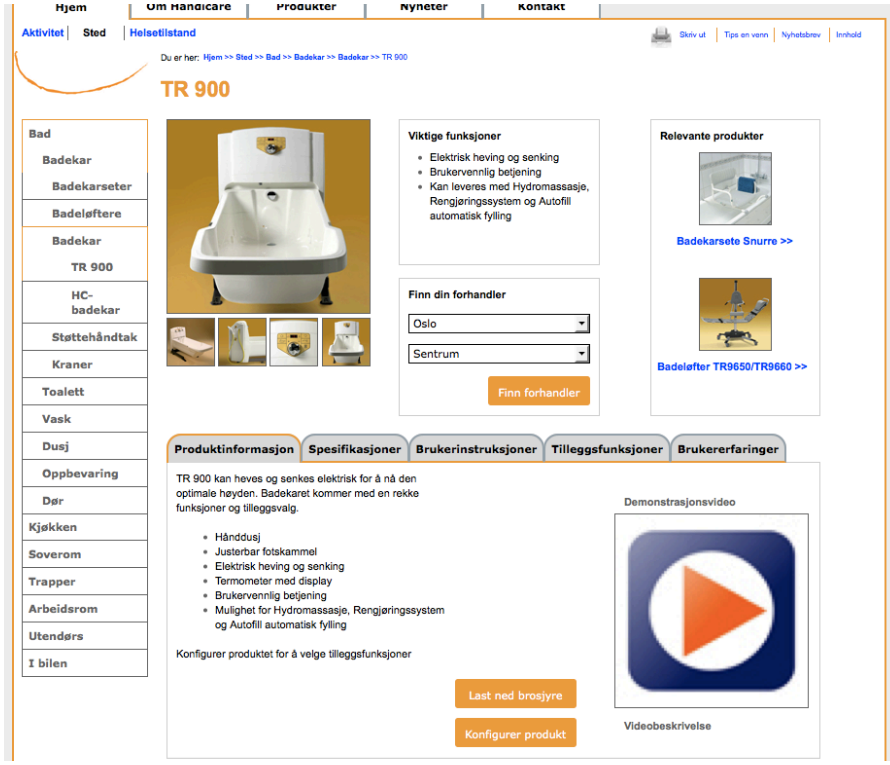

Figure 14 - The product page.

In addition to the menu on this page, explained earlier, several other differences between the high- and low-fidelity versions are found. The first of these is the image browsing functionality.

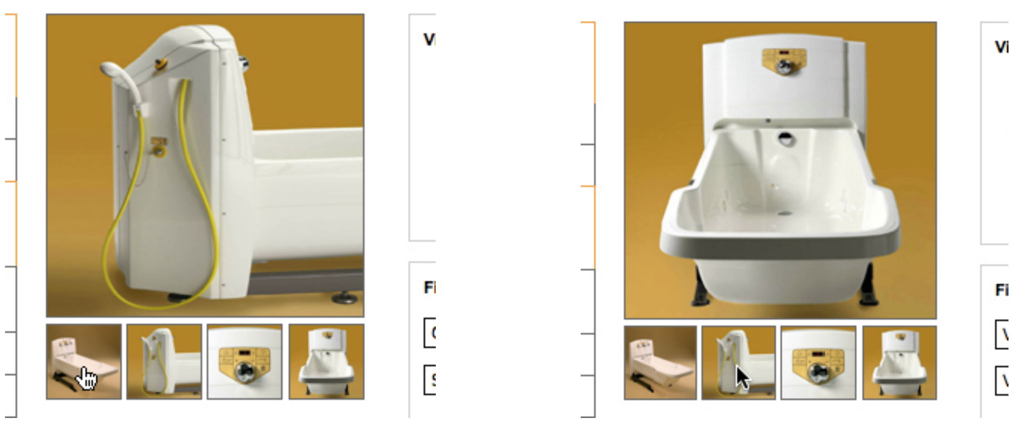

Figure 15 & 16 – The differences in image browsing functionality.

In the high-fidelity version, figure 15, it was possible to browse through images. Clicking on one of the four smaller images resulted in this image being enlarged and displayed in the square above. This functionality was disabled in the low-fidelity version, shown in figure 16.

The level of interaction in drop-down menus on these pages was also different in the two prototypes. In the high-fidelity version these were fully functional, shown in figure 17. This allowed participants to actively use this feature. In the low-fidelity version, illustrated in figure 18, these were disabled.

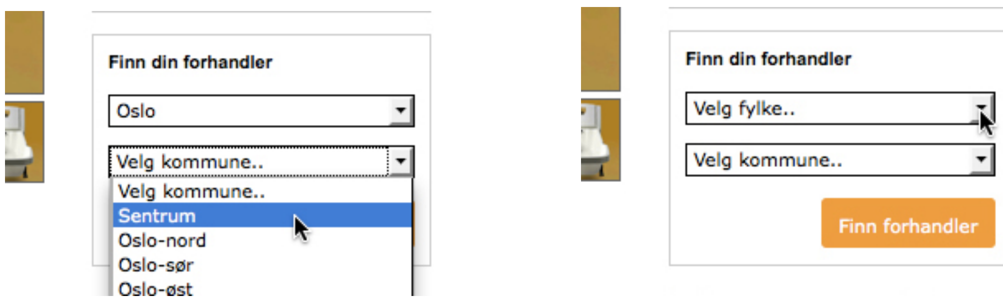

Figure 17 & 18 – Drop-down menus on the product page.

Disabling the functionality in the low-fidelity prototype forced participants to explain how they understood the functionality of this element. The high-fidelity version allowed for explanation, and also observation of how this was used.

Another difference is found in the functionality of tabs. This applied for the tabs on both the home- and product pages. Whilst participants using the high-fidelity prototype were able to navigate to the pages and content connected to tabs, seen in figure 19, this was not possible with the low-fidelity prototype, shown in figure 20.

Figure 19 & 20 – The functionality of tabs.

In addition to the change in cursor identifying that it was possible to interact with the element, the tabs' borders also changed when the cursor was placed over them in the high-fidelity version. Also, tabs that were open were displayed without a bottom border in the high-fidelity prototype.

Level of visual refinement

As to ensure that the results were not affected by differences in the visual refinement, the two prototypes were made to look identical. Following specifications developed in an earlier NetLife Research project, both prototypes were built with a high level of visual refinement.

Using CSS, the different elements were modified according to specifications. However, as some components used in building the prototypes proved difficult to style in the desired manner using CSS, both prototypes deviated somewhat from the design specifications. Nevertheless, these elements looked identical in both prototypes, ensuring that no differences in user feedback were based on the graphical details.

Breadth of functionality

Some differences between the two prototypes can be identified in this dimension due to the deactivation of some features in the low-fidelity version. In the prototype with high-fidelity richness of interactivity, every action performed by the participant resulted in a reaction from the system, whilst several of the features were disabled in the low-fidelity prototype. This resulted in a minor difference between the two with regards to the breadth of functionality. These differences are found when comparing the home- and product pages from the two prototypes. As tabs were disabled in the low-fidelity prototype, participants who tested the prototype with a high level of interaction had a somewhat broader field to navigate. The high-fidelity prototype consisted of 51 separate pages participants could navigate to, whilst the low-fidelity offered 47 due to the deactivated tabs.

Depth of functionality

In both prototypes, the exact same sequences were represented to their conclusion. Three product pages were created allowing participants to reach the bottom of the navigational tree in three different ways. However, due to restrictions in available time to produce the prototypes, several hyperlinks only allowed users to navigate one step down in the hierarchy before they encountered a page that politely asked them to navigate backwards. In total the prototype hierarchy consisted of six layers of pages.

Richness of data model

The contents of the prototypes were gathered from the previous work done by NetLife Research as well as from visits to the existing Handicare website. As the report from the former Handicare project was in English, the contents were translated somewhat freely to Norwegian as the prototypes were tailored for the Norwegian part of the Handicare website. Some category headlines were copied off the existing website, as well as all the product information for the three product pages that were made.

Due to requests made by Handicare, the prototypes are not publicly available.

4.3 Testing the prototypes

Following the DECIDE framework (described in 3.2, page 25), the usability tests were planned and carried out for the two prototypes.

4.3.1 Planning and preparing

In order to carry out the usability testing, much planning and preparing had to be done in order to ensure a smooth implementation of the tests.

Overall goal

The overall goal for these tests was to uncover any possible differences in the participants' ability to provide feedback for the two prototypes. As such, it was important to be able to uncover the expectations and understandings users had of the functionality.

A separate goal was also to uncover the usability issues found in this design, to provide Handicare with valuable feedback on usability for future work.

Usability test tasks

The tasks used in testing were based on the actions users were expected to perform when visiting the pages, e.g. general navigation on the pages, locating specific information about certain products, and finding the closest dealer.

To be able to answer the research questions posed in 1.2, and also provide data about the usability of the new Handicare website design, the tasks were constructed to gather and identify any differences in participants' expectations and understanding of the design.

In total, there were 10 tasks for participants to complete. In addition, the screening and debriefing sessions consisted of 13 questions. Although most of the questions posed during debriefing were meant to sum up, one final task was presented here. Participants were presented with the opposite prototype and asked to locate a product they had studied in a previous task. This was included in order to be able to observe any changes in expectations or performance. It also helped exploring whether or not the expectations participants using the low-fidelity version had to functionality were correct.

The questions were tailored so that the participants were halted, and their expectations to the different features on pages were collected, before they were allowed to navigate on the page.

The complete usability test plan, in Norwegian, is included in the appendix of this thesis.

Paradigm and techniques

The evaluation paradigm used for this project was usability testing. For gathering of data, all the techniques described in 3.2.1 were used.

Practical issues

Usability test team

As the lab used, described later, contained the necessary equipment to record test data, the team ended up consisting of only two persons. One would act as an observer, taking notes during the tests. This was also the team member who evaluated the recorded data when testing was completed. The other member of the team would possess the role of the moderator, presenting the participants with the tasks and accompanying them throughout the duration of the test. The moderator would also question the participants on issues noticed during testing, and perform screening/welcoming- and debrief interviews.

To try to avoid any biased data it was decided that the author of this thesis should inhabit the role of the observer, hereby removing any chance to steer the participant in any way and biasing the outcome of the tests. However, there is also a possibility of ending up with biased data when only having one person go through the collected data, analyzing and evaluating the outcome. This is further discussed in chapter 7.1 Limitations.

Although observation of the testing allowed for first hand experience of how the new design performed, and was something that was encouraged, representatives from Handicare Norge were only able to observe 2 of the 16 sessions.

Recruiting participants

The recruitment of participants for the usability testing was done using a set of criteria defining the qualities they should possess. These criteria were based on results from a previous project NetLife Research had done for Handicare. The complete participant recruitment criteria, in Norwegian, are included in the appendix of this thesis.

Using the personas created in the previous NetLife/Handicare project, three segments of test participants were created. Common for all categories was the age span, from 18 to 65. An even distribution of participants with regards to gender, age and Internet skills was also listed as recruitment demands for all groups. The Internet skills desired ranged from a medium to high skill level. Segment one consisted of people with family members or relatives with mobility problems, not living in the same residence. Segment two was constrained to people with disabled family members or relatives who were living in the same residence. The third segment consisted of people whose mobility was limited due to disabilities.

However, as it proved difficult to recruit participants from segment one, it was decided to change this to people with no relation to mobility issues, but who still fulfilled the requirements in regards to age, gender and Internet skills. Some uncertainties about whether or not to include actual users of Handicare products as participants were also considered, but in the end it was decided that including these users would be a good reference for future work on further adaption of the website to this user group.

The company Norstat[5] handled the recruitment of suitable participants for these tests. The specified criteria were sent to them, providing them with constraints to follow throughout the process.

In total, 16 participants were recruited for these tests. In addition, two participants were recruited as stand-ins in case some of the others did not show up. A complete list of participants is included in the appendix.

The participants were divided into two groups, trying to balance the number of people from each segment, gender and age evenly. One half performed the test tasks using the prototype with high-fidelity interaction, whilst the other half tested the prototype with low-fidelity interaction. This was done to prevent learning effects in having the same people test both prototypes.

The group testing the prototype with high-fidelity richness of interaction consisted of four men and four women, ranging from 26 to 61 years of age. Two participants belonged in the first segment, three in segment two, and the remaining three in segment three. Everyone had experience using the Internet, with an average of about 3 hours online a day. Each participant spent from 0,5 to 8,5 hours using the Internet every day. All eight used Internet banking and had experience purchasing merchandise using web pages. One half also used their cell phones to go online.

[5] For information about the data collection company Norstat, visit: http://www.norstat.no/no/

Three men and five women tested the prototype with low-fidelity richness of interactivity. Out of the eight there were 3 persons from the first segment, 3 from the second, and 2 participants from the third segment. Spanning from 18 to 63 years of age they each spent from 0,5 to 8 hours a day surfing the Internet, averaging about 2,7 hours. Out of the eight, all of them used Internet banking, and had experience with online shopping. Only one participant used her cell phone to go online.

Each participant was presented with an incentive upon completion of the test. This incentive was given in the form of a gift certificate, worth 500 Norwegian kroner. As all the recruited participants showed up for testing at the scheduled time, there was no need for using the extras. Nevertheless, these were also given the incentive.

Location

The usability lab at NetLife Researches offices in Oslo was used for these tests. Situated in downtown Oslo, the location is easily accessible for participants and observers. Permanently set up to cater for usability testing, it proved the ideal location for this purpose.

The lab set up consists of two adjacent, but physically separate rooms, keeping the participant separated from the observers, eliminating the risk of them having any effect on the results. Set up with video cameras, microphones and software that allows for screen capture, both audio and video is transmitted from one room to the other. This kind of usability lab set up is referred to as an *electronic observation room setup* (Rubin and Chisnell, 2008).

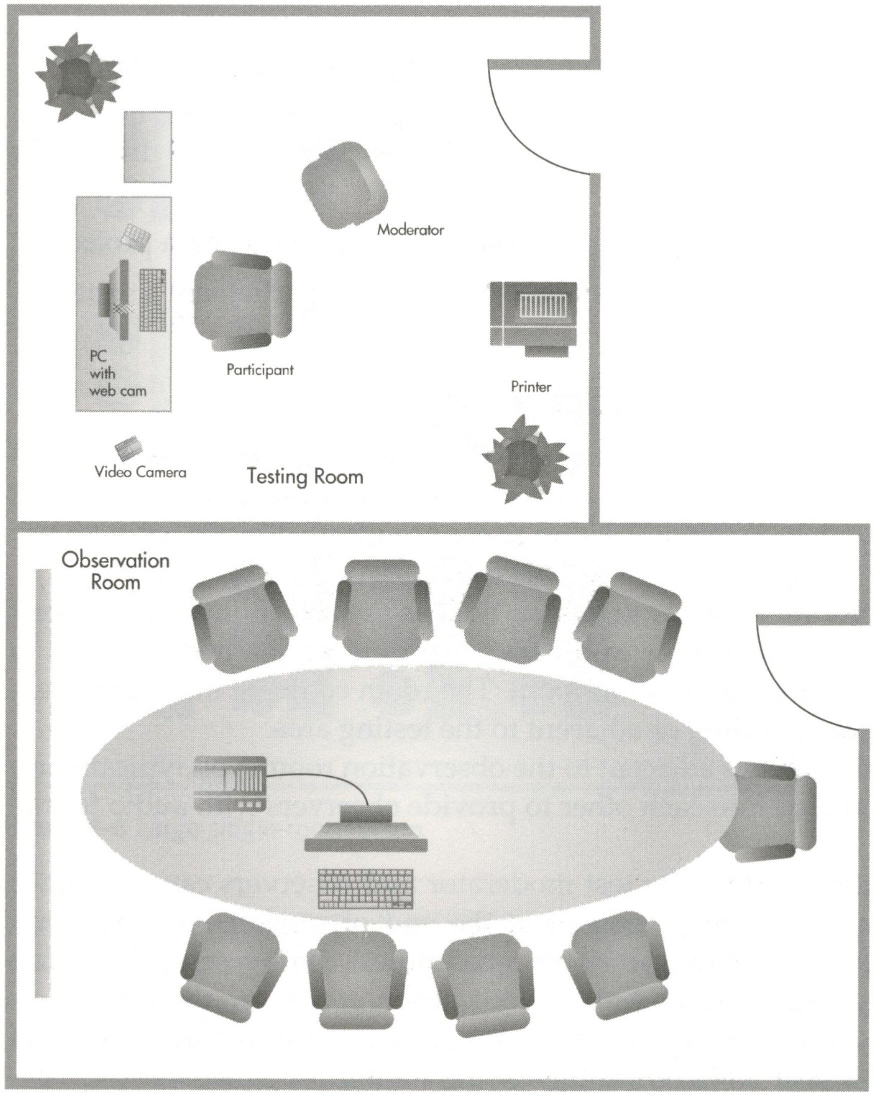

Figure 21 - The set up of an electronic usability test lab. (Preece et al., 2002, pg 108).

In the test room, a camera is mounted in the ceiling, capturing the participant and the moderator during testing. Two computer screens are situated in the room, one for the moderator to use whilst the other is for the participant. The participants screen is also equipped with eye tracking technology, allowing tracing and recording of the participants' eye movements and focus point throughout the test. The room is also equipped with a microphone, recording the conversation between the participant and the moderator. The picture below shows the test room at the NetLife Research office.

Picture 1 – The test room.

The observer room consists of a table situated in the middle of the room. On one of the walls, a projection of the participants screen is shown so that observers can follow their navigation. When eye tracking is used to gather data, observers can also follow the eye movements of the participant on this screen, represented by a red dot moving across it. Next to the projection, a television set is showing an overview of the room where the moderator and participant are located. This allows for the observers to see how participants behave when performing the tasks. There is also a sound system placed in the room, allowing observers to follow the conversation between the moderator and the participant. The picture below shows the observer room at the NetLife Research office.

Picture 2 – The observation room.

Pilot testing

A pilot test was carried out to identify possible issues in the tasks or execution of the testing. Utilizing a NetLife employee, this session helped avoid unforeseen problems with the prototype or tasks during testing, and also identified issues with the software and hardware that was used.

It was discovered that Protoshare required the use of Mozilla Firefox to correctly visualize the alterations in element styles done using CSS. Therefore, an alteration in the software set up had to be made. As the eye tracker, Tobii[6], only worked with Internet Explorer when gathering data of the participants' gaze, the solution ended up being to run Mozilla Firefox in combination with the screen-capturing tool Morae[7]. This tool made it possible to record the participants' point of gaze, but would not support generation of heat maps. Nevertheless, as the production of heat maps require a higher number of participants to provide reliable data, the decision to carry on without these was made.

Ethical issues

Every participant was asked to sign an informed consent form when testing began. It provided information on what was being tested, how recordings were made and that all recorded data would be kept confidential and not tied to any personal data. This was done to ensure the rights of those participating in these tests.

The complete informed consent form can be found in the appendix of this thesis.

4.3.2 Testing the prototypes

The actual testing of the two prototypes was performed in four days, over a period of two weeks. To make analysis of gathered data as easy as possible, testing of the high-fidelity richness of interactivity prototype was completed the first week and the low-fidelity richness of interactivity prototype the second week.

[6] For more information on the eye-tracking equipment, see: http://www.tobii.com/corporate/start.aspx
[7] The Morae software is described in further detail on their website: http://www.techsmith.com/morae.asp

Welcome/Screening

As described in chapter 3.2.2, the welcoming/screening allows for a brief insight into participants background, skills and experience allowing the test team to cross reference with the recruitment criteria. Upon arriving, the participants were greeted by the moderator, who made an effort to release the tension many participants experience when taking part in usability tests.

Some participants were reliant of a wheelchair to get around, and as the building had no equipment for getting wheelchairs in through the main entrance, these had to be led in through the parking garage. Following these participants in and up the elevator presented a chance to investigate their background briefly.

As participants entered the test room, they were informed of their rights when participating in the tests, and asked to sign an informed consent. When this was completed, the moderator started a recording of the participants' computer screen, sound, and a web camera capturing the participants' facial expressions.

To get a brief understanding of the participants' skill level in using the Internet, some screening questions were asked. Questions about prior knowledge of Handicare and their website were also posed. This allowed for a cross reference to the recruitment criteria to ensure that they matched the established user profile.

Testing

When the initial questions were completed, participants were presented with the first task they were to solve. The moderator allowed the participants to navigate freely, offering clues and pointers only when the participant expressed signs of wanting to give up. As participants completed one task, they were presented with the next. Whilst performing them, the moderator urged the participants to think aloud.

Tasks were presented to the participants one at the time, and they were allowed plenty of time to complete them. Following a semi-structured interview guide, the moderator was able to question the participants about interesting observations during testing, rather than having to do this in the debriefing when participants might not have been able to provide any good answers.

When starting the test on the home page, and upon reaching the virtual bathroom and the product pages, the moderator would stop the participants and ask them to explain how the different elements on the page functioned. The participant would then describe, without using the mouse, what they thought was possible to interact with and their expectations as to what would happen.

Although having performed a pilot test to rid the test set up of any problems, small changes were made to the tasks during testing so that they better catered for delivering usable data in relation to the research questions presented earlier in this thesis.

Debriefing

In this part of the testing, the participants were asked about their overall impression of the pages, how they felt navigation worked and other questions about the look and feel of the design.

The participants were also presented with a final task during the debriefing. Prior to this task, the moderator would switch prototypes so that the ones testing the high-fidelity richness of interactivity prototype would now be using the low-fidelity one, and vice versa. Asking the participants to navigate to a product page visited in an earlier task allowed for observations of changes in navigation. Also, participants were asked whether they noticed any differences on the page since the last visit, and which design they favored of the two.

4.3.3 Analysis of usability test data

Upon finishing the testing, the gathered data was analyzed. Observational notes recorded during testing were reviewed along with the captured video. Both reviewing the captured video and analysis of the observational notes had to be done several times, based on the focus of review.

One review was done in order to collect data forming a basis for a report to be delivered to Handicare. This report was focused on the usability issues discovered during testing, and included no information about differences in outcome for the two prototypes. Due to requests made by Handicare, this report is not publicly available.

Keeping the research questions in mind, analysis of the recorded video was done again, this time focusing on discovering differences in navigation, users' expectations and other details that could help uncover whether or not the differences in fidelity of richness of interactivity had any effect on the outcome of usability testing.

As differences in participants' experiences with the two prototypes were expected to be more prominent in some parts of the tests, a selection of tasks to focus the analysis on was made. Focus was placed on the tasks involving the prototypes' home-, product- and virtual bathroom pages.

The choice to focus the attention to these pages was based on the fact that pages differed the most in terms of interaction. As such, it was supposed that this would allow for easier recognition of differences or similarities for the two prototypes.

When the review of the recorded video and observer notes was completed, the task of compiling them and making summaries was performed. Aggregating all the responses and observed actions into two separate summaries allowed for comparison of the individual tasks, and made discovering possible differences or similarities easier. Through revision of these along with the recordings, findings that could help shed light to the research questions posed in 1.2 were observed.

5 Results

This chapter presents the results gathered through analysis of data from the usability testing. Chapter 5.1 presents the findings made during usability testing of the high-fidelity richness of interactivity prototype, whilst chapter 5.2 is a description of observations and findings made during testing of the low-fidelity prototype.

As the usability of the design is not the focus of this research, all identified usability issues are not described in detail here. The participants' navigation, expectations and thoughts using the prototypes are much more important, and to provide answers to the research questions only the issues related to these are described. The review of gathered data with this focus in mind revealed some interesting findings.

5.1 Prototype with high-fidelity richness of interactivity

The observations from the usability testing of the high-fidelity prototype provided a benchmark for participants' knowledge, understanding and expectations to the functionality of the design. Through extracting important data from these sessions a comparison with the low-fidelity results was possible.

5.1.1 Familiarity

A majority of the participants used the text hyperlinks to navigate to other pages in the prototype. Most participants were able to identify and use these quickly, using the differences in color, underline or other form of decoration as clues. One participant stated that: "*color changes is something I am used to. When the color changes, the hyperlink is active*" (translated). Another participant stated that he was "*used to hyperlinks having underlines*" (translated). All the participants from this group were able to describe the functionality of these. In addition, a few of the participants also used the images as hyperlinks on several occasions. This was functionality that was only present in the high-fidelity prototype.

Also the hover effects such as changes in color or text decoration used in the high-fidelity prototype was something several participants seemed familiar with. Some stated that they identified elements that were interactive by using these effects. One participant also stated "*I use the mouse to look*" (translated).

In contrast, it was observed that although the home page's category links had large areas highlighted using a hover effect, and clicking anywhere inside the highlighted field would suffice (the functionality is described in 4.2.2 – Figure 7), several participants carefully placed the cursor over the text in these prior to clicking.

Also when encountering the virtual bathroom, many participants became unsure of the functionality of the page. Some described it as modern and new, whilst two specifically commented on not having seen this before. Not having encountered this feature anywhere, participants found it hard to explain its functionality.

The features found on the product page were also familiar to most participants. Having seen the use of images in similar fashion elsewhere, they could quickly identify and describe how this functionality worked (explained in 4.2.2 – Figure 15). Several stated that this was functionality they recognized from other web sites.

All the participants in the high-fidelity tests also understood the tabs, and how they functioned. On questioned on where they would find different information, users navigated quickly and easily to the correct tab. Only one tab seemed to cause problems for some users. The tab *Tilleggsfunksjoner* was sometimes mistaken for containing supplemental products. In fact, this tab was intended to give a presentation on different extra features offered for some products. However, this is a problem related to labeling or content, as the navigation via tabs was unproblematic.

All participants were able to explain how the drop-down menus worked. Without using the feature, they stated that a list of alternatives would appear when it was clicked and that they could select one of these.

5.1.2 Expectations

For most of the functionality found on these pages, the participants' expectations were quite accurate to the functionality found in the prototype. As mentioned, participants quickly identified the hyperlinks and tabs on the home page. Expecting to be able to interact with these they provided explanations as to what they thought would happen. Several participants also expected the images that were displayed on the pages to function as hyperlinks.

However, several participants failed to identify the headers *Mobilitet*, *Sikkerhet & håndtering* and *Sykehusutsyr* on the home page as hyperlinks. Appearing as plain headings on the page due to their color, as described on page 46 (figure 5 and 6), they did not demand attention as hyperlinks regardless of the inherent hover effects. As such, participants did not expect there to be any interaction with these. Observations show that several participants struggled to identify and use these as hyperlinks.

Being asked to describe how they thought the virtual bathroom worked showed that participants had several different expectations to its functionality. The virtual bathroom acted as an alternate form of navigation, showing a list of products associated with the area or item in the bathroom. Each button would display a box containing a list of product categories associated with that area or feature upon interaction. Clicking the hyperlinks presented in these boxes would bring the participant to another page, identical to the result of using the menu on the left hand side. Although some descriptions matched the pages functionality, others were less sure and came up with suggestions that were far from the functionality found on this page. One participant said, "*it allows me to design a functional bathroom*" (translated), suggesting that it was possible to move items around and change the layout of the room.

Participants expected to be able to interact with the highlighted areas of text in the virtual bathroom, i.e. the buttons placed here. However, they were not certain as to how this interaction would play out. Several expected that they had to click on these for something to happen. Many were also observed doing this on their first encounter with the virtual bathroom, even though this happened without clicking. Other participants did not click on their first encounter, but would later do this when navigating the room. One participant stated, "*I did not think about it, I just clicked because it looks like a button*" (translated). Several participants later commented on that they liked the hover functionality, and that they thought this would make it easier for users to navigate the virtual room.

There also existed some uncertainties as to the functionality of the hyperlinks in the info boxes in the virtual bathroom. Participants were unsure as to exactly what would happen when clicking the links that appeared in the info boxes. Some thought this would result in a presentation of images whilst others expected a presentation of product accessories. A few correctly expected it to be just another method of navigating the products offered by Handicare.

Most participants described the functionality of the images on the product page correctly. They assumed that it was possible to click on the smaller images, and that they would become enlarged.

All the participants expected the drop-down menus to display a list of alternatives when they interacted with these. Expecting to be able to click on the different tabs, they assumed that new information or pages would be displayed.

5.1.3 Switching fidelity

When being presented with the low-fidelity richness of interaction prototype and told to solve a task using this, 5 of the 8 participants commented on the missing functionality. Although starting from the virtual room, participants did not seem to notice the change in fidelity prior to reaching the product page and trying to interact with features here. As their interaction did not provide them with the results they expected, some participants thought the system was lagging whilst others simply remarked that this feature had been clickable earlier in the tests. The remaining participants did not comment on this lack of functionality, but observations indicate that they noticed the difference. Some participants seemed to react to there being no change in cursor display when placing this over an image.

One participant stated that it is easier to understand that there is something more to a feature if there is some kind of feedback, but that she would still understand that text with a different color or decoration were hyperlinks. Another also noticed that there were no changes in color or other visual clues as to hyperlinks being interactive, but commented that "*you can see that they are interactive by the change in cursor, from an arrow to a hand, so I would still click on them*" (translated). Several of the 8 participants commented that they preferred the high-fidelity version.

5.2 Prototype with low-fidelity richness of interactivity

The gathering of data from tests utilizing the low-fidelity interaction prototype allowed for comparison of results. As with the high-fidelity recordings, participants' expectations, previous knowledge, comparisons, and reactions were studied.

5.2.1 Familiarity

When questioned about the functionality on the pages, most of the participants were quickly able to identify the features they could interact with. One participant even described the pages as quite standard, commenting, "*there was nothing revolutionizing about this*" (translated). Stating that text hyperlinks were identified by their color and decoration, participants revealed having prior knowledge to the functionality of these. One person also remarked that it would have been nice to be able to use the images as hyperlinks after attempting to do so during one of the tasks. It was later revealed that this person was familiar with this kind of interaction from other websites.

The division of content in tabs was also a feature most of the participants recognized. It was easy for them to explain their inherent functionality. They were generally able to identify where different content would be placed for most cases, although some difficulties were observed in regards to the previously mentioned tab *Tilleggsfunksjoner*. Participants quickly identified and explained the functionality of drop-down menus. However, only one person commented on this being familiar.

Several of the participants were also familiar with the functionality of the images on the product pages. Only one participant had never encountered this feature previously. Her comments on this feature were therefore aimed at the images themselves rather than their functionality. Commenting on the small size of the four images placed underneath the main image, she stated that it was difficult to see details.

Being familiar with most features used on the home- and product pages in the prototype, the virtual bathroom was something of a contrast for the participants. Not having seen this feature elsewhere resulted in several participants being unable to describe its functionality. This was also noticeable in one participant who took a long time to start interacting with the design.

5.2.2 Expectations

Being able to identify hyperlinks, tabs, drop-down menus and image functionality through previous knowledge, most of the participants were correct in their expectations as to how the different features on the pages would function. Nevertheless, some differed significantly from the possible outcome.

Explaining that visual clues such as color and decoration identified hyperlinks, participants expected to be able to interact with page elements that stood out. In addition, several mentioned using the cursor to identify interactive elements on the pages.

Although only one person commented on the lack of hyperlink functionality in images, several others attempted to navigate the pages by clicking on these expecting to be brought to a new page. As nothing happened when they tried using these for navigation, participants were observed trying to click again before they used the text as hyperlinks. Several also paused significantly with the cursor placed over images before moving it and clicking the text to navigate.

Participants seemed particular in placing the cursor on top of the text prior to clicking when using text as hyperlinks. This was also the case for the category hyperlinks on the home page, that allowed the user to click anywhere inside the square. For the low-fidelity prototype, the only visual clue that a larger area was clickable was the change in cursor visualization. The functionality is described in 4.2.2 – figure 8.

Whilst a few participants were able to identify that the main category links *Mobilitet*, *Håndtering & sikkerhet* and *Sykehusutstyr* on the home page were in fact hyperlinks, the majority failed to identify these as such. This was later confirmed when these users failed to use the hyperlinks when trying to solve one of the tasks. These hyperlinks were not differentiated in color, and did consequently not stand out in the design.

When encountering the virtual bathroom, participants became uncertain as to what functionality they expected from the design. The majority quickly identified the different buttons displayed in the image of the bathroom, and expected some kind of reaction when clicking on these. It should be noted that upon transferring the participant to this page in the prototype, the prototyping tool often displayed content without the participant having to interact.

One participant expected that some information would be displayed or that some virtual action would happen within the picture. As an example, she expected the shower door to open when the button *Dusj* was clicked. Upon actually clicking the button, she seemed surprised, and said, "*the image was very detailed, and so I expected something interactive to happen, perhaps an animation of some sort*" (translated). Other participants also expected some kind of information being presented, but most were somewhat uncertain as to exactly what would happen. One participant, the youngest of the 16, understood the functionality thought implemented immediately when encountering the page. Several were also quite unsure about the functionality of the information boxes in the virtual bathroom.

Seven of the participants expected to be able to interact with the images on the product page. They all knew this feature from previous encounters on other sites, and assumed that the small images placed below the bigger one would be magnified and displayed in the larger square when interacted with. One person also expected this to happen when the cursor was placed on top of the small images without having to click on them. He stated that they would be displayed when hovering over them, but that they had to be clicked in order to remain enlarged when the cursor was moved. Only two persons remarked that this functionality was missing.

Also the tabs on both the home- and product pages were disabled. Most participants expected these tabs to display information when they were clicked, but only one person commented on this lack of functionality. One participant also expected the contents of the tabs to appear on the far right hand side of the screen when clicked. This was not the intended functionality for this feature and quite strange, as this person later stated having experience in using tabs for navigation.

Participants assumed they could click on the drop-down menus, and that these would present them with a list of options to choose from. One participant thought that he would have to manually type in the information in these.

5.2.3 Analogies

Throughout the tests, several participants mentioned that they had seen similar functionality on other websites.

When encountering the virtual bathroom, two of the participants immediately associated it with the IKEA Home Planner[8] software. One participant commented on the resemblance, whilst another thought that it was possible to "*select parts from somewhere and try them out in the room. Try sizes and such*" (translated). Both expected to be able to arrange and plan the layout of the room.

[8] The IKEA Home Planner is a tool allowing users to drag and drop furniture into a room, and try out different furniture configurations. The GUI of this software is fairly similar to the virtual bathroom found in the prototypes. It is available online at: http://www.ikea.com/ms/no_NO/rooms_ideas/splashplanners.html

During testing, three participants also commented on the image functionality on the product pages being quite similar to the one found on the online marketing service Finn.no[9]. One of these three also referred to eBay[10] as having similar functionality.

5.2.4 Switching fidelity

As with the group testing the high-fidelity prototype, the participants using the low-fidelity version were also given one task to solve using the opposite prototype.

Presenting the participants with the high-fidelity richness of interactivity prototype, few mentioned the difference without being asked. One participant commented on that the images on the product page were now possible to interact with, whilst another commented on the possibility to use a larger field as hyperlinks on the home page. However, this functionality was also present in the low-fidelity richness of interactivity prototype only without the visualization of this through hover effects.

The remaining participants did not comment on the extended interaction although they all used one or more features they had not been able to interact with in the previous prototype e.g. clicking on images to navigate. When questioned about the visual feedback the high-fidelity prototype provided, a majority of the participants stated that this would improve the design, but that it did not matter that much to them. One mentioned that, "*it could perhaps be useful*" (translated). Some also stated that they were able to navigate more quickly when using the high-fidelity richness of interactivity prototype, but that they had not paid any attention to what was causing this.

When asked about the functionality of the features in this prototype, most participants explained that it matched the expectations they had to the features in the low-fidelity prototype. However, there were two exceptions. The contents of the tabs on the product page were not presented the way one participant thought. He expected these to show up

[9] Finn.no is a Norwegian online marketing place. Visited by more than 4.1 million users each month, it is the largest of its kind in Norway. Visit Finn.no at: http://www.finn.no/

[10] eBay.com is the world's largest online marketplace. Practically anyone can buy and sell practically anything. With more than 90 million active users globally, eBay is a familiar site for many Internet users. Found at: http://www.ebay.com/

on the far right, which they did not. Neither did the drop-down menus require manual input of data as was assumed by the same participant.

6 Discussion

In this chapter, the research questions and results from the usability testing are addressed. Chapter 6.1 discusses the findings made in regards to how richness of interactivity in prototypes influences participants' ability to provide feedback. In 6.2, the question of how the findings from this research will potentially influence the future process of prototyping and usability testing is discussed. The prototyping tool is briefly evaluated in chapter 6.3.

6.1 Influence on participants' ability to provide feedback

In line with previous research comparing high- and low-fidelity prototypes (Catani and Biers, 1998, Virzi et al., 1996, Walker et al., 2002), this study also found there to be little difference in the results from the two prototypes that were tested. Findings suggest that the level of interaction in features that are *familiar* to participants has little effect on their understanding of these. Accordingly, *participants are able to provide feedback on these regardless of the fidelity of interaction in the prototype*. It would seem there is a greater need for a high level of interaction in *unfamiliar* features to provide users with the information they need to construct the correct mental models. *Without an existing mental model of the functionality in features, participants are less able to comment on these and more dependent on a clear and precise system image*. Comparing data from the two groups of participants, there are several findings supporting these views.

6.1.1 Familiar features

Research has shown that users who have prior experience with a feature or functionality are able to understand, and use these, quicker than the ones without this experience (Blackler et al., 2007b). In relation to this, findings from this study suggest that users who are familiar with a feature are less dependent of a high level of interaction to be able to provide feedback. Jonassen and Henning (1996) state that people make use of their existing mental models from similar situations to form an understanding when encountering novel features. But when people are familiar with a feature's system image, they have already constructed the correct mental model of how it works, and are not dependent on additional information to understand.

This was seen during the usability testing, where all the users identified most of the text hyperlinks on the pages, regardless of prototype fidelity. Both groups expected to be able to click on these, and that they would lead to a different page. The use of a different color or decoration on text hyperlinks is common practice to maximize the affordance of these and increase their clickability (Nielsen, 2004). By using these clues participants easily identified the hyperlinks, and could utilize their existing mental models to understand the functionality. Although the level of interaction was different in the two prototypes, and the high-fidelity prototype had more visual feedback on participants' actions, there was little difference in the ability to provide feedback on these hyperlinks.

Familiarity with hyperlinks may also explain why several participants failed to identify the main category links *Mobilitet*, *Håndtering & sikkerhet* and *Sykehusutstyr* as such. As described earlier in this thesis (figure 5 & 6, pg 46), these hyperlinks were black by default, and would change color and become decorated with an underline on mouseover in the high-fidelity prototype. The only noticeable difference to these in the low-fidelity version was the change in cursor visualization. Nevertheless, several participants from *both* groups failed to indentify these as hyperlinks and consequently could not comment on their functionality, suggesting that designers should be particular in differentiating text hyperlinks using a different color than body text in prototypes. Furthermore, it would seem that the level of interaction in these text hyperlinks did not have any effect on the participants' ability to provide feedback on the design, and consequently is something interaction designers should not spend valuable time and resources developing.

The observations made of participants using the category links on the home page (figure 7 & 8, pg 46) supports this view. In the high-fidelity version, a large area was highlighted to indicate a hyperlink when the cursor was placed over these. The low-fidelity version had the same size area hyperlink, but no highlighting. The fact that several of the users from both the high-fidelity and low-fidelity groups carefully placed the cursor over the text prior to clicking, indicates that the level of interaction is irrelevant as their mental models of how hyperlinks work makes them ignore the visual clues. Although claiming to use the change in cursor visualization to identify interactive features, both groups disregarded this quite often. The high-fidelity group also overlooked the color highlighting. As such, this extra interaction did not cause any differences in the participants' understanding, expectations or ability to comment on the design feature. As they are familiar with the appearance of hyperlinks and have existing mental models of their functionality, participants will place the cursor on top of the text prior to clicking. Therefore, there seems to be no need for a higher level of interaction in the hyperlinks used in prototypes.

On a separate note, as participants testing the high-fidelity prototype stated that hyperlinks changing color and decoration style on mouseover was something they were familiar with and used to, this increased interaction is a positive feature in regards to the usability of the design. The low-fidelity group also commented on this when they were presented with the high-fidelity prototype. Even though they all felt this visual feedback was something that would improve the usability of the design, it was not important for them in order to understand the functionality of these hyperlinks. Although it is not vital to ensure results from usability testing, such functionality should be implemented in the finished design to ensure a site that is usable, and consequently successful (Nielsen, 2000).

The level of interaction in the *image feature* on the product pages was also different in the two prototypes (figure 15 & 16, pg 49). Still, there were few observed differences in the participants' ability to understand and comment on this functionality. Both groups of participants were able to describe how this functionality worked through having previous experience with similar features. The majority expected to be able to browse the images by clicking on them. Participants from the low-fidelity group confirmed their familiarity with this feature through their referrals to sites using images in a similar fashion, such as FINN.no and eBay.com. Recognizing this feature from other sites on the Internet they can use previously established mental models of functionality to understand (Namahn, 2010). Therefore, they are able to comment on these regardless of the level of interaction.

This further supports the view that a high level of interaction is not required in features that are familiar to participants in order for them to provide feedback.

However, that a feature is *familiar* to a participant does not entail that they have the *correct* mental model of how this works. Describing the functionality of the images found on the product pages, one of the participants from the low-fidelity group stated having prior knowledge of this functionality. Accordingly, he expected the images to change when the cursor was placed over them, even if he did not click on these. He expected that clicking an image would result in this being the default image in the larger square. This shows that although familiarity with a feature enables participants to provide feedback regardless of the level of interaction, their mental model of the functionality might not be identical to that of the designer.

In this case the lack of interaction may have caused the participant to mentally compensate for the missing functionality, using an existing mental model and imagining a more advanced interaction. Such mental compensation for lack of detail has already been seen in other studies (Sauer and Sonderegger, 2009), and is a possible pitfall because users envision a different functionality than what is actually available and base their comments on this.

This *deficiency compensation* is something interaction designers should be aware of if they are to employ prototypes with a low fidelity of interaction in usability testing. As users might compensate for missing functionality, this could potentially result in invalid feedback. On the other hand, interaction designers might be able to turn this into something positive. As the features that are commented on are of a low fidelity, designers will be able to incorporate new ideas gathered from participants without having to spend a lot of time and resources to make changes in the design. As a consequence, the user will have a greater influence on the finished design, contributing to a more user-centered design process.

Similar to the image functionality, the *tabs* participants encountered on the home- and product pages (figure 19 & 20, pg 50) did not reveal any difference in the ability to provide feedback for the two groups. Although the tabs were disabled in the low-fidelity prototype, all of the participants expected to be able to click on these, and most were also able to describe their functionality. As all the participants who engaged in these tests were users of Internet banking and also had experience with online shopping, it is highly probable that participants were familiar with this functionality. Knowing how these elements will function in the finished design, they do not need to try their functionality in order to understand how they will work. In other words, when participants recognize features from previous encounters, they can use their existing mental models of how that feature works. As such, they are able to provide feedback on these, regardless of the level of interaction.

Also the *drop-down menus* on the product pages (figure 17 & 18, pg 50) were something the participants from both groups were familiar with. Accordingly, the different level of interaction in the two prototypes had little influence on the ability to provide feedback also for this feature. Even though the drop-down menus were disabled in the low-fidelity prototype, participants using this version had few difficulties explaining the functionality inherent in these.

From the examples above, it would seem that high-fidelity interaction might not be necessary in features that are familiar to users. Having seen and used a feature, participants inhere an existing mental model of how it will perform. Consequently, they are not dependent on experiencing the functionality in order to provide feedback. One possible explanation as to why several of the participants were familiar with the features used in the prototypes might be that building blocks such as drop-down menus, tabs and images placed in certain ways are features frequently used both in software as well as on the World Wide Web (WWW).

There has been rapid expansion in the number of personal computers, personal digital assistants (PDA) and smartphones such as the iPhone throughout the last couple of years. With an increase in the number of Internet users from 360 million in year 2000 to more than 1,8 billion at the end of 2009 (Miniwatts Marketing Group, 2010), the Internet has become a public domain. The number of users and pages on the WWW has increased exponentially, giving everyone from small children to senior citizens experience in the functionality of commonly used web elements. This has equipped many users with mental models of how certain features work (Norman, 2002), removing the need for designers to present a detailed system image to the user.

But for users to be able to make use of their previous experiences, they are reliant of being able to recognize the different features. Several of the examples suggest that users are able to recognize and understand features based on their appearance. It is possible that a decrease in the level of interaction inherent in features increases the need for a higher fidelity of aesthetical refinement. A low level of interaction in combination with a low aesthetical refinement may cause participants in usability tests to become uncertain as to whether or not they are in fact employing the correct mental model. This is a question that can be addressed if further research is performed on this topic.

In contrast, findings also point to that new and unfamiliar features have to be of a high fidelity in order for participants to create the correct mental model, and consequently be able to provide feedback on these.

6.1.2 Unfamiliar features

As described earlier in this thesis (chapter 1.3, pg 8), users mental models are created through combining stored information from previous experiences with information available through the look and feel of a design (Morris, 2002). These mental models allow for a prediction of the effects caused by an action, but without a good model it is impossible to understand what will happen (Norman, 2002). As such, encountering novel features or functionality it is impossible to know what to do, resulting in a need for a *more detailed system image* to gain a better understanding. Being faced with unfamiliar features participants look to prior experiences that are in some way similar for understanding (Hill, 2009, Jonassen and Henning, 1996).

This was observed when participants encountered the virtual bathroom (figure 9, pg 47) for the first time. Several participants from both groups stated that this was something they had not seen before, and were unfamiliar with. Not being sure as to what to expect from this, they started comparing the feature to others they had previously experienced in order to understand how they worked. Participants who tested the low-fidelity prototype compared the virtual bathroom with the IKEA Home Planner software, expecting there to be similar functionality in this prototype. Although none of the participants from the high-fidelity group made such analogies, some still expected to be able to move elements around and make changes to the layout of the room in a similar fashion as in the IKEA software. Even if there was no gathered data on whether or not they had previous experiences with this, their expectations indicate that they were familiar with the functionality found in this software.

In this case, their previous experiences caused participants to create invalid mental models of the functionality inherent in the feature. There seems to be few differences in the expectations the two groups had to the virtual bathroom. The uncertainty they experienced when facing new and unfamiliar features and functionality limited *both* groups from being able to provide feedback on this. Therefore, it would seem there is a need for a higher level of interaction in features and functionality that is unfamiliar to participants. By providing a more detailed system image, designers enable interaction with the interface, and consequently provide the participants with the information they need to develop the correct mental model.

Further support for this view is found in another example from the virtual bathroom. One participant based her expectations on the high level of detail in the background image on this page, and expected an animation to take place when interacting with the buttons. Not having an existing mental model of how this feature worked, she made use of the knowledge that was available to her in the world. Basing her understanding on this limited system image, she created an invalid mental model of the functionality and was consequently unable to provide relevant feedback on this feature. As such, this supports the view that a higher level of interaction has to be implemented in features that users are unfamiliar with and do not have existing mental models for.

In addition to a high level of interaction, designers should also incorporate a high level of *visibility* and *affordance* (Norman, 2002) into unfamiliar features, inviting participants to explore the functionality and thereby enabling them to create a suitable mental model. The buttons on the virtual bathroom page is a good example of this. As described earlier, participants from both groups became quite unsure as to the functionality found on this page, and their expectations differed widely from the how the feature actually worked. Nevertheless, all the participants were able to identify the buttons, although somewhat uncertain as to what would happen when these were clicked. As the buttons in both prototypes "invited" the participants to interact, they were able to create a more suitable mental model through interaction with the system image.

Similarly, users from both groups were unsure as to the functionality in the information boxes that appeared when they clicked the buttons. Also for this feature both prototypes allowed for interaction, and provided participants with the information they needed to create the correct mental models of functionality. Without this opportunity to explore the feature, through interaction with the system, participants would not have been able to provide feedback on this.

As described in the previous chapter, most participants from both groups were familiar with, and could therefore provide feedback on, the image functionality found on the product pages. However, one participant from the low-fidelity group failed to understand the image functionality (figure 15 & 16, pg 49), commenting on the size of the smaller images and how they made it difficult to look at details. Later commenting on this being an unfamiliar feature to her, although spending 6 to 8 hours online every day, this supports the view that users who are unfamiliar with functionality cannot understand or use these as intuitively as those with prior experience (Blackler et al., 2007b). Not being able to draw conclusions about the functionality from previous knowledge, or having the ability to experience it first hand, this participants' feedback was fairly irrelevant to the design. Therefore, to enable participants without an existing mental model of a feature to provide feedback, it would seem there is a need for a high fidelity of interaction in order for them to understand the functionality.

Further supporting this view, one participant from the low-fidelity group expected the contents of the tabs on the product page (figure 19 & 20, pg 50) to appear next to the information that was already being displayed. Even though he claimed having seen and used this functionality before, his mental model of how it worked was not correct. For that reason, he was unable to provide relevant feedback on this feature.

Also for the drop-down menus on the product pages (figure 17 & 18, pg 50) this same participant had a somewhat different expectation than the others. Although stating he was familiar with this feature, he thought he would have to manually input the data rather than selecting an option from a list. Accordingly, his mental model of the functionality in the feature was incorrect.

As this participant stated having prior knowledge of these features, this might suggest that it can be difficult for some participants to employ the correct mental model when a feature is lacking interaction. Another possible explanation to this may be that the participant confused the drop-down menu with a text input field, as they are quite similar in appearance. Furthermore, his level of experience in using the web might explain why he had difficulties explaining these. Spending 45 minutes a day online, he is below the Norwegian daily average of 73 minutes (Vaage, 2010).

The discussion above points to that when participants are unfamiliar with a feature or functionality, they are less able to provide feedback if prototypes with a low fidelity interaction is used. Therefore, interaction designers should provide users with a clear and precise system image for these features, inviting the user to interact and form a correct mental model. It also suggests that if such prototypes are used, it is vital that interaction designers are familiar with the participants experience and knowledge of features used in the prototype, i.e. the level of “interaction competence” in the target audiences.

Although these findings might enable interaction designers to adjust the fidelity of interaction in prototypes based on participants’ familiarity with these, there are still some considerations to be made. The implications these findings have on the process of prototyping and usability testing are discussed in the following chapter.

6.2 Implications for prototyping and usability testing

The discussion in chapter 6.1 showed that the level of interaction in prototypes could have an effect on participants' ability to provide feedback on design, depending on their inherent mental models of the prototypes features and functionality. It seems unnecessary to implement high-fidelity interaction in the features participants are already familiar with, but in contrast they require a higher level of interaction in functionality they have not seen before. To enable development of correct mental models of the functionality, and consequently be able to provide designers with relevant feedback, users are reliant on being presented with the complete functionality and chain of events, i.e. a detailed system image. These findings can certainly have an effect on the future of prototyping and usability testing.

Choosing the correct level of fidelity in prototypes is essential to produce results that can be used further in the iterative process (Olsen, 2005). Still, the available resources have often been a basis for this decision. Badre (2002) stated that the main reason for building high-fidelity prototypes, is to test the flow and ease of navigation in a design. But as shown in the discussion above, it would seem that the level of interaction in features that are familiar to participants does not necessarily have to be of a high fidelity in order for them to understand the functionality in these, and consequently be able to provide feedback. The findings from this study can help interaction designers distribute their time and resources more wisely, and consequently enable them to make decisions based on needs rather than available funds.

The ability to explore ideas and experience how users interact with a design without spending much resources, is one of the best reasons for prototyping (Berkun, 2000). Using mixed fidelity prototypes, it is possible for interaction designers to apply resources more precise, and where they are needed, to support the end goals (McCurdy et al., 2006). Knowing that users are familiar with a feature and its functionality, it is possible for interaction designers to spend less time and money developing this. Instead they can focus on development of new and unfamiliar forms of interaction, where participants do not have existing mental models of the functionality and are reliant on more knowledge in the world to be able to develop them. Elements that are not that commonly used, e.g. the virtual bathroom, may require more exploration to understand.

Being able to prototype features using low-fidelity interaction is an advantage for interaction designers. As described earlier (chapter 2.1.1), the low-fidelity prototypes are not only cheaper and faster to produce, but are also easier to make changes to, and even discard completely if they do not work as they are supposed to. Being able to resolve to a low-fidelity interaction for many features also decreases the amount of work needed to develop them, and consequently removes the resistance many developers have towards making changes to a high-fidelity design they have spent a lot of time building (Rettig, 1994).

Another possible effect is that development projects can reduce the risk of running out of resources, having to cut short or even abandon some steps of the process. When a lack of resources occurs in a design process it is often the focus on usability that suffers (Gulliksen et al., 2006), and as such it is possible that these findings can help ensure that more projects are in fact able to work with the usability of their design, resulting in a product that is better for the users.

The ability to experience real users interacting with a design is one of the main reasons for doing usability testing (Badre, 2002, Dumas and Redish, 1999, Hoekman Jr., 2009, Tognazzini, 2000), and will provide interaction designers with data for the next step of the design process. Through recordings of observational data combined with comments from the participants, interaction designers are able to create an understanding of the users experiences and focus on these in further development. But in removing the interaction from features in a design, you also remove the participants' ability to interact with these and consequently the ability to make observations of how they use these. As such, the only way to get feedback on these features is to have participants comment on how they normally interact with these. This could present a major pitfall as there is often a difference in what people *say* they do, and what they *actually* do (Blomberg et al., 1993). Accordingly, it is important for interaction designers to be aware of this when utilizing low-fidelity interaction prototypes in testing, and not rely blindly on the verbal feedback provided by participants.

Changes also have to be made in the process of planning and preparing for usability testing if interaction designers are to enjoy the benefits of the findings made throughout this research. The process of recruiting participants has to be modified, taking into account that people have to be familiar with the features that are presented with a low fidelity of interaction. It is no longer enough to recruit participants based on that they represent the typical users of the design, as these might not be familiar with the different features that are used in the prototypes, and consequently might not be able to provide feedback.

Nor will it be enough to base the participants' knowledge on the amount of time they spend online, as they may still be unfamiliar with features and functionality. An example of this is the participant who had no prior knowledge of the image functionality on the product page (discussed in chapter 6.1), although she stated that she spent 6 to 8 hours online every day. It is possible that individuals who spend a lot of time online have come across several different features and functionalities, but assuming that they are familiar with the features used in a prototype may result in them being incapable of providing feedback.

Findings also show that although participants claim to be familiar with features, their mental models might not match the actual functionality. The participant who expected the contents of a tab to open next to the information already being displayed, or having to manually type in information in the drop-down menus, is a good example of this (described in 6.1). Although stating that he was familiar with these features, his expectations did not match the actual functionality of these. As such, for interaction designers to base recruitment on participants' own statements about which features they are familiar with may result in users that are unable to provide relevant feedback on the design.

To enable testing of prototypes with a low fidelity of interaction, the criteria for recruitment of participants have to examine user backgrounds even more thoroughly to be certain that they have existing mental models of the features and functionality that is used. Having an existing mental model of a feature enables participants to understand and use functionality intuitively (Blackler et al., 2007b), and in turn allows them to understand and comment on these. Recruiting people without an existing mental model of how a feature works may therefore result in them being unable to provide feedback.

Participants encountering the virtual bathroom support this view, as they were unable to understand the functionality of the feature just by looking at it, and consequently could not provide any feedback on this. Participants from both groups understood that there was some kind of functionality inherent here, but were quite uncertain as to what they could expect. In this case, they were reliant on being able to interact with the system to understand its functionality. Accordingly, it is vital that interaction designers have knowledge of participants existing mental models to ensure that they are capable of providing feedback on the design. Getting this information will involve a lot of work for the designers (Blackler et al., 2007a), and is far from easy.

One way of doing this could be to include more screening questions in the recruitment process to help uncover prior knowledge. A possible approach would be to only recruit participants who are users of certain websites. Getting an understanding of which sites the participants are familiar with, designers can cross-reference these with their own designs, making sure that the features and functionality found on these match those used in the prototypes.

An effect of only recruiting users that are familiar with certain features for usability testing is that you only get to test your site on experienced users. While experienced users may uncover a higher number of usability issues in a design, novices find more serious problems (Sauer et al., 2008). As such, in utilizing only experienced users, one is liable to overlook issues that might cause serious problems for new users of the design. Therefore, including at least one person with little or no skills, but who would still potentially use your system, in the participant group would allow for detection of fundamental errors in the design (Rubin and Chisnell, 2008).

6.3 About the prototyping tool

As described in 4.2.1, the two prototypes were built using the application Protoshare.

Building the basic wireframes and adding the desired functionality to the prototypes was fairly easy using this tool. Even though it offers a functional solution, it is not without issues and hiccups.

The possibility to add *styles* to the different elements is a valuable feature in concept, but getting it to work as desired proved difficult. As most of the building blocks available in the application are rather low-fidelity aesthetically, the opportunity to add CSS to the design was an important feature. Nevertheless, this proved to be one of the pitfalls of Protoshare. As all the elements have predefined styles, located in one of four extensive stylesheets, finding the correct bit of code to alter was a strenuous task. However, through email correspondence with Protoshare support it was pointed out that the application is meant as a means to quickly and easily put together a mock-up of the design, giving it interactive capabilities. Therefore, applying a high level of visual refinement was a laborious task.

The possibility for communication amongst members of the design team and clients within the tool was previously mentioned. It was thought that this functionality would allow for most of the communication in regards to design details to be handled during development of the prototypes, rather than when the prototype was finished and ready for testing. However, as the client was occupied with other projects at the time, this feature was never fully tested. Nevertheless, it is probable that this feature could help interaction designers reduce the resources spent on development by allowing the clients to participate more actively in the design process. This would allow for resolving minor design issues early in the process when they are easy to correct, rather than later in the process where it might be to late to do alterations to the design.

One major drawback was discovered during the course of this research. On numerous occasions, content meant hidden for participants unless they interacted with elements on the page was displayed when pages were loaded. This is further discussed in chapter 7.1.

7 Conclusion

This thesis has explored the implications of richness in interaction fidelity in prototypes used for usability testing. Two research questions were presented in the introduction.

The first research question was concerned with how differences in the fidelity of interaction in prototypes affect participants in usability testing:

1. In what way does the level of interaction in prototypes influence the participants' ability to provide feedback in usability testing?

It was found that the participants' prior knowledge and experiences are closely related to the influences caused by the level of interaction in prototypes.

Results show that in features and functionality that are *familiar* to the participants, the level of interaction in prototypes has *little effect on the ability to provide feedback*. This due to the fact that they already have an inherent mental model of the functionality in these features, and are not reliant on having to interact with them to understand their purpose and effects.

In contrast, the findings also show that participants are more dependent on having a high-fidelity interaction in features and functionality they are *not familiar* with in order for them to understand and be able to comment on these. As they do not have existing mental models of how these work, they form wrong expectations and theories based on experiences from similar designs. As such, they require more input from the design on how it functions in order to develop the correct mental models of its functionality. This can be done through increasing the level of interaction in these elements, presenting the participants with a clear and precise system image they can explore.

This thesis therefore concludes that the level of interaction in prototypes has little effect on participants' ability to provide feedback on familiar features, whilst it is a major influence when encountering unfamiliar features and functionality.

The aim of the second research question was to investigate the implications of these findings:

2. *How do the findings from this research influence the process of prototyping and usability testing?*

Results show that interaction designers do not have to build every single feature in a prototype with a high level of interaction fidelity, but that they can *focus their efforts and available resources on the unfamiliar features and functionality* they want to test. But in using this newfound knowledge, interaction designers base their analysis on participants' verbal statements rather than actual observations of their actions. This is a possible source of error, and is something to be aware of when choosing to apply this in a design process.

The findings also points to that there is a need for *stricter recruitment criteria* when testing prototypes with a low fidelity of interaction. Designers have to ensure both that users are in the target audience and representative of the typical user, and also that they are in fact familiar with the features that are low-fidelity in the design. The possible influence on results caused by participants' level of experience is also something that should be taken into account when recruiting for usability testing of prototypes with a low level of interaction.

This thesis concludes that these findings enable interaction designers to focus their prototyping efforts and resources on new and unfamiliar features, provided they take the necessary precautions in planning and carrying out usability testing.

7.1 Limitations of this study

There are several limitations that have to be considered when looking at the results presented in this thesis.

Possibly the biggest one is the fact that only one person performed the recording of data, and subsequent analysis. Only getting one perspective on the gathered data puts it at risk of being biased. Results from usability testing are highly subjective, and are affected by the researcher's background, experience level and perception. It is possible that a preconception of what was going to be found already existed when data was gathered and analyzed, and that this may have influenced the results. Although trying to remain objective there is the danger of only having observed the things one is hoping to find, or emphasizing the wrong findings. Similar studies have used a much higher number of evaluators to ensure a correct interpretation of data (Walker et al., 2002), and others state that it is highly questionable to use a think-aloud study with only one evaluator as an authoritative statement (Hertzum and Jacobsen, 2003). Truly impossible to eliminate completely, this *evaluator effect* can be reduced by utilizing multiple evaluators (Hertzum and Jacobsen, 2003). Nevertheless, usability issues identified, and the severity of these, has been shown to vary also for usability teams due to differences in method and use of resources, regardless of the number of evaluators used (Molich and Dumas, 2008, Molich et al., 2004).

The number of participants in this study is also an issue that needs to be addressed. Due to limitations in available resources, this study did not allow for more participants. Although six participants may be sufficient to keep the iterative process going (Molich, 2010), a much higher number is needed to uncover all usability issues in a design. It is possible that the results from this study might have been different if a larger number of participants were recruited for testing. Previous studies comparing high- and low-fidelity prototypes have employed a higher number of participants (50-60) (Sauer and Sonderegger, 2009, Wiklund et al., 1992), but there are also examples of studies utilizing a similar sized group (15 participants) (Lim et al., 2006). Although the number of participants in this study may have been too low to make solid conclusions about the influence caused by the level of interaction, it has revealed a tendency and formed a basis for further research on the area.

The participants' level of Internet experience might also have been a limitation. They were all experienced users of the web, resulting in that they quickly understood the functionality intended for the finished design, regardless of which prototype they used. As discussed previously, the results gathered through testing pointed to that the level of interaction in elements can be adjusted based on the participants level of experience and prior knowledge of such functionality. However, not having a control group of participants to measure against, there is no way of being certain that this is in fact a valid conclusion. As such, the validity of this research can be questioned.

There is also the question of whether or not the difference in richness of interaction in the two prototypes was in fact large enough. This is especially the case for the virtual bathroom pages. Although the high-fidelity version had more advanced interaction, the low-fidelity also allowed participants to experience the functionality of the feature. It is therefore possible that a further reduction of interaction on this page in the low fidelity version could have resulted in better data.

The moderator's performance during testing can influence the participants (Mayhew, 1999), and may also have had an effect on the collected data in this study. Although remaining fairly consistent throughout the course of usability testing, there were noticeable differences in both how the tasks were presented to the participants and the amount of guidance provided. This is especially visible in the recorded data when participants are presented with the opposite prototype. As the participants were presented with the task, the moderator sometimes provided them with guidance on where to click in order to find the page they were looking for. As such, participants had limited time and opportunity to actually notice the difference in level of interaction. Although some participants remarked on the changes, it is possible that it would have been noticed by additional participants were they given the freedom to explore the opposite prototype more.

The prototyping tool used in this research was also the source of a possible limitation. When participants navigated between pages during testing, the prototyping tool sometimes displayed items that were supposed to be hidden unless a button or other element was interacted with. This was especially noticeable on the virtual bathroom page, where these incidents might have given the participants ideas as to the functionality on this page. As the prototyping tool showed them content upon loading the page, this may have given them further clues as to how the interaction on this page worked. As such, it is possible that these participants might have responded differently to the questions posed by the moderator if they had not been exposed to the content prematurely. Consequently, this may have had an effect on the results and observations made during this study, and might also have influenced the conclusions.

7.2 Future work

Although the results gathered throughout this study, and the analysis of these, suggest a lesser need for high-fidelity interaction in familiar features, more work is needed to confirm the findings. Further research on the subject would provide valuable data. By including the factors described below in research following up on this study, it may be possible to increase the accuracy of the gathered results.

To reduce the *evaluator effect*, additional studies on the subject should include two or more evaluators. As this study was carried out using only one evaluator, the possibility of having biased results is definitely present. By including more people in the evaluation process it will be possible to increase the validity of the results.

Furthermore, the number of participants should be increased to be certain that the findings are universally applicable. As the sample size used in this study was quite small, there is no way to be certain that the findings made here are valid. Also, this study was lacking a control group. Adding a number of people with little or no skills in using the Internet could possibly provide conclusive data as to whether or not familiarity is in fact a determinant factor for the required level of interaction in prototypes. Although finding such participants in the fast paced, technological world we inhabit might be a difficult task.

Earlier research has shown that aesthetics has a significant influence on the human mind in usability testing (Lavie and Tractinsky, 2004, Sonderegger and Sauer, 2010). As such, it would be interesting to find out to what degree the prototypes' level of visual refinement influenced the results from this research. A possible approach to this could be to create four prototypes for usability testing. Having one pair with a high level of visual refinement, and one pair with a contrasting low-level refinement would allow for comparison and evaluation of the effect of this dimension on the ability to provide feedback. Each pair should have one version with a high level of interaction, and one version where the interaction was limited. This would allow for exploration of the importance of aesthetical refinement and its relation to interaction fidelity in prototypes.

References

BADRE, A. (2002) *Shaping Web usability: interaction design in context,* Boston, Addison-Wesley.

BARNUM, C. M. (2002) *Usability testing and research,* New York, Longman.

BEELDERS, T. R., BLIGNAUT, P., MCDONALD, T. & DEDNAM, E. (2008) Novice Word Processor User Performance with Pictorial and Text Icons. *Proceedings of the 8th Asia-Pacific conference on Computer-Human Interaction.* Seoul, Korea, Springer-Verlag.

BERKUN, S. (2000) #12 – The art of UI prototyping. [Online] Avaliable: http://www.scottberkun.com/essays/12-the-art-of-ui-prototyping/ [Accessed: 3/2-2010]

BLACKLER, A. L. (2009) Applications of high and low fidelity prototypes in researching intuitive interaction. *Proceedings of the Design Research Society Conference 2008.* Sheffield Hallam University, Sheffield.

BLACKLER, A. L., POPOVIC, V. & MAHAR, D. P. (2003a) Designing for Intuitive Use of Products: An Investigation.

BLACKLER, A. L., POPOVIC, V. & MAHAR, D. P. (2003b) The nature of intuitive use of products : an experimental approach. *Design Studies,* vol. 24**,** 491-506.

BLACKLER, A. L., POPOVIC, V. & MAHAR, D. P. (2004) Studies of Intuitive Use Employing Observation and Concurrent Protocol. *Design 2004 8th International Design Conference.* Dubrovnik, Croatia.

BLACKLER, A. L., POPOVIC, V. & MAHAR, D. P. (2005a) Intuitive Interaction Applied to Interface Design. *International Design Congress - IASDR 2005.* Douliou, Taiwan.

BLACKLER, A. L., POPOVIC, V. & MAHAR, D. P. (2005b) Intuitive Interaction with Complex Artefacts. *Design Research Society International Conference 2004.* Melbourne, Australia.

BLACKLER, A. L., POPOVIC, V. & MAHAR, D. P. (2007a) Developing and Testing a Methodology for Designing for Intuitive Interaction. *International Association for Societies of Design Research 2007: Emerging Trends in Design Research.* Hong Kong, The Hong Kong Polytechnic University School of Design.

BLACKLER, A. L., POPOVIC, V. & MAHAR, D. P. (2007b) Empirical investigations into intuitive interaction: a summary. *MMI-Interaktiv,* vol. 13**,** 4-24.

BLOMBERG, J., GIACOMI, J., MOSHER, A. & SWENTON-HALL, P. (1993) Ethnographic field methods and their relation to design. IN SCHULER, D. & NAMIOKA, A. (Eds.) *Participatory Design: Principles & Practices.* New Jersey, Lawrence Erlbaum.

BONNER, J. & VAN SCHAIK, P. (2004) The use of high and low level prototyping methods for product user interfaces. IN HANSON, M. (Ed.) *Contemporary Ergonomics 1998.* London, Taylor and Francis.

BOSTON, I. (2009) Transcription of a Recording - Factors Which Influence How Long Will it Take. [Online] Avaliable: http://ezinearticles.com/?Transcription-of-a-Recording---Factors-Which-Influence-How-Long-Will-it-Take&id=3367680 [Accessed: 22/4-2010]

CALABRIA, T. (2004) An introduction to personas and how to create them. [Online] Avaliable: http://www.steptwo.com.au/papers/kmc_personas/index.html [Accessed: 22/4-2010]
CATANI, M. B. & BIERS, D. W. (1998) Usability evaluation and prototype fidelity: Users and usability professionals. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* vol. 2.
CORSINI, R. (2002) *The Dictionary of Psychology,* New York, Brunner-Routledge.
COTTRELL, N. B., WACK, D. L., SEKERAK, G. J. & RITTLE, R. H. (1968) Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of personality and social psychology,* vol. 9**,** 245-250.
DENZIN, N. K. & LINCOLN, Y. S. (2005) *The Sage handbook of qualitative research,* Thousand Oaks, Calif., Sage Publications, Inc.
DIX, A., FINLAY, J., ABOWD, G. D. & BEALE, R. (2004) *Human-computer interaction,* Upper Saddle River, NJ, Pearson.
DUMAS, J. S. & REDISH, J. C. (1999) *A practical guide to usability testing,* Exeter, Intellect LTD.
EIE, T. F. (2010) Gjør det selv: 5 trinn til brukertestet nettsted. [Online] Avaliable: http://www.iallenkelhet.no/gjør-det-selv-5-trinn-til-brukertestet-nettsted [Accessed: 30/3-2010]
GULLIKSEN, J., BOIVIE, I. & GÖRANSSON, B. (2006) Usability professionals-current practices and future development. *Interacting with Computers,* vol. 18**,** 568-600.
HARPER, D. (2010) Online Etymology Dictionary. [Online] Avaliable: http://www.etymonline.com/index.php?term=interaction [Accessed: 28/4-2010]
HERTZUM, M. & JACOBSEN, N. E. (2003) The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International journal of human-computer interaction,* vol. 15**,** 183-204.
HEWETT, T. T., BAECKER, R., CARD, S., CAREY, T., GASEN, J., MANTEI, M., PERLMAN, G., STRONG, G. & VERPLANK, W. (2009) ACM SIGCHI Curricula for Human-Computer Interaction : CHAPTER 2: Human-Computer Interaction. [Online] Avaliable: http://old.sigchi.org/cdg/cdg2.html [Accessed: 28/4-2010]
HILL, B. R. (2009) Mental models. [Online] Avaliable: http://mentalmodelassessment.org/mental-models/ [Accessed: 19/4-2010]
HOEKMAN JR., R. (2009) The Myth of Usability Testing. [Online] Avaliable: http://www.alistapart.com/articles/the-myth-of-usability-testing/ [Accessed: 12/11-2009]
HOUDE, S. & HILL, C. (1997) What Do Prototypes Prototype? IN HELANDER, M., LANDAUER, T. & PRABHU, P. (Eds.) *Handbook of Human-Computer Interaction.* 2nd ed. Amsterdam, Elsevier Science.
JACOBSEN, N. E., HERTZUM, M. & JOHN, B. E. (1998) The evaluator effect in usability studies: Problem detection and severity judgments. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting.* Chicago.
JONASSEN, D. H. & HENNING, P. (1996) Mental models: knowledge in the head and knowledge in the world. *Proceedings of the 1996 international conference on Learning sciences.* Evanston, Illinois, International Society of the Learning Sciences.

LANDAUER, T. K. (1995) *The trouble with computers: usefulness, usability, and productivity,* Cambridge, Mass., The MIT Press.
LAVIE, T. & TRACTINSKY, N. (2004) Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies,* vol. 60**,** 269-298.
LIM, Y.-K., PANGAM, A., PERIYASAMI, S. & ANEJA, S. (2006) Comparative analysis of high- and low-fidelity prototypes for more valid usability evaluations of mobile devices. *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles.* Oslo, Norway, ACM.
MAYHEW, D. J. (1999) *The usability engineering lifecycle: a practitioner's handbook for user interface design*, Morgan Kaufmann Publishers Inc.
MCCURDY, M., CONNORS, C., PYRZAK, G., KANEFSKY, B. & VERA, A. (2006) Breaking the fidelity barrier: an examination of our current characterization of prototypes and an example of a mixed-fidelity success. *Proceedings of the SIGCHI conference on Human Factors in computing systems.* Montreal, Quebec, Canada, ACM.
MINIWATTS MARKETING GROUP (2010) World Internet Usage Statistics News and World Population Stats. [Online] Avaliable: http://www.internetworldstats.com/stats.htm [Accessed: 14/4-2010]
MOLICH, R. (2010) CUE (Comparative Usability Evaluation), English - DialogDesign. [Online] Avaliable: http://www.dialogdesign.dk/CUE.html [Accessed: 8/4-2010]
MOLICH, R. & DUMAS, J. S. (2008) Comparative Usability Evaluation (CUE-4). *Behaviour & Information Technology,* vol. 27.
MOLICH, R., EDE, M., KAASGAARD, K. & KARYUKIN, B. (2004) Comparative Usability Evaluation. *Behaviour & Information Technology,* vol. 23**,** 65-74.
MORRIS, P. E. (2002) Memory: Mental Models. [Online] Avaliable: http://www.answers.com/topic/memory-mental-models [Accessed: 15/4-2010]
NAMAHN (2010) Mental models. [Online] Avaliable: http://www.namahn.com/method/mental-models [Accessed: 18/4-2010]
NEILL, J. (2007) Qualitative versus Quantitative Research: Key Points in a Classic Debate. [Online] Avaliable: http://wilderdom.com/research/QualitativeVersusQuantitativeResearch.html [Accessed: 7/4-2010]
NIELSEN, J. (1993) *Usability engineering,* Boston, Mass., Morgan Kaufmann.
NIELSEN, J. (2000) *Designing web usability: [the practice of simplicity],* Indianapolis, Ind., New Riders Publishing.
NIELSEN, J. (2004) Guidelines for Visualizing Links. [Online] Avaliable: http://www.useit.com/alertbox/20040510.html [Accessed: 31/3-2010]
NIELSEN, J. & MACK, R. L. (1994) *Usability inspection methods,* New York, Wiley.
NORMAN, D. A. (2002) *The design of everyday things,* New York, Basic Books.
OLSEN, H. (2005) Balancing fidelity in prototyping. [Online] Avaliable: http://www.guuui.com/issues/03_05.php [Accessed: 30/3-2010]
OSVOLD, L. (1996) Internett en flopp! *Dagens Næringsliv.* Oslo.
PREECE, J., ROGERS, Y. & SHARP, H. (2002) *Interaction design: beyond human-computer interaction,* New York, John Wiley & Sons, Inc.
RETTIG, M. (1994) Prototyping for tiny fingers. *Commun. ACM,* vol. 37**,** 21-27.
RUBIN, J. & CHISNELL, D. (2008) *Handbook of usability testing: how to plan, design, and conduct effective tests,* Indianapolis, Indiana, Wiley Publishing, Inc.

RUDD, J., STERN, K. & ISENSEE, S. (1996) Low vs. high-fidelity prototyping debate. *interactions,* vol. 3**,** 76-85.
SALVENDY, G. (2006) *Handbook of human factors and ergonomics,* Hoboken, N.J., John Wiley & Sons.
SAUER, J., FRANKE, H. & RUETTINGER, B. (2008) Designing interactive consumer products: Utility of paper prototypes and effectiveness of enhanced control labelling. *Applied Ergonomics,* vol. 39**,** 71-85.
SAUER, J. & SONDEREGGER, A. (2009) The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics,* vol. 40**,** 670-677.
SEFELIN, R., TSCHELIGI, M. & GILLER, V. (2003) Paper prototyping - what is it good for?: a comparison of paper- and computer-based low-fidelity prototyping. *CHI '03 extended abstracts on Human factors in computing systems.* Ft. Lauderdale, Florida, USA, ACM.
SNYDER, C. (2003) *Paper prototyping: the fast and easy way to design and refine user interfaces,* Amsterdam, Morgan Kaufmann Publishers.
SONDEREGGER, A. & SAUER, J. (2010) The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics,* vol. 41**,** 403-410.
TOGNAZZINI, B. (2000) If They Don't Test, Don't Hire Them. [Online] Avaliable: http://www.asktog.com/columns/037TestOrElse.html [Accessed: 21/10-2009]
VAAGE, O. F. (2010) Norsk mediebarometer 2009. Oslo, Statistisk Sentralbyrå.
VIRZI, R. A. (1992) Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 34**,** 457-468.
VIRZI, R. A., SOKOLOV, J. L. & KARIS, D. (1996) Usability problem identification using both low- and high-fidelity prototypes. *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground.* Vancouver, British Columbia, Canada, ACM.
WALKER, M., TAKAYAMA, L. & LANDAY, J. A. (2002) High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes. *Proc. Human Factors and Ergonomics Society 46th Annual Meeting.* Baltimore, USA.
WICKENS, C. D. (2004) *An Introduction to human factors engineering,* Upper Saddle River, N.J., Pearson Prentice Hall.
WIKLUND, M. E., THURROT, C. & DUMAS, J. S. (1992) Does the fidelity of software prototypes affect the perception of usabililty? *Proceedings of the Human Factors Society: 36th Annual Meeting.* Santa Monica, CA, Human Factors and Ergonomcs Society.

Appendix

A - Criteria for recruitment

Om testene

Antall testpersoner	16 totalt - 8 personer til hver test (bør rekruttere en reserve til hver test)
Tidspunkt	Tirsdag 23./Onsdag 24.februar og Tirsdag 3./Onsdag 4.mars 2010 (evt. Torsdag 5.mars 2010)
Hva skal testes	Prototyper av Handicare nettsider
Sted	Storgaten 2, 6.etg, 0155 Oslo , hos NetLife Research XXL- Huset ved Kirkeristen
Hva kan sies til testpersonene?	Som deltaker vil du få tildelt noen oppgaver som du skal løse ved hjelp av en ”*nettside”* mens du tenker høyt. Mens du løser oppgavene vil det sitte en person sammen med deg som stiller oppfølgingsspørsmål.
Hva kan ikke sies til testpersonene?	Det er viktig å IKKE si til den man rekrutterer som reserve at de er reserve. Da har de mye mindre motivasjon for å møte opp.
Premiering	Gavekort på 500 kroner

Rekrutteringskriterier

Generelle kriterier

Kjønn	50/50 kvinner og menn
Alder	18-65 år. Jevn spredning i alder(ønsket fordeling: 4 stykker: 20-29 år, 4 stykker: 30-39 år, 4 stykker: 40-49 år, 4 stykker: 50+ år)
Internetterfaring	Jevn spredning av erfarne og meget erfarne brukere av internett.
Ekskluderings-kriterier	Ingen av deltakerne skal jobbe med webutvikling eller på andre måter ha erfaring med utvikling/design av nettsider eller programvare.

Segment

Antall personer	**Beskrivelse**
6 (2+2)	**Gruppe 1:** Personer uten familiemedlemmer med mobilitetsproblemer, og som ikke har disse problemene selv
6 (2+2)	**Gruppe 2:** Personer med partnere/foreldre som har mobilitetsproblemer og behøver hjelpemidler for å lette hverdagen. Bosatt i samme bolig som personen/e som behøver hjelpemidler
4 (2+2)	**Gruppe 3:** Personer med mobilitetsproblemer
2	Reserve (en reserve for gruppe 1 og en person for gruppe 2)

Med mobilitetsproblemer menes mennesker som behøver hjelpemidler for å kunne bevege seg fritt. Eksempler på dette kan være rullestolbrukere og mennesker som behøver krykker, stokk eller annen støtte (håndtak etc.) for å gå.

Sjekkliste for rekruttering

1. Arbeider eller har du arbeidet med webutvikling eller utvikling/design av nettsider?
    a. Ja – (Deltaker kvalifiserer ikke som testperson)
    b. Nei – (Fortsett)
2. Har du personer med mobilitetsproblemer i din nærmeste omgangskrets/familie, eller har selv problemer med mobilitet?
    a. Ja – (Fortsett)
    b. Nei – (Fortsett)
3. Deltar du i innkjøp/anbefalinger i forbindelse med hjelpemidler til denne /disse personen/e, eller benytter deg av denne typen produkter selv?
    a. Ja – (Fortsett)
    b. Nei – (Fortsett)
4. Hva slags relasjon har du til personen/-e med mobilitetsproblemer?
    a. Foreldre/familiemedlemmer/partner (Fortsett)
    b. Meg selv (Stopp – rekrutteres 4 testpersoner av denne gruppen)
    c. Ingen tilknytning (Stopp – 6 personer fra denne gruppen)

    (Her skal det rekrutteres fra alle kategorier)
5. Er du bosatt i samme bolig som personen med mobilitetsproblemer?
    a. Ja (Rekruttere 6 personer fra denne gruppen)
    b. Nei (Deltaker kvalifiserer ikke)

Datoer og tider for gjennomføring

Tirsdag 23. februar

Nummer	**Tid**	**Profil for testperson**
Testperson 1	16.00 – 16.50	*Kjønn, alder, segment, andre kriterier*
Testperson 2	17.00 – 17.50	
Testperson 3	18.00 – 18.50	
Testperson 4	19.00 – 19.50	

Onsdag 24. februar

Nummer	**Tid**	**Profil for testperson**
Testperson 5	16.00 – 16.50	*Kjønn, alder, segment, andre kriterier*
Testperson 6	17.00 – 17.50	
Testperson 7	18.00 – 18.50	
Testperson 8	19.00 – 19.50	

Tirsdag 2. mars

Nummer	**Tid**	**Profil for testperson**
Testperson 9	16.00 – 16.50	*Kjønn, alder, segment, andre kriterier*
Testperson 10	17.00 – 17.50	
Testperson 11	18.00 – 18.50	
Testperson 12	19.00 – 19.50	

Onsdag 3. mars

Nummer	**Tid**	**Profil for testperson**
Testperson 13	16.00 – 16.50	*Kjønn, alder, segment, andre kriterier*
Testperson 14	17.00 – 17.50	
Testperson 15	18.00 – 18.50	
Testperson 16	19.00 – 19.50	

Torsdag 4. mars

Nummer	**Tid**	**Profil for testperson**
Reserve 1	16.00 – 16.50	*Kjønn, alder, segment, andre kriterier*
Reserve 2	17.00 – 17.50	

Generelle krav til rekrutteringsprosessen

Komplett liste over rekrutterte personer skal sendes kontaktperson hos NetLife Research senest dagen før brukertesten hvis ikke annet er avtalt. Kontaktpersonen bør holdes oppdatert om hvordan det går med rekrutteringen fram til dato for testen.

NetLife Research forutsetter at alle testpersoner har fått informasjon om hvor og når brukertesten skal foregå på e-post (se eget dokument), og at de har svart bekreftende på at de har mulighet til å delta.

Norstat bør sørge for at alle deltakere får en påminnelse om deltakelse på testen dagen før testen skal gjennomføres, gjerne både på e-post og SMS.

Krav til liste over rekrutterte deltakere

Oversikten over rekrutterte personer bør inneholde

- Navn
- Alder
- Stilling
- Telefonnummer, helst mobil slik at man kan sende påminning på SMS
- Hvilket segment personen tilhører
- Svar på spørsmålene i rekrutteringskriteriene
- Svar på spørsmål gitt i forhold til spesielle rekrutteringskriterier

B - Participants

No.	Segment	Age	Gender	Hours/day online	Internet banking	Internet shopping	Internet via cell phone	Prototype version
1	3	55	Male	1	Yes	Yes	No	High-fi
2	3	52	Male	2-3	Yes	Yes	No	High-fi
3	2	61	Male	5	Yes	Yes	Yes	High-fi
4	1	26	Female	2-3	Yes	Yes	Yes	High-fi
5	2	37	Female	2	Yes	Yes	No	High-fi
6	2	59	Female	0,5	Yes	Yes	No	High-fi
7	3	35	Female	2	Yes	Yes	Yes	High-fi
8	1	29	Male	8,5	Yes	Yes	Yes	High-fi
9	1	34	Male	1,5	Yes	Yes	No	Low-fi
10	1	63	Female	6-8	Yes	Yes	No	Low-fi
11	1	33	Female	0,5	Yes	Yes	No	Low-fi
12	2	18	Male	5-7	Yes	Yes	No	Low-fi
13	3	49	Female	2	Yes	Yes	No	Low-fi
14	2	39	Female	1-3	Yes	Yes	No	Low-fi
15	3	56	Mann	0,75	Yes	Yes	No	Low-fi
16	2	46	Female	2	Yes	Yes	Yes	Low-fi
17	1	23	Male	-	-	-	-	Extra
18	2	17	Female	-	-	-	-	Extra

C - Test tasks

OBS! Moderator/observatør merker seg spesielle hendelser underveis, og utforsker disse i debriefing fasen av undersøkelsen.

Start Morae (Shift+F9) **Innledende spørsmål / screening**		
Nr.	**Oppgave**	**Merknad**
1.	Hvor mange timer per uke/(dag) bruker du på Internett? Har du bestilt ting på nett? Bruker du nettbank? Bruker du internett på mobilen?	
2.	Har du noen i familien som er bevegelseshemmet? Kan du fortelle litt rundt dette? Evt. Kan du fortelle litt om det å være bevegelseshemmet?	
3.	Hvilket forhold har du til Handicare? • Kjenner til produkter, kjøpt produkter, hørt om firma?	
4.	Har du besøkt handicare.no før? Hvis ja: • Hvor ofte er du inne? • Er det noe du savner? • Hva synes du om sidene? Hvis nei: • Hva forventer du å finne på nettsidene?	

Informere om at det er en prototype som testes – ikke alt fungerer

Kalibrere Tobii

Testoppgaver – HUSK REF.PUNKT FOR TOBII		
Nr	**Oppgave**	**Merknad**
1.	**Start fra Hjem:** Hva tror du det er mulig å gjøre på dette nettstedet? Hva finnes her? Hva kan du klikke på/hva skjer ved klikk?	
2.	**Start fra Hjem:** Tenk deg at du skal hjelpe en person som har vanskeligheter med å bevege seg med å finne hjelpemidler til baderommet. Kan du finne en oversikt over hvilke deler av badet Handicare tilbyr produkter til? **STOPP TESTPERSON NÅR DENNE NÅR VIRTUELT ROM!!**	
2.1	**Start fra virtuelt rom:** Hva synes du om denne siden? • Hva forventer du det er mulig å klikke på? • Hva tror du er hensikten med denne? • Hva tror du er mulig å gjøre her?	
2.2	**Start fra virtuelt rom:** Hvilke produktkategorier finnes for Dusj? Hva synes du om at det gis videopresentasjon? - Hva slags video tror du dette er?	
2.2	**Start fra Virtuelt rom:** Personen du skal finne hjelpemidler til ønsker seg nytt badekar på badet. Kan du finne en oversikt over hvilke badekar Handicare tilbyr? Hvilket av disse er aktuelt? **På produktsiden:** Hva kan man klikke på her? Hva skjer ved klikk?	

	Hva synes du om informasjonen som oppgis om produktet på produktsiden? **Om produktsiden:** Hva er bredden og lengden på dette badekaret? Kan du finne en bruksanvisning for dette karet? Hva tror du menes med brukererfaringer? Finnes det andre produkter som kan være relevante til badekaret du ser på?	
2.3	**Start fra Virtuelt rom:** Tenk deg at denne personen også behøver hjelpemidler for å komme seg inn og ut av badekaret. Kan du finne en oversikt over de tilgjengelige badeløfterene?	
3.	**Start fra Hjem:** Tenk deg at du har et familiemedlem med multippel sklerose (MS). Kan du finne en side med hjelpemidler for personer som har denne lidelsen? Hva synes du om denne måten å gruppere produkter for mennesker med en bestemt sykdom på?	
4.	**Start fra Hjem:** Kan du finne rullatoren **Volaris S3**?	

5.	**Start fra Volaris S3-siden:** Du bestemmer deg for å kjøpe dette produktet. Hvordan vil du gå frem for å finne en forhandler i nærheten?	

Debrief		
1.	Hva synes du om innholdet på sidene? Mengden info, ord/uttrykk, oversikt over hvor man var?	
2.	Hva synes du om utseendet på sidene?	
3.	Kan du si noe som du synes var spesielt bra? Kan du si noe som du synes var spesielt dårlig?	
4.	Hvordan opplevde du å navigere på sidene? Var det slik du forventet da du først så siden?	
5.	Hva synes du om - lenkene på siden? - infobokser i virtuelt rom? - arkfaner - bilder - etc.?	
6.	Hvor ferdig tror du navigasjonen på nettstedet er? Rangeres på en skala fra 1-5 Fortell litt rundt hva som gjør at du rangerer slik	
7.	Har du forslag til hva som kunne gjort sidene enda bedre?	
8.	**Start fra <u>Low-fi</u> virtuelt rom:** Bruk midten av siden til å finne badekaret du så på tidligere? Legger du merke til noen forskjeller fra i sted?	

	Opplevde du at ting du trodde var mulige å klikke på ikke var det?	
9.	Moderator utforsker andre ting som ble lagt merke til underveis av observatør/moderator	
10.	Annet	

Stopp Morae (Shift+F9)

D - Informed consent

Dine rettigheter som deltaker i undersøkelsen

1. Det er ikke du som testes, men løsningen. Det er derfor **ingen** krav til prestasjoner eller ferdigheter

2. Det er frivillig å delta og du kan når som helst avbryte testen

3. Videobilder av skjermen og testsituasjonen med lyd sendes til et annet rom

4. Din anonymitet sikres ved at ingen personlige data knyttes til notater eller observasjoner som gjøres

5. Du skal behandles høflig og med respekt

Jeg bekrefter at jeg har fått den informasjon som er angitt ovenfor

Oslo

Underskrift