# The Usability Aspects of Innovative Design Concepts on Mobile Phones - with Focus on Messaging

Johan Garnolf & Nils Hallberg

Sony Ericsson

## Abstract

This report contains the result of a Master Thesis in which the usability aspects of innovative software design concepts on mobile phones are studied. More precisely, the study is divided in the following two main parts:

1) *Threaded Messaging*; the investigation of message threading as an innovative way of presenting text messages in a conversation like manner on mobile phones.

2) *Usability versus Aesthetics*; the investigation of how aesthetics of mobile phone Graphical User Interfaces (GUIs) influence peoples' judgements of usability and how people admit different levels of innovation and aesthetics in graphic design of mobile phones.

In the study, three different mobile GUI prototypes were developed, which had different message handling solutions and varying levels of usability and aesthetics. The prototypes were then tested thoroughly in a usability lab using 16 participants from different age groups.

The results elicited from the tests included that Threaded Messaging is a usable concept that most people find to be easy to learn and are interested in to use on their mobile phones in their everyday lives.

Furthermore, the test results corroborated most similar studies in the computer GUI area, concluding that there is a strong correlation between perceived usability and perceived aesthetics even in mobile GUIs, and that this correlation is bi-directional. In other words, a person's apprehension of a mobile GUI's "ease of use" is strongly affected by his/her apprehension of the mobile GUI's aesthetics, and vice versa.

Moreover, the weakness of using traditional usability properties such as heuristics alone to evaluate usability was accentuated. The importance of involving users early in GUI design processes was emphasized, to ensure that 1) the aesthetics of a GUI does not affect the perceived usability in any undesired way and/or 2) the usability of a GUI does not affect the perceived aesthetics in any undesired way.

## Contents

3

# 1      Introduction

This document is the official report of a Master Thesis for the MMS and iMode Messaging unit at Sony Ericsson, in cooperation with the Ergonomics and Aerosol Technology department at the Faculty of Engineering, Lund University, Lund, Sweden.

The purposes of the thesis were:

• To examine new ways to handle and present messages on mobile phones. The far most common manner to present SMS, MMS and email messages on mobile phones today is to place them in *inbox* and *sent items* directories sorted by date. We have developed prototypes where the users can browse through whole conversations made with specific persons. In this way users will get instant knowledge about the background to each incoming message. These prototypes are tested and evaluated in comparison with existing solutions. This part of the study is referred to in this document as *Threaded Messaging*.

• To examine how people from different age groups react to different levels of innovative and unconventional design and how they admit different levels of aesthetics of mobile phone graphical user interfaces. It is also quantified to what extent aesthetics influence people's judgements of usability. This part of the study is referred to in this document as *Usability versus Aesthetics*.

To concretise the purposes with the study, the following problem statements were formulated:

  **Q1**. Are users positive to the concept of threaded messaging?

  **Q2**. Do users find threaded messaging as usable as the inbox/sent items solution?

  **Q3**. Are users willing to switch from inbox/sent items solutions to threaded messaging?

  **Q4**. To what extent does aesthetics influence people's judgements of usability?

  **Q5**. How do people from different age groups react to different levels of innovative graphic design in mobile phones?

The results of the study were yielded from a usability test, where 16 participants got to test and evaluate Java prototypes of menu systems with different message presentation solutions, and varying levels of usability and aesthetics.

## 2 Background

### 2.1 Threaded Messaging

The far most popular way to present SMS, MMS and email messages on mobile phones is to use *Inbox* and *Sent items* folders. Threaded messaging; the idea of presenting incoming and outgoing messages together as conversations, grouped on specific persons, is a new way of presenting messages on mobile phones.

In email messaging, the concept of threaded messages has been around for a while. Google was first to present threaded messaging to a wide audience through Gmail, released April 1 2004. Gmail piles all messages that belong to certain conversations together.  However, if the subject of a message is changed, the message is no longer classified by Gmail to belong to the conversation, and the new message will not be a part of any existing message pile. The main purpose with the threaded messaging is to provide a quick way for users to see what a conversation with a specific individual has been about. To some extent, threaded messaging has been around as long as people have sent email. When answering an email message, the original message text is normally saved in the new message. If the user does not remove the original message text, the original text is sent together with the new text, thus forming a conversation history that enables the conversers to, in a quick way, go back and see how the conversation has evolved.

As mentioned earlier, threaded messaging in mobile phones is a new concept. However, some implementations have recently been released on the market. The Apple iPhone, released in the summer of 2007, provides the threaded messaging feature by presenting SMS conversations with single messages placed in speech bubbles connected to specific senders. The messages from the iPhone holder are in bubbles with white background that emerges from the left side of the screen, and the messages from the other person have green background bubbles that emerge from the right side of the screen.



**Figure 1**
*Apple iPhone Messaging*

Nokia has also released a message threading solution as a downloadable application for their E60 series (Figure 2). The application is called *Conversation* and was released as a beta version Sept. 21, 2007. In Nokia's application, all messages from and to a specific person are presented in a vertical list under a name and an optional picture of the other person. *Nokia Conversation* coincidently has a big resemblance to the prototype *Threaded Messaging Menu* made by us. (Figure 6-8).
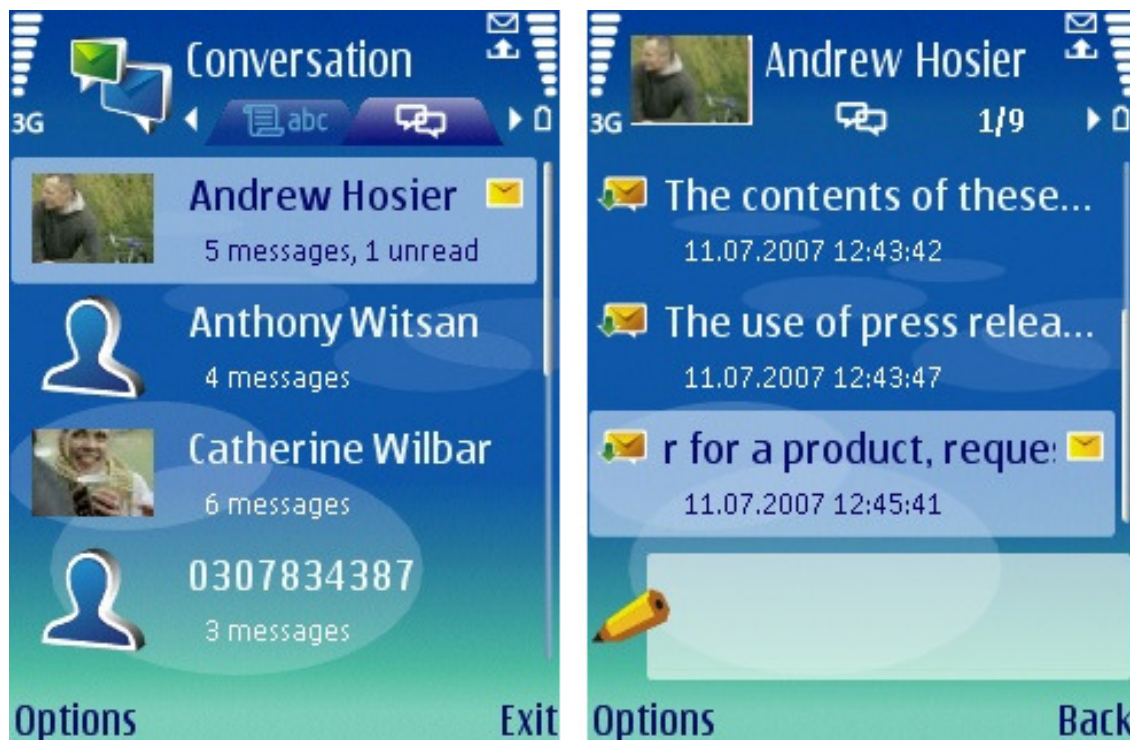
**Figure 2**
*Nokia Conversation*

## *2.2      Usability*

The concept of usability originates from the field of human factors, ergonomics, which has its roots in psychology. The subject arose during World War 2 within the US military, with the purpose of making the weapon systems safe and easy to use. Usability is considered a multidisciplinary field and its importance when applied to Human Computer Interaction, HCI, has been widely accepted through the work of Norman and Nielsen 1984-1993 [1] [2] [4]. The ISO 9241-11 document defines usability as: *The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.* Donahue, 2001, showed that *$1 spent on Usability offers a return of $30.25* [5]. This result really provides a tangible proof of the importance of usability. The benefit, according to Donahue, consists of reduced development and maintenance costs, increased customer satisfaction, lower support costs and improvement of end user productivity.

To get a further description of the concept of usability, the rest of this section will be used to present a number of *design principles*, *usability principles, usability goals* and *user experience goals.* A lot of suggestions of usability guidelines and definitions have been elicited throughout the last decades. We have decided to present the ones originating from the work of Norman, Nielsen and Price, Rogers and Sharp.

1998, Norman coined his design principles [2]. These design principles are intended to help designers to explain and improve the design. They are not intended to be used as a specification to follow when designing an actual interface. Rather, they are meant to act as reminders to the designers, ensuring that certain things are provided at the interface [10]. Note that these design principles are applicable to all interactive products, such as doors, elevators etc, and not just computer interfaces. Norman's design principles are presented below, with brief explanations:

*Visibility:* It should be easy to get an overview of what is possible to achieve with an interface. The vital functions should be easy for the user to find.

*Feedback:* The interface should send acknowledgements to the user when carrying out tasks, so that the user knows and understands what the system is doing.

*Constraints:* The interface should constrain the possible ways to carry out a certain task. I.e. the interface should make users understand that certain actions cannot be made at a certain time, and thus prevent the users from making errors.

*Mapping:* The system controls should be related to their effects in the real world. E.g. an up-arrow is a good mapping for moving something upwards on a screen.

*Consistency:* An interface should have similar operations and similar elements for achieving similar tasks. An example of consistency in the real world can be that all main doors within a building open inwards.

*Affordance:* The objects of an interface should be designed to give a clue of what is possible to do with them. E.g. a door that lacks a door handle gives the clue to people that it needs to be pushed to open; in other words, the door affords pushing.

Jacob Nielsen has recommended 10 usability principles for practical use. The usability principles are quite similar to Norman's design principles. However, the usability principles are more prescriptive, and are better fit to be used as basis for usability evaluation. The usability principles are often referred to as *heuristics*, when used in evaluations. This is discussed further in the next section. Nielsen's usability principles are defined as: [18]

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

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

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

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

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

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

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

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

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

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

Moreover, 2002, Preece, Rogers and Sharp presented the goals of interaction design in terms of usability goals and user experience goals. The usability goals are quite clearly defined and describe what a system is supposed to do: [10]

- *Effectiveness:* The system should be effective to use. I.e. the system should do what it is supposed to do.

- *Efficiency:* The system should provide efficiency so that the users can carry out their tasks with a minimal number of steps.

- *Safety:* The system should prevent the users from making serious errors, and provide the possibility to recover when an error has occurred.

- *Utility:* The system should provide the right kind of functionality so that the users can perform the tasks that they need or want to perform.

- *Learnability*: The system should be easy to learn.

- *Memorability*: It should be easy to remember how to use the system.

The user experience goals are less clearly defined. They describe in general how users should experience a system. Rogers, Preece and Sharp's user experience goals suggests that, part from reaching the usability goals above, a system should also be: satisfying, enjoyable, fun, entertaining, helpful, motivating, aesthetic, creativity supporting, rewarding and emotionally fulfilling [10].

## *2.3      Usability Evaluation Methods*

There are a lot of different techniques that can be used to evaluate the usability of interactive products. The key to be able to develop usable products, in short, is to involve actual users in evaluations early in the process and repeat the usability evaluations iteratively throughout the design process [11]. In the following subsections, the usability evaluation methods that are used in this study are presented.

### 2.3.1      Heuristic Evaluation

The heuristic evaluation is a predictive type of usability evaluation, typically conducted early in the development process. In this kind of evaluation, the users do not need to participate. Instead, persons with great knowledge about usability, human-computer interaction, and the product domain conduct the evaluation by analyzing the products guided by heuristics.   The heuristics are often some kind of usability principles; Nielsen's 10 usability principles, mentioned earlier in the usability section, are commonly used as heuristics. The advantage with heuristic evaluations is that they can be very quick, and thus inexpensive, since no user involvement is required. Moreover, Preece, Rogers and Sharp imply that around 75% of the total usability problems can be found by using only five evaluators. Documented disadvantages with heuristic evaluation include that all usability problems might not be found. Additionally, the

11

evaluators may find irrelevant problems, i.e. problems that, if ignored, do not turn out to be real problems for the actual users. [10]

### 2.3.2      Usability Testing

The main idea with usability testing is to observe users while carrying out tasks on products. Usability tests are usually conducted on a small number of persons, typically 8-16 participants. The collected data can be both qualitative, through comments, open-answer questionnaires et cetera, and quantitative through for example completion times and questionnaire check box questions. Usability tests can be conducted in different types of environments and under varying types of conditions. They can vary from field studies where users are observed performing natural everyday tasks in their natural environment, to very controlled conditions in laboratory environments. A disadvantage of conducting usability tests in controlled laboratory environments is that the participants may feel uneasy. Cameras, one-way mirror and so forth, can make people feel nervous, invoking them to actions that they normally would not take. However, there are advantages as well. One advantage is that it is fairly easy to conduct the tests in the exact same manner with each and every one of the participants, with minimal or no outside interference. In addition to this, it is possible to get detailed video recordings that are good for analysis.

The goal with usability testing is typically to ensure that a product meets the usability requirements. The results can be used as benchmarks for future products. It can also reduce support costs, yield a higher customer satisfaction and sale rate et cetera.

In this study, however, a usability test is conducted with purpose to discover patterns of user behaviours. The method has its roots in classic empirical controlled environment research. However, it is not possible with usability testing to retrieve results that contribute to the research in the area. The reason for this is, for example, that research contributions have very strict requirements to achieve credibility. For example, there must be a specific hypotheses formulation, the participants must be chosen randomly using systematic methods and the sample of participants must be of sufficient size to achieve statistically significance and avoid erroneous conclusions [11]. It is possible though, to see patterns of user preferences by combining quantitative data such as completion times and check-box answers with qualitative data such as user comments during the test.

Rubin, 1996, stated six basic elements that can be used as basis when developing and conducting usability tests: [11]

- *Development of problem statements or test objectives rather than hypotheses.*

- *Use of a representative sample of end users which may or may not be randomly chosen.*

- *Representation of the actual work environment.*

- *Observation of end users who either use or review a representation of the product. Controlled and sometimes extensive interrogation and probing the participants by the test monitor.* (The test monitor is referred to as *test leader* in the usability test of this study)

- *Collection of quantitative and qualitative performance and preference measures.*

- *Recommendation of improvements to the design of the products.*

The usability test in this study was designed with strong influence of the elements above. The exact test method of this study is presented in the *Usability Test* section later in this document.

## 2.4    Usability versus Aesthetics

### 2.4.1    Aesthetics

The concept of aesthetics or beauty is seemingly impossible to give a publicly accepted definition. Since ancient time, people have discussed this issue thoroughly and sometimes violently without settling the issue. Professor of Philosophy, D.W. Prall, 1932, wrote in the foreword to E.F Carritt's "Philosophies of Beauty" that: *"Aesthetics is in fact only a pseudo-science or pseudo-philosophy, a study that no self-respecting member of an academic faculty can safely devote himself to exclusively, or even mainly"* [6]. Norman blames the agreement problem on people's difference in terminologies and backgrounds [3]. He draws parallels with English physicist and novelist C.P. Snow's "The two cultures" [7]. Snow describes the breakdown of conversation between the culture of humanity and literature on one side and the culture of science on the other, which according to Norman is applicable to the debate of what beauty really is. In the 1990s, a lot of textbooks and articles about HCI completely ignored the term of aesthetics. Nielsen, 1993, defined usability goals in HCI as *learnability*, *efficiency*, *memorability*, *error recovery and prevention*, and *subjectory satisfaction* [4]. The aesthetics, he hid in the term of subjectory satisfaction. Even if there are exceptions to this, Tractinsky, 1997, claimed that the majority of the HCI community neglects the term of aesthetics [8]. However, as mentioned in the *Usability* section, Nielsen later defined the 10 principles of usability with aesthetics and minimalistic design as one of them, and Preece, Rogers, Sharp, 2002, defined the more subjective user experience goals, which embody aesthetics [10].

### 2.4.2     Correlation between Usability and Aesthetics

1972, Dion et al found that people that are physically attractive are assumed by other people to posses more desirable personality than people that are physically unattractive [12]. Does this apply also to graphical user interfaces? Do people, before trying to use a system, think that it is more usable if it's pretty? A short glance at the system's aesthetics gives the user a first impression that may bias the attitude towards the system in terms of usability. Is there perhaps a difference between how people apprehend the system's aesthetics and usability before and after the actual usage of the system? Overbeeke, Djjadiningrat, Hummels and Wensveen, 2000, studied research project and thesis results at a department of Delft University of Technology and their conclusion was a quite unusual contribution to HCI. They suggested that HCI should be more like human-human interaction, which people prefer to be a fun, challenging and beautiful experience. They even implied that HCI designers should put less emphasis on making interactive systems easy to use, and instead try to make them resemble a human-human interaction [13]. If HCI would benefit from similarities with human-human interaction, maybe Dion et al's research about human aesthetics might be applicable to HCI design. Kurosu and Kashimura, 1995, conducted a Japanese study where they let participants test and evaluate ATM machines with different level of aesthetics. They found a strong correlation between users' pre-use judgments of the interface's aesthetics and its perceived usability [14]. The result of this study was later confirmed in an Israeli study by Tractinsky in 1997 [8]. Tractinsky, 2000, conducted a more extensive research study with engineering students in the US. The products that were tested were, again, ATM layouts with different levels of aesthetics and usability. The report was given the rather provoking title "What is beautiful is usable". It corroborated the former studies and added a new interesting finding: When the users evaluated the systems usability after the tests, the most attractive GUI was chosen to be the most usable, not the system with the best usability properties. This indicates that the strong correlation between aesthetics and perceived usability remains intact even after the actual using of a system [9].

Hassenzahl and Norman, 2004, injected new dimension tendencies to the HCI community. Hassenzahl opposed to Tractinsky's "What is beautiful is usable" where he in his "The interplay of beauty, goodness and usability in interactive products" found no relationship between judgments of aesthetics and usability when testing different levels of usability. He did, however, concur with Tractinsky's finding that people find an attractive GUI more usable than an unattractive one, when the actual usability is the same. Hassenzahl says though, that the difference is not in beauty, but in goodness. He defines goodness as the result from impressions of hedonic identification, pragmatic values and mental effort. In other words, goodness is a combination of: perceived pleasure, perceived usability and actual usability. Beauty, he means, is solely based on perceived pleasure [15]. Norman acclaims goodness as a new dimension in HCI, and further he presents three different levels of beauty: visceral, behavioural and reflective. The visceral level is described as biologically determined and triggers immediate judgments or feelings about the product (or person). The

behavioural level is said to be driven by expectation, meaning that an experience with a product not reaching the expectations yields a negative effect. Both the visceral and the behavioural levels are subconscious. The reflective level however, is regarded conscious and intellectually driven. People uses their past experiences and personal meanings to determine whether something (or someone) is beautiful. Furthermore, Norman implies that the visceral and behavioural levels do influence people's opinions of beauty and goodness, but they can only be recognized after they have been interpreted by the reflective level [3].

Using this reasoning one can conclude that when people are asked to fill out a questionnaire about systems they have tested, they use reflective thinking based on their own intellect, prior experiences et cetera. If, for example, a decision about a product's usability made in a questionnaire does not make any difference to the participant, e.g. if the participant does not intend to use the product in the future maybe his/her opinions about the aesthetics will bias on the opinions about the usability, then perhaps the result of the evaluation will be misleading. The possible weakness of questionnaires is one of the conclusions Ben-Bassat, Meyer and Tractinsky, 2006, made in their research leading to the paper "Economic and Subjective Measures of the perceived value of aesthetics and usability". In that study, participants tested computerized phone books with varying levels of usability and aesthetics. The main difference between this test and the ones conducted earlier was the user evaluation process. Part from regular questionnaires, the participants where asked to bid on the different systems in an auction, and were given monetary rewards when completing tasks with the systems quickly. The results showed no differences in the usability versus aesthetics evaluations whether or not the users were given performance rewards. It did show, however, that when using questionnaires, people rated aesthetics higher than usability, and when using auctions, people bid higher to usability than to aesthetics. I.e., the market value is higher for usable systems than for aesthetical systems. This indicates that the use of questionnaires may be a weak manner of evaluating what the users really think [16].

This study is different in many ways from the ones made earlier. First, applications on mobile phones are tested, and not applications on computers. This context difference might invoke different user preferences regarding usability and aesthetics. Katz and Sugiyama, 2006, made a survey among collage students in the US and Japan, about young people's mobile communication behaviour. The outcome of the study was the implication that young people, in a big extent, use their mobile phones to express their self image. The mobile phone acts as a fashion widget and the choice of mobile phone affects the way young people perceive each other [17]. The notion that a mobile phone, similarly to clothing, acts almost as an extension to the human body, may certainly affect the users' preferences about usability and aesthetics. Additionally, young people's view on mobile phones might tangibly differ from older people's view. The earlier studies about usability versus aesthetics mentioned above have very narrow target groups; mostly young computer experienced students have been

15

participating. This study has its focus on different age groups. The age groups' attitudes towards aesthetic and innovative graphic design are elicited and quantified.

# 3 Method

## *3.1 Prototype Development*

Essential for this study to be carried out, were the presence of testable prototypes with varying levels of usability and aesthetics and with different message presentation solutions. For this purpose, we solely developed three different prototypes using Java Micro Edition (ME). The elicitation of ideas regarding looks and functions of the different prototypes was an iterative process. We developed several low fidelity suggestions that were discussed with interaction designers at Sony Ericsson. The final prototypes were developed as Java MIDlet applications that are able to run on mobile phones.

All three prototypes use the same icons to the items that they have in common, and all three prototypes have the same level of implementation details, namely:

- Main menu

- Messaging menu

- Message presentation solution

The prototypes have been given names that correspond with their properties. The names, which will be used throughout this report, are: *Standard Menu*, *Threaded Messaging Menu* and *Aesthetic Menu.* Further descriptions of these are given in the following subsections.

### 3.1.1 *Standard Menu*

*Standard Menu* is designed to be a simple menu system with a high level of usability. The look of the prototype is very simple, colourless and with no animations or other elements to enhance the visual experience. *Standard Menu* is designed to resemble existing solutions, but made simpler to enhance the difference between this prototype and *Aesthetic Menu.* The main colours of *Standard Menu* are black text on white background, with a grey highlight on the selected menu item.

### 3.1.1.1 Main menu

The main menu is a simple nine option grid, similar to what has been very common on especially Sony Ericsson phones for a long time. The grid is navigated through by using the navigation buttons. All four buttons: up, down, left and right, can be used in the navigation. If the user navigates left when the top left menu item is highlighted, the new position will be down right, and vice versa. Furthermore, if the user navigates up from the top row of the grid, the new position will be the corresponding icon item on the bottom row on the grid, and vice versa.



**Figure 3**
*Standard Menu and Threaded*
*Messaging Menu*
Main menu

### 3.1.1.2 Messaging menu

The messaging menu is a four option list with the following items: *write new*, *inbox*, *sent items* and *drafts*. The navigation in the messaging menu is made in the same manner as in the main menu.
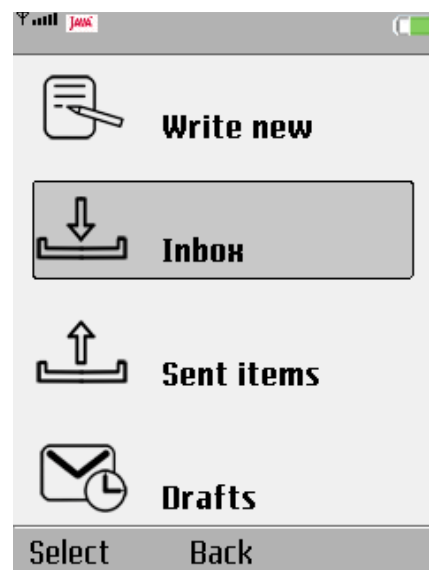


**Figure 4**
*Standard Menu*
Messaging menu

### 3.1.1.3    *Message presentation*

The message presentation is implemented as a standard *inbox/sent items* solution, with the incoming messages placed in the *inbox*, and the sent message in the *sent items* folder.



**Figure 5**
*Standard Menu*
Message presentation

## 3.1.2    *Threaded Messaging Menu*

*Threaded Messaging Menu* is also designed to be a simple menu system with poor aesthetics but with a high level of usability. *Threaded Messaging Menu* uses the same colours as *Standard Menu*. The message presentation solution on the *Threaded Messaging Menu,* however, we designed as an idea of a new way of presenting messages.

### 3.1.2.1    *Main menu*

The main menus are exactly the same on *Standard Menu* and *Threaded Messaging Menu*, see Figure 3.

### 3.1.2.2 Messaging menu

The messaging menu on *Threaded Messaging Menu* is similar to the corresponding menu on *Standard Menu*, but the *inbox/sent items* options are replaced with a single option called *Conversations.*

**Figure 6**
*Threaded Messaging Menu*
Messaging menu

### 3.1.2.3 Message presentation

In the *Conversations* menu, all the conversations are listed, sorted on last date of contact. The message presentation solution is of threaded messaging type, designed to be able to represent both messages (SMS, MMS and email) and phone calls. The list presents the names of people with which conversations have been made. Under each name, the last contact with the specific individual is presented with its date and time and an icon representing the type of the contact (message or phone call), and an arrow representing the direction of the message or phone call, i.e. if the message or phone call was ingoing or outgoing.
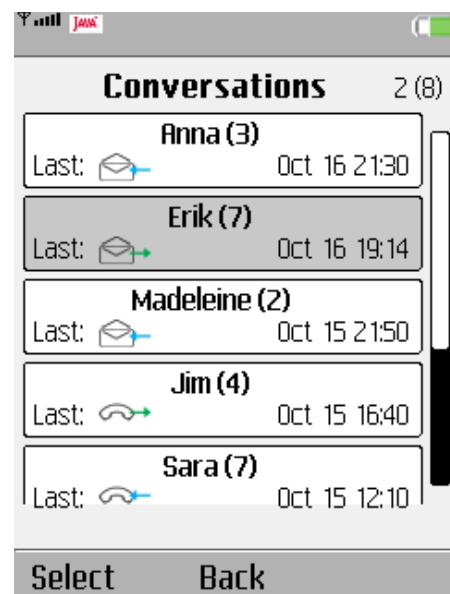
**Figure 7**
*Threaded Messaging Menu*
Conversations

19

When entering a conversation with a specific individual, a list of all contacts made with that individual is presented. The previews of the contacts that are not highlighted have the same information as in the former menu. The highlighted message is enlarged and its content (text, details of the phone call et cetera) is presented.
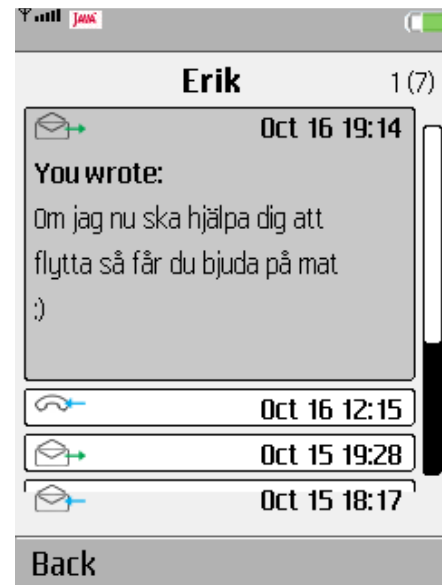


**Figure 8**
*Threaded Messaging Menu*
Specific conversation

### 3.1.3    *Aesthetic Menu*

The *Aesthetic Menu,* we designed to be very innovative both in its general menu structure appearance and regarding its message presentation solution. It is a result of several low fidelity prototype level ideas. The current look of the *Aesthetic Menu* was chosen because it differed most from existing solutions, and also because it was the one with the most exciting look according to us. *Aesthetic menu* is supposed to be aesthetically attractive throughout the whole menu system. However, it has been given a lot of scarcities regarding its usability. This will be discussed later on.

### 3.1.3.1    Main menu

The main menu of the *Aesthetic Menu* is represented by a wheel that is only partly visible on the screen. The menu items are placed on the wheel and are navigated by using the left and right buttons. When an icon is highlighted, a larger representation of the icon is made visible in a bubble on the top left part of the screen, and a text explanation of the choice appears inside the wheel at the bottom left part of the screen. To enhance the aesthetics of the main menu, some animations are added. First, the icon movement when navigating is animated to create a pleasant smoothness in the transitions. Second, the text description is faded in and out when navigating. Third, three lines are emerging from the top right part of the screen every time the user navigates to a new icon.
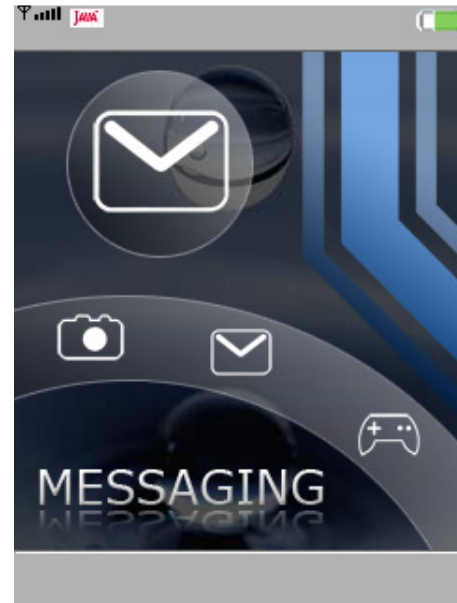


**Figure 9**
*Aesthetic Menu*
Main menu

### 3.1.3.2    Messaging menu

The messaging menu of the *Aesthetic Menu* has the same colours and style, but is built up differently from the main menu. The options, *Conversations, Write new* and *Drafts* are represented as icons glued to lines that emerge from the top right corner of the screen. The navigation on the messaging menu is done with the up and down buttons. When an icon is highlighted, a large representation of the icon and explanatory text is displayed, just as on the main menu. The animation on the messaging menu consists of gradually fading the line that holds the active icon to a brighter colour. Also the explanatory text is faded in and out when navigating.
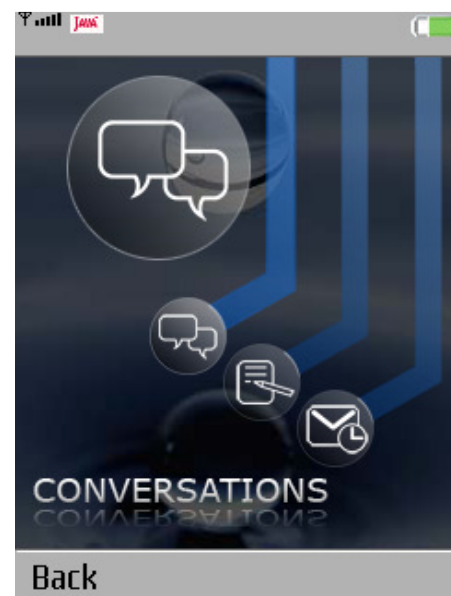


**Figure 10**
*Aesthetic Menu*
Messaging menu

### 3.1.3.3 Message presentation

The message presentation in the *Aesthetic Menu* is of threaded messaging type. The conversations in the solution are represented line-wise on the screen and navigated through in a two dimensional manner. In a bubble to the left on each conversation, the name or picture of the converser is displayed. At the right side of the picture, icons are placed representing messages and phone calls. These icons can be browsed through using the left and right navigation buttons. When a message or a phone call is highlighted, a preview of the specific message or phone call is displayed in a small textbox. When pressing *view* on a preview, the entire message or information about the phone call is displayed. To change the active converser, the up and down navigation buttons are used.



**Figure 11**
*Aesthetic Menu*
Conversations
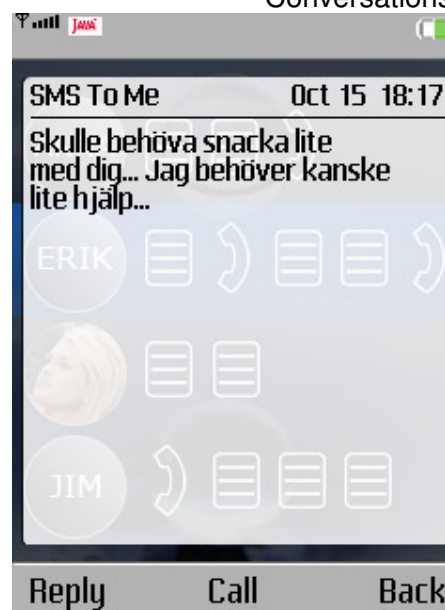


**Figure 12**
*Aesthetic Menu*
Selected Message

22

### 3.1.4 Heuristic Usability Evaluation

To get a hunch of the level of usability of the three prototypes before the tests, a small heuristic evaluation was performed using some of the usability principles defined by Nielsen, mentioned earlier in the usability section of this report. The heuristics used are adjusted to better fit this study:

*Visibility*: The system should inform the users about what is going on, and provide adequate feedback.

*Consistency*: The menu items, language and navigation manner should be the same throughout the system.

*Standards*: The system should follow platform conventions.

*Flexibility and efficiency*: Navigation manner flexibility should be provided to make the system perform fast and efficient regardless of the users' experience of the system.

*Minimalistic design*: The system should not contain text and/or graphics that deteriorates the visibility of the system, and is irrelevant for the usage of the system.

The evaluation was conducted by us, and three interaction designers at Sony Ericsson concurred with our assessments. The results of the evaluation are presented in the results section of this document.

## 3.2 Usability Testing

### 3.2.1 Test Purpose

To gather both qualitative and quantitative data about the different prototypes, leading to the results of this study, a usability test was conducted. The test was designed to retrieve results regarding:

- User preferences about Threaded Messaging.

Are users positive to the idea of presenting in- and outgoing messages and phone calls in a conversation like manner? Do users find threaded messaging as usable as the widely accepted *inbox/Sent items* manner to present messages? To what extent are users willing to switch from other message presentation solutions to the threaded messaging solution? This part of the test will be referred to in the rest of this report as the *Threaded Messaging Test.*

- The correlation between Usability and Aesthetics.

To what extent does aesthetics influence people's judgements of usability? How do people from different age groups react to different levels of innovative and unconventional design in mobile phones? What attributes affect people's choices when they purchase mobile phones? What do people consider their mobile phones to be? A communication device? An entertainment system? A fashion accessory? Or something

else? This part of the test will be referred to in the rest of this report as the *Usability versus Aesthetics Test.*

### 3.2.2 Participants

There were 16 persons participating in the test. The participants were divided into four different age groups: (1) 12-18 years, (2) 19-35 years, (3) 36-55 and (4) over 56 years. There was an equal amount of male and female participants within the different groups.

### 3.2.3 Test Environment

The test was conducted in a controlled lab environment. The lab setup can be seen in Figure 13 and Figure 14. The lab had two rooms: a test room and an observation room. The rooms were separated with a one way mirror. In the test room the participant was seated at a table, facing the mirror. There were two cameras recording the test. One camera was located in the ceiling, recording the mobile phone display, catching the participant's actions. The other camera was located in a 45 degree angle behind the participant, recording his/her facial expressions through the mirror. On the opposite side, behind the participant, the test leader was seated, leading and guiding the participant throughout the test. The test leader had a monitor with a split screen showing the images from both cameras. In the observation room, the test monitor was located along with the technical appliances. The test monitor controlled the recordings, and timed the different tasks. The test monitor also had a split screen monitor where he could follow the actions made by the participant.
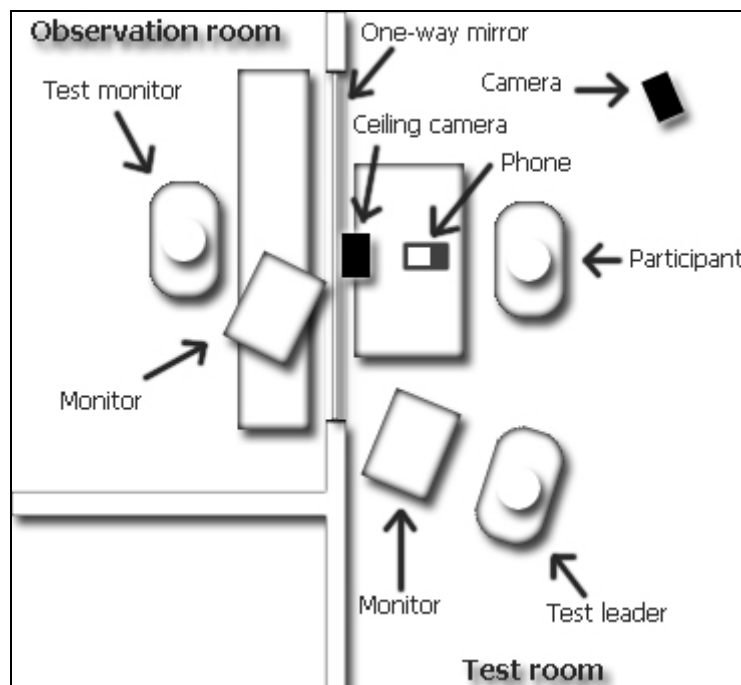


**Figure 13**
Test Lab Overview

The participants performed the test tasks on the prototypes, which were run on black Sony Ericsson w910 phones. Prior to the test, the participants were informed about what buttons that were supposed to use during the tests. The purpose of providing this information was to minimize the effect of hardware influences in the evaluations.



**Figure 14**
Test Lab Setup
Test Room

### 3.2.4 Test Implementation

In this subsection, the test implementation is presented. The documents used in the test are presented in the Appendices B.1-B.7. Note that the original documents were in Swedish, and the appendices contain translated versions. Here follows a description of the test implementation:

The participant was welcomed, seated in the test room and asked to fill out a pre-test questionnaire (Appendix B.1). This questionnaire was designed to gather information about the participant's gender, age and technical skills. Also, information about the participant's current and previous mobile phone brand usage was quantified, along with reasons for changing phones, and important attributes when purchasing new phones. In addition, data about how the participant sees his/her mobile phone was collected. After the participant had finished the pre-test questionnaire, the test leader

read an orientation script (Appendix B.2) informing the participant why he/she was there, and explaining briefly how the test were to be conducted. To provide the different participants with the same background information, the orientation script was the same for all participants.

After the orientation script had been read, the participant started working on the tasks belonging to the *Threaded Messaging Test.* The prototypes used in this test were *Standard Menu* and *Threaded Messaging Menu*. There were three tasks to be completed with each prototype. The tasks were of varying levels of difficulty, and consisted in finding sent and received messages in the message presentation solutions. The tasks can be found in Appendix B.3. The test order of the two prototypes was changed between the different participants, to even out the negative effects of transfer of learning. Transfer of learning is the possibility of users retrieving knowledge about a prototype from having tested a similar prototype before. The tasks were the same for the two prototypes, meaning that the participant first carried out the three tasks with the first prototype, and then the same three tasks with the second prototype.

When the *Threaded Messaging Test* was completed a questionnaire was filled out by the participant with the purpose to gather information about the participants' opinions about the usability of the prototypes (Appendix B.4). Considering the possible weakness of regular questionnaires mentioned by Ben-Bassat and Tractinsky 2006 [16], the participants had to make the decision of which one of the prototypes they would choose if they had to use one of them on their phones in their everyday life.

After this, the *Usability versus Aesthetics Test* began. Similarly to the former test, the *Usability versus Aesthetics Test* contained three tasks which were the same for the two prototypes (Appendix B.5), and with an order that was changed in the same manner as in the *Threaded Messaging Test*. The prototypes used in this part were *Threaded Messaging Menu* and *Aesthetic Menu*.

After the *Usability versus Aesthetics Test* another questionnaire was filled out by the participant designed to collect information about how the participant experienced the usability and the aesthetics of the two prototypes. (Appendix B.6) In similarity with the post-*Threaded Messaging Test* questionnaire, the users had to make a choice between the two prototypes regarding which one they would prefer to use.

Finally the user was asked to fill out a post-test questionnaire (Appendix B.7) gathering more information about what attributes the participant found important when purchasing mobile phones. After this, the participant was handed a cinema ticket as thanks for his/her time and was escorted out of the test lab.

# 4    Results

## *4.1    Heuristic Usability Evaluation*

The usability heuristic evaluation that was made by us in accordance with three interaction designers at Sony Ericsson roughly indicated that the prototypes indeed had different levels of usability, and that the desired usability levels were met. *Standard Menu* and *Threaded Messaging Menu* were designed to be usable, and appropriately they received high average usability grades (4.4 out of 5). *Aesthetic Menu* was designed to possess some usability flaws, and consequently, the usability grades in the evaluation were low (2.4 out of 5).

Table 1 below shows the result of the evaluation. A five scale grade system is used, where 1 is poor, and 5 is excellent. The grades are presented together with an explanation.

**Table 1**
Usability Heuristic Evaluation
Grades and description of heuristics

| | *Standard Menu* | *Threaded Messaging Menu* | *Aesthetic Menu* |
|---|---|---|---|
| **Visibility** | 5<br><br>The active item is highlighted and the soft keys are given explanations about the options available. Moreover, all of the options available are shown in each menu. | 5<br><br>The active item is highlighted and the soft keys are given explanations about the options available. Moreover, all of the options available are shown in each menu. | 2<br><br>The visibility of the messaging menu is fairly good. In the main menu, however, all options are not visible; the user has to use the wheel to browse through the options. In the Conversations it is not possible to get a quick overview of how many conversers there are, and how many contacts that has been made with each converser. No scroll bar or anything similar is indicating that there are more messages and conversers than the ones that the user can see. |
| **Consistency** | 5<br><br>The navigation manner is the same and the highlighted items have the same appearance throughout the whole system | 5<br><br>The navigation manner is the same and the highlighted items have the same appearance throughout the whole system | 3<br><br>The navigation manner differs between the menus. In the main menu, the left and right buttons are used, whilst in the messaging menu the up and down buttons are used. Part from this, the general appearance differs tangibly between the different menus. |

| Standards | 4 | 3 | 2 |
|---|---|---|---|
| | The system is designed with a grid menu, and inbox/sent items message presentation, in a way that is very common among the large mobile phone manufacturers. | The main and messaging menus are built up in the same manner as *Standard Menu*. However, the message presentation is using message threading, which is not widely used. | The system does not look like anything that is common on the mobile phone market. Also the threaded message presentation is a new concept. However, the icons are to some extent designed according to standards. |
| Flexibility and efficiency | 3 | 4 | 3 |
| | All of the navigation buttons can be used to navigate in all menus. The ends of the lists and the grid can be rounded, i.e. if navigating up from the top of a list the new highlighted item will be the bottom item. However, there are no shortcuts for experienced users. | All of the navigation buttons can be used to navigate in all menus. The ends of the lists and the grid can be rounded, i.e. if navigating up from the top of a list the new highlighted item will be the bottom item. Furthermore, the threaded message presentation makes it faster for users to browse through their message history. However, there are no shortcuts for experienced users. | Different navigation buttons are used in different menus. The ends of lists cannot be rounded. The threaded message presentation, however, provides a fast way for users to browse through their conversations. |
| Minimalistic design | 5 | 5 | 2 |
| | The icons are black and minimalistic, and there is neither unnecessary text nor graphics in the system. | The icons are black and minimalistic, and there is neither unnecessary text nor graphics in the system. | The icons are white and minimalistic. However, there are a lot of unnecessary graphics and animations in the menus. |
| Average Usability grade | 4.4 | 4.4 | 2.4 |

## 4.2    Usability Test

This section presents the results of the performed Usability test. As mentioned earlier, the test consisted of two parts: the *Threaded Messaging Test* and the *Usability versus Aesthetics Test.* The overall results common for both tests and the results for each test part respectively, are presented in the following subsections. The results are compiled with former studies and own reflections in the conclusion section later in the report. All quantitative data can also be found in Appendix C.1 – C.8.

### 4.2.1 Common for both tests

Figure 15 below shows the average completion times for the different prototypes on each age group. When comparing the different age groups it is obvious that younger generally means faster. Also, a tangible transfer of learning effect on the *Threaded Messaging Menu* can be seen: since *Threaded Messaging Menu* is used in both tests, the participants have already tried it and, to some extent, learned how to use it. This makes them perform the tasks faster when they use it again in the *Usability versus Aesthetics Test*. The average completion time for *Threaded Messaging Menu* has decreased radically the second time it is tested, reaching times that are similar to those of *Standard Menu*. This indicates a good level of learnability of the *Threaded Messaging Menu*. Thus, it also shows that people adapt very quickly to the concept of threaded messaging in general.
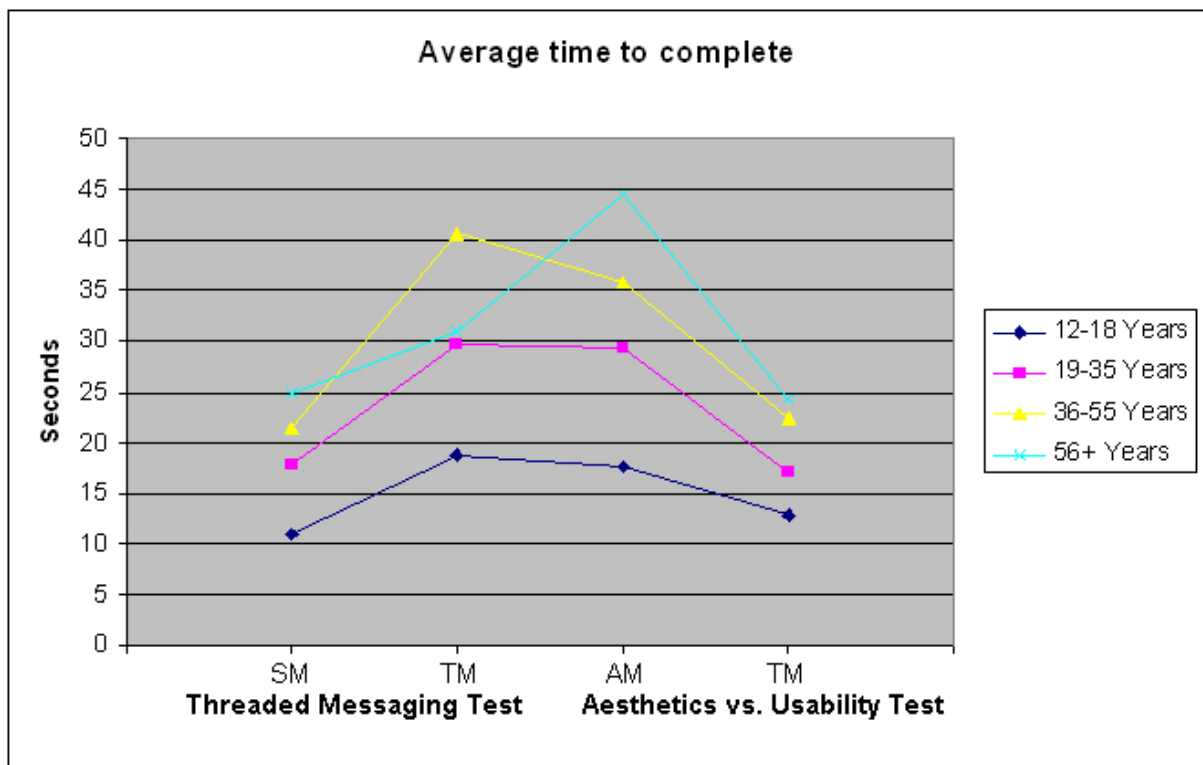


**Figure 15**
Average completion times
SM = *Standard Menu*
TM = *Threaded Messaging Menu*
AM = *Aesthetic Menu*

The data collected in the pair wise comparison in Appendix B.7 was evaluated with Analytic Hierarchy Process (AHP) to create a prioritised order among the attributes. The only certain conclusion that can be drawn from these answers is that *ease of use* is the most important and *brand* is the least important. The prioritisation of the other attributes is more uncertain because of an identified problem regarding the way we

constructed the form. The attribute *design* should have been called *looks* or similar to avoid confusion. The attribute *design* is simply too extensive to be used as an attribute here. The average values from this prioritisation are presented in Table 2 below where the sum of all values is 1 and a higher number indicates higher prioritisation. We did not identify any significant differences in how different age groups prioritised, nor did we see much of a difference in prioritising when grouping the participants on choice of user interface according to Appendix B.6.

**Table 2**
Prioritisation based on pair wise comparison

| Attribute | Design | Price | Functions | Brand | Ease of use |
|---|---|---|---|---|---|
| Average: | 0,18 | 0,17 | 0,20 | 0,11 | 0,35 |

### 4.2.2    *Threaded Messaging Test*

This subsection presents the results of the *Threaded Messaging Test*, in which the concept of *Threaded Messaging* was evaluated. The prototypes tested in this test were *Standard Menu* versus *Threaded Messaging Menu*.

After completing the tasks of the *Threaded Messaging Test*, the participants were asked to compare the usability on a five grade scale of the two prototypes (Appendix B.4). The answers were later translated into a score system where a strong preference of the *Standard Menu* was translated into -100 points and a mild preference of *Standard Menu* was translated into -50. If the participant instead preferred the Threaded Messaging, the scores were the same but positive. If a participant found the prototypes to be equally usable a score of 0 where given. They where also given the question: "If you had to use one of the prototypes on your mobile phone, which one would you choose?" Table 3 below presents the participants' scores and prototype choices.

**Table 3**
Usability score
*Threaded Messaging Test*

| | | Grouped by prototype Choice | |
|---|---|---|---|
| **Participant** | **Score** | ***Standard Menu*** | ***Threaded Messaging Menu*** |
| **1** | -50 | -50 | |
| **2** | -100 | | -100 |
| **3** | 50 | | 50 |
| **4** | -100 | -100 | |
| **5** | 100 | | 100 |
| **6** | -100 | -100 | |
| **7** | 50 | | 50 |
| **8** | 100 | | 100 |
| **9** | -50 | | -50 |
| **10** | -100 | -100 | |
| **11** | 50 | | 50 |
| **12** | -50 | | -50 |
| **13** | -50 | -50 | |
| **14** | -100 | -100 | |
| **15** | -100 | -100 | |
| **16** | -100 | -100 | |
| **Average:** | **-34,4** | **-87,5** | **18,8** |

The average score for all participants was -34,4. This score indicates that, for being the first time using *Threaded Messaging*, the participants did not find it that difficult to use, compared to a standard solution with inbox/sent items. 50% of the participants chose that they would prefer to use the threaded messaging solution, and the average score grouped by choice indicates a distinct difference between the groups. The participants that chose threaded messaging also found the threaded messaging solution to be slightly more usable (18,8) and the ones that chose the standard solution considered the standard solution to be a lot more usable (-87,5).

An explanation to why the participants made the choices they did, could probably be found in the average completion time in the test. As can be seen in Table 4 below, the ones that chose *Standard Menu* was considerably faster at performing the different tasks than the ones that chose *Threaded Messaging Menu*. One obvious reason for this difference is experience; the group with the fastest average time is probably more used to communicate with messages. The reason for why this group chose *Standard Menu* could be the fact that they are used to the traditional message presentation solution and have learned to perform tasks very efficiently on it.

**Table 4**
Completion times based on choice of prototype
SM = *Standard Menu*
TM = *Threaded Messaging Menu*
1, 2, 3a and 3b = the different tasks

| Completion times based on choice of prototype | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Standard Menu* **chosen** | | | | | | | |
| **Participant** | **SM 1** | **SM 2** | **SM 3a** | **SM 3b** | **TM 1** | **TM 2** | **TM 3a** | **TM 3b** |
| **1** | 55 | 21 | 34 | 15 | 39 | 22 | 23 | 9 |
| **4** | 21 | 18 | 16 | 10 | 126 | 22 | 21 | 3 |
| **6** | 8 | 13 | 17 | 7 | 103 | 12 | 28 | 2 |
| **10** | 19 | 27 | 20 | 28 | 47 | 47 | 23 | 2 |
| **13** | 12 | 27 | 16 | 4 | 15 | 8 | 9 | 2 |
| **14** | 14 | 13 | 9 | 7 | 33 | 80 | 16 | 2 |
| **15** | 8 | 15 | 11 | 5 | 32 | 12 | 12 | 5 |
| **16** | 9 | 7 | 11 | 6 | 25 | 20 | 28 | 2 |
| **Average:** | 18,3 | 17,6 | 16,8 | 10,3 | 52,5 | 27,9 | 20,0 | 3,4 |
| **Average Total:** | **15,7** | | | | **25,9** | | | |
| *Threaded Messaging Menu* **chosen** | | | | | | | |
| **Participant** | **SM 1** | **SM 2** | **SM 3a** | **SM 3b** | **TM 1** | **TM 2** | **TM 3a** | **TM 3b** |
| **2** | 26 | 20 | 18 | 25 | 104 | 39 | 37 | 5 |
| **3** | 18 | 21 | 21 | 13 | 20 | 14 | 18 | 8 |
| **5** | 26 | 24 | 19 | 15 | 77 | 59 | 26 | 3 |
| **7** | 21 | 17 | 13 | 8 | 147 | 27 | 32 | 5 |
| **8** | 13 | 14 | 18 | 9 | 25 | 10 | 15 | 4 |
| **9** | 43 | 43 | 31 | 18 | 136 | 59 | 35 | 9 |
| **11** | 38 | 22 | 27 | 20 | 30 | 13 | 14 | 3 |
| **12** | 18 | 21 | 32 | 23 | 73 | 22 | 22 | 3 |
| **Average:** | 25,4 | 22,8 | 22,4 | 16,4 | 76,5 | 30,4 | 24,9 | 5 |
| **Average Total:** | **21,7** | | | | **34,2** | | | |

When looking at the different age groups, there was a lower age average on the participants choosing the standard messaging solution. All of the participants in the youngest age group chose the standard solution. In the other age groups, the choices were quite evenly distributed, with the exception of the oldest group that had the surprisingly high 50 % preference towards the Threaded Messaging solution.

Apart from the quantitative data analysed in this section, a lot of qualitative data regarding the *Threaded Messaging Test,* such as user comments, open-answer questionnaire answers and discussion about why they made the choices they did, have been collected as well. The most important, and also the most explicit user comments, were the ones explaining why they made the choices they did. Table 5 below shows the most common reasons for the participants' choices.

**Table 5**
Qualitative data about prototype choice
*Threaded Messaging Test*
SM = *Standard Menu*
TM = *Threaded Messaging Menu*

| Reasons for choosing SM | Reasons for choosing TM |
| --- | --- |
| - The inbox/sent items solution is very common, and most people are used to it, and know how to use it. | - No need to go back and forth in the menu system when browsing messages, thus easier to follow a conversation. |
| - Easier to distinguish incoming messages from sent messages. | - Easy to learn how to use. |

Moreover, some of the participants thought that it was unnecessary, some even annoying, to mix messages with telephone calls in the conversations, as is done on the *Threaded Messaging Menu* (and also on the *Aesthetic Menu,* which they had not yet tested at this point). They implied that it would be better to separate the phone call conversations from the text message conversations.

As mentioned earlier, eight of the participants (50%) chose the prototype with the threaded messaging solution. However, six of the participants that chose the prototype with the standard messaging solution were positive to the idea of threaded messaging.

### 4.2.3 *Usability versus Aesthetics Test*
This subsection presents the results of the *Usability versus Aesthetics Test.* In this test, the prototypes *Threaded Messaging Menu* and *Aesthetic Menu* were tested against each other.

After the participants had completed the tasks of the *Usability versus Aesthetics Test*, they where asked to grade the prototypes on how easy they were to use and how aesthetically pleasing they were. They were also asked to compare the two prototypes on the same attributes; ease of use and aesthetics. Finally they were asked the question: "If you had to use one of the prototypes on your mobile phone, which one would you choose?" (Appendix B.6) Table 6 below shows the participants' grades (1-5) of usability and aesthetics of the different prototypes. The results are grouped based on choice of prototype.

**Table 6**
Usability and Aesthetics grades grouped on choice of prototype
*Usability versus Aesthetics Test*
AM = *Aesthetic Menu*
TM = *Threaded Messaging Menu*

| **Participants choosing *Aesthetic Menu* (AM)** | | | | |
|---|---|---|---|---|
| **Participant** | **Usability AM** | **Aesthetics AM** | **Usability TM** | **Aesthetics TM** |
| **3** | 4 | 5 | 5 | 2 |
| **5** | 3 | 5 | 5 | 3 |
| **9** | 3 | 4 | 2 | 3 |
| **10** | 4 | 4 | 4 | 3 |
| **13** | 4 | 5 | 4 | 3 |
| **14** | 4 | 5 | 4 | 3 |
| **15** | 3 | 4 | 2 | 3 |
| **16** | 3 | 4 | 4 | 3 |
| **Average** | **3,5** | **4,5** | **3,8** | **2,9** |

| **Participants choosing *Threaded Messaging Menu* (TM)** | | | | |
|---|---|---|---|---|
| **Participant** | **Usability AM** | **Aesthetics AM** | **Usability TM** | **Aesthetics TM** |
| **1** | 4 | 4 | 5 | 3 |
| **2** | 2 | 3 | 5 | 3 |
| **4** | 3 | 2 | 5 | 3 |
| **6** | 3 | 3 | 3 | 3 |
| **7** | 2 | 5 | 3 | 3 |
| **8** | 2 | 3 | 5 | 5 |
| **11** | 2 | 4 | 4 | 2 |
| **12** | 3 | 2 | 5 | 3 |
| **Average** | **2,6** | **3,3** | **4,4** | **3,1** |

50% of the participants answered that they would choose *Aesthetic Menu.* The results show that the participants that chose the *Aesthetic Menu* did not experience much of a difference in usability between the prototypes (The *Aesthetic Menu* got an average grade of 3,5, compared to *Threaded Messaging Menu*'s average of 3,8). The same users on the other hand gave the aesthetics of *Aesthetic Menu* much higher grades than *Threaded Messaging Menu* (4,5 for *Aesthetic Menu* compared to 2,9 for *Threaded Messaging Menu*).

The results are the exact opposite for the other 50% that chose *Threaded Messaging Menu*. They did not see much of a difference in aesthetics between the prototypes (3,3 for *Aesthetic Menu* and 3,1 for *Threaded Messaging Menu*). But when it came to usability they saw a great difference (2,6 for *Aesthetic Menu* and 4,4 for *Threaded Messaging Menu)*. These results indicate a strong correlation between perceived usability and aesthetics, in both directions. In other words: people that find a GUI aesthetically attractive seem to also find it usable, and people that do not find the same GUI aesthetically attractive do not find it that usable. Vice versa, people that find a GUI very usable also find it to be aesthetically pleasing, while people not finding the same GUI very usable do not find it that aesthetically pleasing.

When looking at the different age groups, all of the teenagers preferred aesthetics over usability. Consequently, the average age for the ones choosing *Aesthetic Menu* was tangibly lower than for the ones choosing the usable menu. However, in the oldest age group, 50 % of the participants chose the aesthetic GUI, which indicates a quite surprisingly high level of admittance towards aesthetics and innovative looking GUIs among older people.

Note also, comparing Table 6 and Table 1, that the average usability grades of the ones that chose the *Threaded Messaging Menu* are quite similar to the ones that were yielded in the pre-test heuristic evaluation (2,6 compared to 2,4 for *Aesthetic Menu* and 4,4 compared to 4.4 for *Threaded Messaging Menu*), while the heuristic evaluation did not at all correspond with the average usability grades of the ones that chose *Aesthetic Menu*. A reason for this may be that people that do not see aesthetics as an important factor in the choice of mobile phone GUIs perceives the usability in terms of classic usability principles. Another way to look at this is to conclude that the classic usability principles, e.g. Nielsen's heuristics, are defined with little or no regard to aesthetics as a usability factor.

The completion times in the *Usability versus Aesthetics Test* are presented in Table 7 below. The average times show that the tasks took longer time to complete on *Aesthetic Menu*. This was the expected result due to the built in usability flaws. To some extent, however, the difference could also be explained with the transfer of learning effect of the *Threaded Messaging Menu* mentioned earlier, causing the time differences between the *Aesthetic Menu* and the *Threaded Messaging Menu* to increase. The table shows no significant difference in average completion times of the *Threaded Messaging Menu* based on the participants' choices of prototype. However, it shows lower average completion times on *Aesthetic Menu* for the ones choosing *Aesthetic Menu*. An explanation for this is that participants that performed good on the *Aesthetic Menu* got a good feeling about it, and consequently the choice of use fell on the *Aesthetic Menu*.

**Table 7**
Completion times based on choice of prototype
*Usability versus Aesthetics Test*
AM = *Aesthetic Menu*
TM = *Threaded Messaging Menu*
1, 2, 3a and 3b = the different tasks

**Completion times based on choice of prototype**

*Aesthetic Menu* **chosen**

| Participant | AM 1 | AM 2 | AM 3a | AM 3b | TM 1 | TM 2 | TM 3a | TM 3b |
|---|---|---|---|---|---|---|---|---|
| 3 | 41 | 34 | 27 | 4 | 19 | 20 | 20 | 2 |
| 5 | 75 | 27 | 78 | 2 | 33 | 24 | 33 | 2 |
| 9 | 96 | 23 | 50 | 13 | 72 | 29 | 36 | 4 |
| 10 | 79 | 25 | 48 | 1 | 42 | 40 | 40 | 3 |
| 13 | 45 | 10 | 23 | 3 | 10 | 8 | 13 | 2 |
| 14 | 35 | 14 | 13 | 2 | 13 | 11 | 19 | 20 |
| 15 | 34 | 11 | 36 | 3 | 18 | 13 | 20 | 2 |
| 16 | 18 | 10 | 21 | 3 | 23 | 10 | 19 | 3 |
| **Average:** | **52,9** | **19,3** | **37,0** | **3,9** | **28,8** | **19,4** | **25,0** | **4,8** |
| **Average Total:** | **28,3** | | | | **19,5** | | | |

*Threaded Messaging Menu* **chosen**

| Participant | AM 1 | AM 2 | AM 3a | AM 3b | TM 1 | TM 2 | TM 3a | TM 3b |
|---|---|---|---|---|---|---|---|---|
| 1 | 112 | 30 | 36 | 9 | 57 | 16 | 23 | 5 |
| 2 | 108 | 14 | 27 | 5 | 32 | 18 | 30 | 5 |
| 4 | 43 | 15 | 29 | 2 | 32 | 16 | 31 | 3 |
| 6 | 20 | 16 | 14 | 3 | 15 | 21 | 26 | 2 |
| 7 | 97 | 12 | 34 | 2 | 54 | 22 | 30 | 4 |
| 8 | 68 | 42 | 34 | 5 | 20 | 8 | 21 | 2 |
| 11 | 91 | 19 | 45 | 2 | 16 | 7 | 13 | 2 |
| 12 | 124 | 21 | 38 | 6 | 29 | 17 | 25 | 3 |
| **Average:** | **82,9** | **21,** | **32,1** | **4,3** | **31,9** | **15,6** | **24,9** | **3,3** |
| **Average Total:** | **35,1** | | | | **18,9** | | | |

The same scoring system was used in the *Usability versus Aesthetics Test*, as in the *Threaded Messaging Test*. When looking at the comparing test scores in Table 8, the same patterns are seen as in the participants' grading. The *Aesthetic Menu* fans see little difference in usability (-12,5) and great difference in aesthetics (81,3.) The

36

*Threaded Messaging Menu* fans are the exact opposites, seeing great difference in usability (-75,0) and no difference at all in aesthetics (0).

**Table 8**
Test scores
*Usability versus Aesthetics Test*
AM = *Aesthetic Menu*
TM = *Threaded Messaging Menu*

| **Test scores** *Aesthetics versus Usability Test* | | | | | | |
|---|---|---|---|---|---|---|
| | **All** | | **AM chosen** | | **TM chosen** | |
| **Participant** | **Score Usability** | **Score Aesthetics** | **Score Usability** | **Score Aesthetics** | **Score Usability** | **Score Aesthetics** |
| **1** | -50 | 100 | | | -50 | 100 |
| **2** | -100 | -50 | | | -100 | -50 |
| **3** | 0 | 100 | 0 | 100 | | |
| **4** | -100 | -100 | | | -100 | -100 |
| **5** | -100 | 100 | -100 | 100 | | |
| **6** | 0 | 50 | | | 0 | 50 |
| **7** | -100 | 100 | | | -100 | 100 |
| **8** | -100 | -100 | | | -100 | -100 |
| **9** | 0 | 0 | 0 | 0 | | |
| **10** | 0 | 100 | 0 | 100 | | |
| **11** | -100 | 100 | | | -100 | 100 |
| **12** | -50 | -100 | | | -50 | -100 |
| **13** | -50 | 100 | -50 | 100 | | |
| **14** | 100 | 100 | 100 | 100 | | |
| **15** | -50 | 100 | -50 | 100 | | |
| **16** | 0 | 50 | 0 | 50 | | |
| **Average** | **-43,8** | **40,6** | **-12,5** | **81,3** | **-75,0** | **0,0** |

In the pre-test questionnaire, the participants recorded all of the different features that they generally use on their mobile phones. In Table 9 below, the number of features different participants used is presented along with which prototype they would choose to use. It is difficult to find a pattern for why different participant chose either *Aesthetic Menu* or *Threaded Messaging Menu* based on the way they use their mobile phones. However, a tendency can be seen: participants that use the most and the very least functions seem to choose *Aesthetic Menu* and the ones in between seem to choose *Threaded Messaging Menu*.

**Table 9**
Number of features in phone used
AM = *Aesthetic Menu*
TM = *Threaded Messaging Menu*

| Number of features in phone used | | |
|---|---|---|
| **Participant** | **AM chosen** | **TM chosen** |
| **1** | | 6 |
| **2** | | 4 |
| **3** | 6 | |
| **4** | | 2 |
| **5** | 1 | |
| **6** | | 2 |
| **7** | | 5 |
| **8** | | 3 |
| **9** | 1 | |
| **10** | 1 | |
| **11** | | 1 |
| **12** | | 4 |
| **13** | 8 | |
| **14** | 9 | |
| **15** | 6 | |
| **16** | 6 | |

Just as in the *Threaded Messaging Test,* a lot of qualitative data was collected in the *Usability versus Aesthetics Test*. Table 10 below shows the most common reasons for choosing to use either the *Aesthetic Menu* or the *Threaded Messaging Menu.*

**Table 10**
Qualitative data about prototype choice
*Usability versus Aesthetics Test*
AM = *Aesthetic Menu*
TM = *Threaded Messaging Menu*

| **Reasons for choosing AM** | **Reasons for choosing TM** |
|---|---|
| - Nice-looking | - Easy to use |
| - Different from what has been seen before. A new way of thinking. | - No unnecessary graphics, "less is more". |
| - Fun to look at and fun to use | |

Moreover, some of the participants thought that *Threaded Messaging Menu* was easier to use in all aspects, except in the conversations, where they thought that *Aesthetic*

*Menu* provided a better overview. However, the feeling of overview may be triggered by the fact that most participants thought that the conversations in *Aesthetic Menu* had much better aesthetics than the conversations on *Threaded Messaging Menu.* Some of the participants also commented that it is a very good idea to provide the possibility to add pictures of the conversers in the conversations, which is possible on the *Aesthetic Menu.* Adding pictures of the conversers may indeed be a good way to enhance the feeling of presence between the conversers. Overall, user comments about usability of both *Aesthetic Menu* and *Threaded Messaging Menu* corresponded surprisingly well with the comments elicited in the pre-test heuristic evaluation.

# 5    Discussion

Like most other studies, this study has limitations; i.e. things that may have deteriorated the test results. As mentioned earlier, the usability test conducted in this study is not to be considered a scientific test. The results yielded are to be considered as hints to findings and recommendations rather than scientific research results. However, there are some things that could have been performed in other ways to increase the result credibility.  These limitations and scarcities are accounted for in this section.

A majority of the people participating in the test were employees at the Ingvar Kamprad Design Centre, working with design and subjects related to the subject of this study. This may have caused the results of the test to differ a bit in some direction; compared to if the participants would have been selected randomly.

The fact that we carried out the tests in a controlled test environment has, as mentioned before, its advantages as well as disadvantages. Testing in a controlled environment made it easy to ensure that the tests were conducted in the same way for all the participants. It also provided us with valuable test data in the form of video and sound recordings. However, there are disadvantages with controlled environment testing that may have impaired the results. First, the controlled test lab is an artificial environment not corresponding to the natural environments in which people perform similar tasks in their everyday lives. Second, there is a risk that participants felt a bit uneasy during the tests due to the one-way mirror, cameras, microphones, monitors et cetera (See Figure 14). These disadvantages may have caused the participants to act in ways that they would not do normally, if the tasks were given in a, for them, natural environment.

Another remark is about the test order and the prototypes tested. *Threaded Messaging Menu* was tested twice which naturally lead to transfer of learning. This transfer of learning effect caused lower average completion times of the *Threaded Messaging Menu* in the second test (*Usability versus Aesthetics Test)* compared to *Aesthetic Menu*. However, although *Aesthetic Menu* had a different appearance, it also had the threaded messaging concept, which may have alleviated this effect a bit. One positive

outcome of testing *Threaded Messaging Menu* twice is that it could be seen how fast the participants adapted to and learned threaded messaging, indicating a fast general acceptance for the threaded messaging concept.

Some of the questions given to the participants in the questionnaires was either poorly formulated, or negligently answered by the participants. The answers received were in some cases very contradictious, forcing us to ignore them in the analysis. An example of contradictious answers that still lead to a conclusion is the answers received by the teenage group to the multi-answer question: "What is your mobile phone for you?" Even though the majority of the teenagers had stated earlier that they commonly use the majority of functions in their mobile phones, such as camera, FM-radio, music player, calculator etc, they still answered that the mobile phone for them is "just a phone". We had expected people that used a lot of mobile phone features to answer that they saw their mobile phone as an entertainment device or similar. However, it might be so that "just a phone" for today's teenagers is a device with all those applications mentioned, and in that case, that is a result of its own.

# 6        Conclusion

## 6.1        *Threaded Messaging*

The results of this study indicated that threaded messaging is a usable concept that most people 1) find to be easy to learn and 2) are interested in using on their mobile phones in their everyday lives. First, the users' usability comparison score of the prototype with the threaded messaging solution was relatively high considering that it was a new concept to them. Second, threaded messaging got much better completion times the second time it was used (equally good as the standard solution) which indicates that people adapt very fast to the threaded messaging concept. Third, and most convincing, is the fact that 50 % of the participants chose that they would prefer the prototype with the threaded messaging solution over the prototype with the standard inbox/sent items messaging solution, and out of the persons choosing the standard solution, 75 % were still positive to the idea of threaded messaging. This is considered very high because 1) it was the first time the participants used the threaded messaging concept on mobile phones and 2) the threaded messaging solution was tested against the very common and publicly accepted inbox/sent items solution. The only disadvantage people saw with the threaded messaging concept, part from the fact that they were not used to it, was that it was a little bit harder to distinguish incoming from sent messages. Advantages included that it is easier to follow conversations with threaded messaging, and that it is an easy concept to learn.

The threaded messaging concept seems to have most of its fans in the middle age groups. It seemed in our test, a bit surprisingly, as though the youngest and fastest writing people are more reluctant to switch from the standard solution. However it is an overall positive response towards threaded messaging that has been observed.

We are confident that the threaded messaging concept would be received favourably on the market, and we recommend mobile phone manufacturers to produce and release implementations of the concept upon the market as soon as possible. Furthermore, to also please the people with doubts, it would be good in a transition phase to provide the possibility to choose between a standard solution and a threaded messaging solution on the phones.

## *6.2      Usability versus Aesthetics*

The results of the *Usability versus Aesthetics Test* showed a strong correlation in both ways between perceived usability and perceived aesthetics in mobile GUIs. In other words, a person's comprehension of a mobile GUI's "ease of use" is strongly affected by his/her comprehension of the mobile GUI's aesthetics, and vice versa. 50 % of the participants chose that they would prefer to use the aesthetic GUI with poor traditional usability properties, and 50 % chose that they would prefer to use the less aesthetic GUI with good usability properties. The participants preferring aesthetics over classic usability perceived no significant difference in usability between the two GUIs (grade average: 3,4 vs. 3,8; score average: -12,5), but instead they thought that the aesthetic GUI was much more aesthetically attractive (grade average: 4,5 vs. 2,9; score average: 81,3). The participants choosing classic usability over aesthetics perceived a big different in usability between the two GUIs (grade average: 2,6 vs. 4,4; score average: -75,0), but they found no significant difference in aesthetics (grade average: 3,3 vs. 3,1; score average: 0). This corroborates with the results of Kurosu and Kashimura, 1995 and Tractinsky 1997 [8] [9] that the pre-test judgement of aesthetics bias persons' estimations of usability. It also coincides with Tractinskys "What is beautiful is usable", 2000, that the correlation remains intact even after the actual use of the system [14]. In other words, it seems to be a strong correlation between perceived usability and perceived aesthetics, both before and after the actual use of the system. Additionally, this study implies that the correlation between usability and aesthetics documented by Tractinsky et al in computer GUIs also exists on mobile phone GUIs. Furthermore, and even more interesting is the results of this study which imply that the usability/aesthetics correlation is bi-directional, meaning that aesthetics affects perceived usability and usability affects perceived aesthetics.

Moreover, the study indicates that teenagers are most positive to innovative and aesthetic graphic design, while middle aged people are most positive to clean design with good classic usability properties. However, the study also shows that there is a surprisingly high curiosity and positivism towards aesthetic design in the oldest age group (50% of participants over 55 years of age preferred the aesthetic GUI).

Also noted, is the fact that the participants that chose the usable GUI graded the different GUIs very similar to the grades yielded in the pre-test heuristic evaluation. The participants choosing the aesthetic GUI however gave the aesthetic GUI much higher usability grades and the usable GUI much lower usability grades, in comparison.  This indicates that classical heuristic usability evaluation has little or no

concern to the influence of aesthetics to the usability. It is here concluded, that aesthetics is a very important aspect to usability, and therefore, classic usability properties such as Nielsen's 10 usability principles should not alone be used to measure usability. This is somewhat an addition to Tractinsky's article in 1997 where he mentioned the HCI community's neglecting of the aesthetics [8]. However, aesthetics is considered by most researchers as extremely subjective, and thus very hard to quantify. Therefore, considering the correlation between usability and aesthetics, it is hard to perform accurate usability evaluations without thorough user involvement. With this reasoning, we emphasize the importance of involving actual users early and often in development processes, to ensure that the aesthetics does not affect the usability in any undesired way, and also to ensure that the usability does not affect the aesthetics in any undesired way.

## 6.3    Problem Statement Answers

To bring this report to a close, this final subsection contains the initial problem statements provided with brief answers.

**Q1**. Are users positive to the concept of threaded messaging?
**A1**. *Yes. 88 % of the participants expressed positivism about the concept.*

**Q2**. Do users find threaded messaging as usable as the inbox/sent items solution?
**A2**. *The completion times were better for the standard solution. However, a high level of learnability was observed for the threaded messaging concept.*

**Q3**. Are users willing to switch from inbox/sent items solutions to threaded messaging?
**A3**. *Yes. Even though it was the first time that the users tried the concept of threaded messaging on mobile phones in our test, 50 % said that they would choose to use that solution over the standard solution.*

**Q4**. To what extent does aesthetics influence people's judgements of usability?
**A4**. *Usability and aesthetics are correlated bi-directionally. Perceived aesthetics affects perceived usability, and perceived usability affects perceived aesthetics.*

**Q5**. How do people from different age groups react to different levels of innovative graphic design in mobile phones?
**A5**. *Teenagers are early adopters of innovative graphic design, while the middle age groups were the most conservative. Furthermore, a curiosity among older people towards innovative graphics was observed.*

# 7     References

[1] Norman, D. A., 1984, Things That Make Us Smart: Defending Human Attributes in the Age of the Machine, Reading, Massachusetts.

[2] Norman, D. A., 1988, The Psychology of Everyday Things, Basic Books, New York.

[3] Norman, D. A., 2004, Beauty, Goodness, and Usability, Special Section of Human-Computer Interaction, Volume 19, Number 4.

[4] Nielsen, J., 1993. Usability Engineering. AP Professional.

[5] Donahue G. M., 2001, Usability and the bottom line, IEEE Software, 18, 1, 31-37.

[6] Carritt, E.F., 1932, Philosophies of Beauty, Clarendon Press, Oxford.

[7] Snow, C.P., 1956, The two cultures (Canto ed.), Cambridge University Press, Cambridge.

[8] Tractinsky, N., 1997, Aesthetics and apparent Usability: Empirically assessing cultural and methodological issues. Proceedings of the CHI 97 Conference on Human Factors in Computing Systems, New York: ACM.

[9] Tractinsky, N., Adi, S.-K., & Ikar, D., 2000, What is Beautiful is Usable. *Interacting with Computers, 13* (2), 127-145.,

[10] Preece, Rogers and Sharp, 2002, Interaction design: Beyond human-computer interaction, John Wiley & Sons, Inc.

[11] Rubin, Jeffrey, 1994, Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests

[12] Dion, K., Berscheid, E., Walster, E. 1972, What is beautiful is good. Journal of Personality and Social Psychology, 24(3), 285-290.

[13] Overbeeke, K., Djajadiningrat, T., Hummels, C. and Wensveen, S., 2000, Beauty In Usability: Forget About Ease Of Use!, Delft University of Technology, The Netherlands.

[14] Kurosu M., and Kashimura, K., 1995, Apparent Usability vs. inherent Usability: experimental analysis on the determinants of the apparent Usability, Conference companion on Human factors in computing systems, p.292-293, May 07-11, 1995, Denver, Colorado, United States

[15] Hassenzahl, M., 2004, The interplay of beauty, goodness, and Usability in interactive products. *Human-Computer Interaction*, *19*,

[16]   Ben-Bassat, T. A., Meyer, J., Tractinsky, N, 2006, Economic and subjective measures of the perceived value of Aesthetics and Usability. ACM Trans. Comput.-Hum. Interact. 13(2): 210-234.

[17]   Katz, J. E. and Sugiyama, S., 2006, Mobile phones as fashion statements: evidence from student surveys in the US and Japan, Sage

[18]   Nielsen, J., 2004b, Ten Usability Heuristics. Also available on: http://www.useit.com/papers/heuristic/heuristic_list.html (2007-12-12)

# 8 Revision history

The table below contains the changes made to this document.

| Date | Change | Responsible | Version |
|---|---|---|---|
| 2007-10-11 | Creation of the document | JG; NH | 1.0 |
| 2007-11-13 | Addition of Prototype Development section | JG, NH | 1.1 |
| 2007-11-15 | Revision of headings and text | NH | 1.2 |
| 2007-11-16 | Addition of prototype screenshots | JG, NH | 1.3 |
| 2007-11-20 | Addition of Test Method | JG, NH | 1.4 |
| 2007-12-05 | Test Method revision and concretization of the usability section. | JG | 1.5 |
| 2007-12-10 | Addition of usability evaluation background | JG, NH | 1.6 |
| 2007-12-11 | Addition of test results | JG, NH | 1.7 |
| 2007-12-12 | Addition of qualitative data and revision of the references. | JG, NH | 1.8 |
| 2007-12-18 | Addition of conclusion and abstract | JG | 1.9 |
| 2007-12-19 | Addition of limitations and scarcities | NH | 2.0 |
| 2007-12-20 | Document revision after external supervisor comments | JG, NH | 2.1 |
| 2008-01-14 | Document revision after internal supervisor comments | JG, NH | 2.2 |
| 2008-01-15 | Correction of Typos | JG, NH | 2.3 |
| 2008-01-17 | Revision of decimals and commas | JG, NH | 2.4 |
| 2008-01-21 | Correction of Typos and addition of Appendix C | JG, NH | 2.5 |
| 2008-01-23 | Document revision after opponents' comments | JG, NH | 2.6 |

| Document Type: | Authors: | Version: | Last Change: |
|---|---|---|---|
| Report | Johan Garnolf and Nils Hallberg | 2.6 | 2008-01-23 |

46

# Appendices

## *A        Project Plan v.1.4*

### A.1        Introduction

The purpose of this document is to cover the overall Master Thesis project plan for the MMS and iMode Messaging unit at Sony Ericsson, in cooperation with the Ergonomics and Aerosol Technology department at the Faculty of Engineering, Lund University, Lund, Sweden. The goals with the project are presented along with a description of the organization with the different stakeholders and descriptions and schedule for the different activities and deliverables.

This document is of a dynamic nature and will be updated as the project proceeds.

### A.2        Goals

The main goals with the project are:

- To examine new ways to handle and present messages on mobile phones. The far most common manner to present SMS, MMS and email messages on mobile phones today is to place them in *inbox* and *sent items* directories. In this project, prototypes are developed in which the users can browse through whole conversations made with specific persons. In this way users will get instant knowledge about the background to each incoming message. This conversation presentation solution is referred to in this document as *threaded messaging.* The threaded messaging prototypes are tested and evaluated in comparison with existing solutions.

- To examine how people from different age groups react to different levels of innovative and unconventional design and how they admit different levels of aesthetics of mobile phone graphical user interfaces. It is also quantified to what extent aesthetics influence people's judgements of usability.

Three prototypes of mobile phone menu systems are developed, where the messaging functions is the part with the highest level of implemented details. One of the prototypes has an innovative and aesthetically pleasing design, but is deliberately given some usability flaws. The other two prototypes are plain and dull graphically, but have good usability properties. The aesthetic prototype and one of the unaesthetic ones have innovative message presentation solutions that implement the threaded messaging solution. The prototypes are developed using Java ME, and are able to run as Java MIDlets on mobile phones. The prototypes include main menus, messaging menus and message handling solutions.

The messaging functions of the three prototypes are tested thoroughly in a usability lab, and the project results are extracted from the analysis of these tests along with a study of related work and research on the topic.
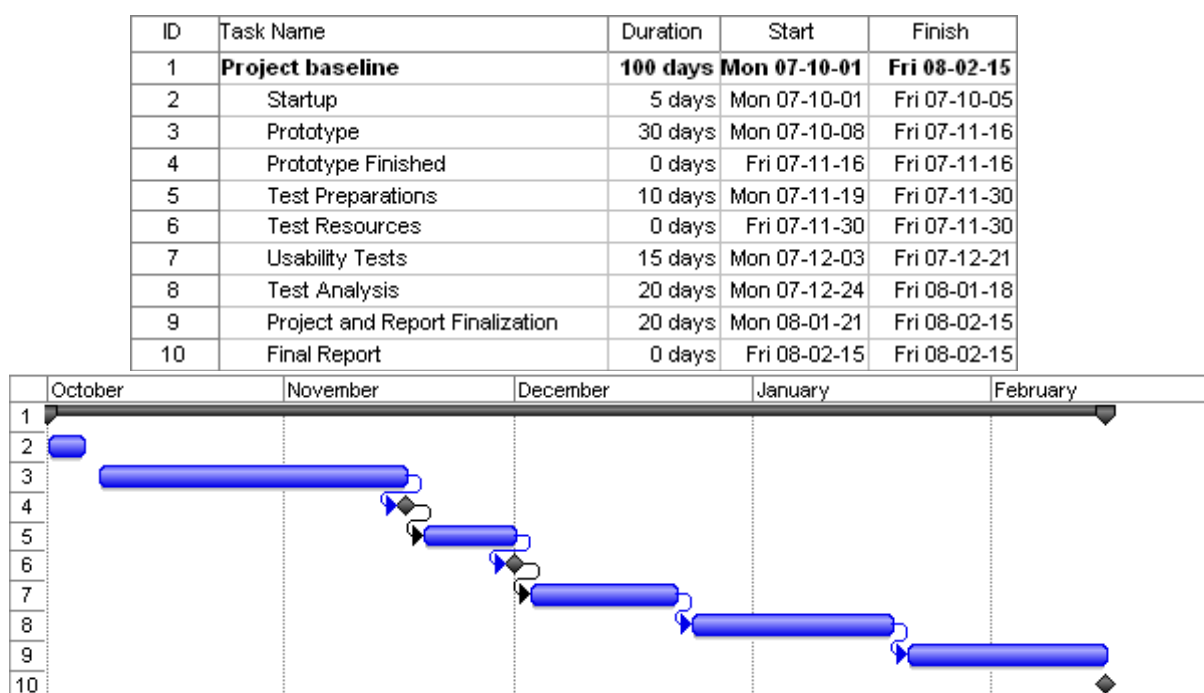
### A.3        Organization

The stakeholders of the project are:

| | |
|---|---|
| Johan Garnolf | Thesis Worker |
| Nils Hallberg | Thesis Worker |
| Elin Andersson, MMS and iMode Messaging section, Sony Ericsson | External Supervisor |
| Maria Rang, MMS and iMode Messaging section, Sony Ericsson | Mission Assigner |
| Joakim Eriksson, Ergonomics and Aerosol Technology department, Faculty of Engineering, Lund University | Internal Supervisor |

### A.4        Activities and Schedule

In the table and diagram below, the time allocated for each activity is presented. The search for and study of literature will be done continuously throughout the project. Note that the deliverable times are deadlines; there is a good possibility that these times will be earlier than expected. Here follows a description of each activity:

1.      Project baseline – Time windows for the whole project.
2.      Start-up – Goals are formulated, the schedule is established and a list of relevant literature is assembled.
3.      Prototype – The prototypes are developed.
4.      Deliverable: Prototype Finished – The prototypes are completed and presented.
5.      Test Preparations – The tests are prepared, i.e. the test method is determined in detail, target groups are established, participants are booked, background questionnaires, user tasks and debriefing questions are formulated et cetera.
6.      Deliverable: Test Resources – All documents and procedures needed during the tests are completed and presented.
7.      Usability Tests – The tests are performed in the usability lab.
8.      Test Analysis – The results from the tests are quantified by analysis of the test data.
9.      Project and Report Finalization – The final report is completed by merging our test results with existing knowledge and a conclusion is formulated based on the initial goals of the project. Also the project presentation is prepared during this period.
10.     Deliverable: Final Report – The final report is delivered to the stakeholders.

| ID | Task Name | Duration | Start | Finish |
|---|---|---|---|---|
| 1 | **Project baseline** | **100 days** | **Mon 07-10-01** | **Fri 08-02-15** |
| 2 | Startup | 5 days | Mon 07-10-01 | Fri 07-10-05 |
| 3 | Prototype | 30 days | Mon 07-10-08 | Fri 07-11-16 |
| 4 | Prototype Finished | 0 days | Fri 07-11-16 | Fri 07-11-16 |
| 5 | Test Preparations | 10 days | Mon 07-11-19 | Fri 07-11-30 |
| 6 | Test Resources | 0 days | Fri 07-11-30 | Fri 07-11-30 |
| 7 | Usability Tests | 15 days | Mon 07-12-03 | Fri 07-12-21 |
| 8 | Test Analysis | 20 days | Mon 07-12-24 | Fri 08-01-18 |
| 9 | Project and Report Finalization | 20 days | Mon 08-01-21 | Fri 08-02-15 |
| 10 | Final Report | 0 days | Fri 08-02-15 | Fri 08-02-15 |



## A.5    Revision history

The following table contains the changes made to this document.

| Date | Change | Responsible | Version |
|---|---|---|---|
| 2007-09-27 | Creation of the document | NH, JG | 1.0 |
| 2007-10-02 | Goals regarding prototypes updated, list of contents and revision history added | JG | 1.1 |
| 2007-10-11 | Goals regarding prototypes updated according to meeting with external supervisor. | NH, JG | 1.2 |
| 2007-12-10 | Revision of the project goals, in accordance with the proceeding work and last meetings with external supervisor. | JG, NH | 1.3 |
| 2007-12-10 | Typos corrected. | JG | 1.4 |

## *B* *Test Documents*

### B.1 Pre-test Questionnaire

Participant nbr: _____

☐ Man　　　☐ Woman

Age: _____

|  | **Poor** |  |  | **Very good** |  |
|---|---|---|---|---|---|
| How would you rate your technology skills? | ☐ | ☐ | ☐ | ☐ | ☐ |

|  | **Not at all** |  |  | **Very** |  |
|---|---|---|---|---|---|
| How interested are you in new technology? | ☐ | ☐ | ☐ | ☐ | ☐ |

What mobile phone do you use today?
_____

How many different mobile phones have you owned?
_____

Which mobile phone brand have you used the most?
_____

If you have mainly used one certain mobile phone brand, what are the reasons for that? Check one or many boxes.
☐ The design ☐ The functions ☐ The battery time
☐ Easy to use ☐ Other: _____

What are your reasons for changing mobile phones?
Check one or many boxes.
☐ Design ☐ New, better functions ☐ Better battery time
☐ "Free" phone ☐ Last phone lost/broken ☐ Easier to use
☐ Variation ☐ Other: _____

What functions do you use on your mobile phone?
Check one or many boxes.
☐ SMS ☐ MMS ☐ E-mail ☐ Camera
☐ Music player ☐ FM radio ☐ Internet ☐ Alarm clock
☐ Calculator ☐ Calendar ☐ Voice calls ☐ Video calls
☐ Games ☐ Notes ☐ Other:_____

What is your mobile phone for you? Check one or many boxes.
☐ Just a phone ☐ Something you need to have because all others have it
☐ Entertainment device ☐ Fashion statement
☐ An indispensable aid (an extension of yourself)
☐ Other:_____

### B.2 Orientation Script

The following script is read out loud to each participant before the test so that each one of them is provided with the same background information

*This test is a part of our Master Thesis for Sony Ericsson. You are here to help us to test new ways of presenting menu systems and text messages.*

*The test will be videotaped so that we will be able to go back and examine details. The movie clips will solely be used by us for test analysis.*

*You may ask questions at any time during the test, but considering the test compilation, it is not sure that we will answer.*

*Do your best, but do not focus too much on your performance. Always remember that it is the products that are being tested, not you.*

After the script is read, the participant is briefed regarding which buttons on the phones that shall be used during the test.

### B.3   *Threaded Messaging Test* Tasks

Participant nbr: _____

Do the following tasks; say the answers out loud as soon as you find them.

1. What did Erik write to you in the SMS that you received October 15, 18:17?

2. Which SMS was the last you wrote to Lisa?

3. You invited Jim to something through SMS.

     a.  What did the invitation regard?
     b.  What did Jim answer?

### B.4 Threaded Messaging Post-test Questionnaire

Participant nbr: _____

Which one of the prototypes was easiest to solve the tasks with? Check the box that seems the most correct when you compare the two prototypes.

If you had to use one of the techniques on your mobile phone in your everyday life, which one would you prefer?



Why?

### B.5 *Usability versus Aesthetics Test* Tasks

Participant nbr: _____

Do the following tasks; say the answers out loud as soon as you find them.

1. You have answered Jim through SMS regarding a tip about something. What did the tip regard?

2. Which SMS was the last you wrote to Anna?

3. You sent an invitation to Sara which did not regard a lunch.

    a. What did the invitation regard?
    b. What did she answer?
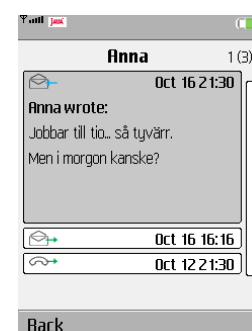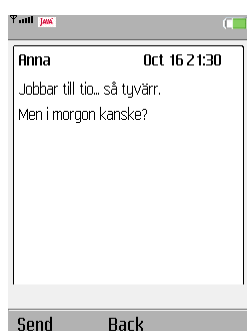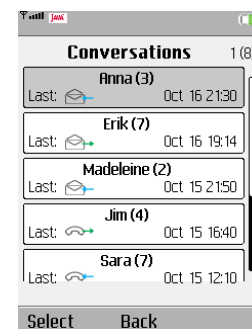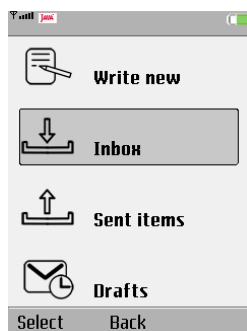
### B.6        Usability versus Aesthetics Post-test Questionnaire

Participant nbr: _____

Which one of the prototypes was easiest to solve the tasks with? Check the box that seems the most correct when you compare the two prototypes.



☐         ☐         ☐         ☐         ☐

Which one of the prototypes did you find the most aesthetically attractive?  Check the box that seems the most correct when you compare the two prototypes.



☐         ☐         ☐         ☐         ☐

How easy to use was the following prototype?



Put a grade between 1 and 5.

Grade: _____

How easy to use was the following prototype?



Put a grade between 1 and 5.

Grade: _____

How aesthetically attractive was the following prototype?



Put a grade between 1 and 5.

Grade: _____

How aesthetically attractive was the following prototype?



Put a grade between 1 and 5.

Grade: _____

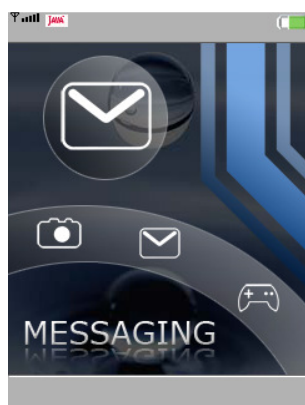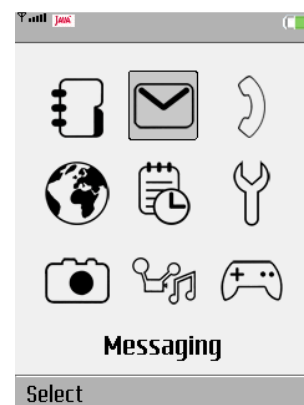If you had to use one of the techniques on your mobile phone in your everyday life, which one would you prefer?



Why?

### B.7       Post-test Questionnaire

Participant nbr: _____

If you would buy a new mobile phone today, please enter how important the following attributes would be for you:

|  | Unimportant |  |  |  | Very important |
|---|---|---|---|---|---|
| Design | ☐ | ☐ | ☐ | ☐ | ☐ |
| Price | ☐ | ☐ | ☐ | ☐ | ☐ |
| Functions | ☐ | ☐ | ☐ | ☐ | ☐ |
| Brand | ☐ | ☐ | ☐ | ☐ | ☐ |
| Easy to use | ☐ | ☐ | ☐ | ☐ | ☐ |

If you would buy a new mobile phone today, please weigh the following attributes against each other:

**Example:** If you find price much more important than design, you would answer like this

Design    ☐    ☐    ☐    ☐    ■    Pris

If both attributes instead are equally important, you would check the middle box.

| Easy to use | ☐ | ☐ | ☐ | ☐ | ☐ | Price |
|---|---|---|---|---|---|---|
| Brand | ☐ | ☐ | ☐ | ☐ | ☐ | Design |
| Design | ☐ | ☐ | ☐ | ☐ | ☐ | Easy to use |
| Price | ☐ | ☐ | ☐ | ☐ | ☐ | Functions |
| Functions | ☐ | ☐ | ☐ | ☐ | ☐ | Brand |
| Easy to use | ☐ | ☐ | ☐ | ☐ | ☐ | Functions |
| Brand | ☐ | ☐ | ☐ | ☐ | ☐ | Easy to use |
| Design | ☐ | ☐ | ☐ | ☐ | ☐ | Price |
| Price | ☐ | ☐ | ☐ | ☐ | ☐ | Brand |
| Functions | ☐ | ☐ | ☐ | ☐ | ☐ | Design |

| Document Type: | Authors: | Version: | Last Change: |
|---|---|---|---|
| Report | Johan Garnolf and Nils Hallberg | 2.6 | 2008-01-23 |

62

# C        Quantitative Data

### C.1 *Threaded Messaging Test* Completion Times

| | Standard Menu | | | | Threaded Messaging Menu | | | |
|---|---|---|---|---|---|---|---|---|
| Participant | Task 1 | Task 2 | Task 3a | Task 3b | Task 1 | Task 2 | Task 3a | Task 3b |
| 1 | 55 | 21 | 34 | 15 | 39 | 22 | 23 | 9 |
| 2 | 26 | 20 | 18 | 25 | 104 | 39 | 37 | 5 |
| 3 | 18 | 21 | 21 | 13 | 20 | 14 | 18 | 8 |
| 4 | 21 | 18 | 16 | 10 | 126 | 22 | 21 | 3 |
| 5 | 26 | 24 | 19 | 15 | 77 | 59 | 26 | 3 |
| 6 | 8 | 13 | 17 | 7 | 103 | 12 | 28 | 2 |
| 7 | 21 | 17 | 13 | 8 | 147 | 27 | 32 | 5 |
| 8 | 13 | 14 | 18 | 9 | 25 | 10 | 15 | 4 |
| 9 | 43 | 43 | 31 | 18 | 136 | 59 | 35 | 9 |
| 10 | 19 | 27 | 20 | 28 | 47 | 47 | 23 | 2 |
| 11 | 38 | 22 | 27 | 20 | 30 | 13 | 14 | 3 |
| 12 | 18 | 21 | 32 | 23 | 73 | 22 | 22 | 3 |
| 13 | 12 | 27 | 16 | 4 | 15 | 8 | 9 | 2 |
| 14 | 14 | 13 | 9 | 7 | 33 | 80 | 16 | 2 |
| 15 | 8 | 15 | 11 | 5 | 32 | 12 | 12 | 5 |
| 16 | 9 | 7 | 11 | 6 | 25 | 20 | 28 | 2 |
| Average: | 21,8125 | 20,1875 | 19,5625 | 13,3125 | 64,5 | 29,125 | 22,4375 | 4,1875 |

### C.2    *Usability versus Aesthetics Test* Completion Times

| | Aesthetic Menu | | | | Threaded Messaging Menu | | | |
|---|---|---|---|---|---|---|---|---|
| **Participant** | **Task 1** | **Task 2** | **Task 3a** | **Task 3b** | **Task 1** | **Task 2** | **Task 3a** | **Task 3b** |
| **1** | 112 | 30 | 36 | 9 | 57 | 16 | 23 | 5 |
| **2** | 108 | 14 | 27 | 5 | 32 | 18 | 30 | 5 |
| **3** | 41 | 34 | 27 | 4 | 19 | 20 | 20 | 2 |
| **4** | 43 | 15 | 29 | 2 | 32 | 16 | 31 | 3 |
| **5** | 75 | 27 | 78 | 2 | 33 | 24 | 33 | 2 |
| **6** | 20 | 16 | 14 | 3 | 15 | 21 | 26 | 2 |
| **7** | 97 | 12 | 34 | 2 | 54 | 22 | 30 | 4 |
| **8** | 68 | 42 | 34 | 5 | 20 | 8 | 21 | 2 |
| **9** | 96 | 23 | 50 | 13 | 72 | 29 | 36 | 4 |
| **10** | 79 | 25 | 48 | 1 | 42 | 40 | 40 | 3 |
| **11** | 91 | 19 | 45 | 2 | 16 | 7 | 13 | 2 |
| **12** | 124 | 21 | 38 | 6 | 29 | 17 | 25 | 3 |
| **13** | 45 | 10 | 23 | 3 | 10 | 8 | 13 | 2 |
| **14** | 35 | 14 | 13 | 2 | 13 | 11 | 19 | 20 |
| **15** | 34 | 11 | 36 | 3 | 18 | 13 | 20 | 2 |
| **16** | 18 | 10 | 21 | 3 | 23 | 10 | 19 | 3 |
| **Average:** | **67,875** | **20,1875** | **34,5625** | **4,0625** | **30,3125** | **17,5** | **24,9375** | **4** |

## C.3    Participant Information

| Participant | Age | Age group | Gender |
|---|---|---|---|
| 1 | 59 | 4 | F |
| 2 | 43 | 3 | F |
| 3 | 32 | 2 | F |
| 4 | 36 | 3 | M |
| 5 | 58 | 4 | F |
| 6 | 26 | 2 | M |
| 7 | 23 | 2 | F |
| 8 | 42 | 3 | M |
| 9 | 51 | 3 | F |
| 10 | 62 | 4 | M |
| 11 | 33 | 2 | M |
| 12 | 57 | 4 | M |
| 13 | 15 | 1 | M |
| 14 | 15 | 1 | M |
| 15 | 15 | 1 | F |
| 16 | 14 | 1 | F |

## C.4  Functions Used

| Participant | Aesthetic Menu chosen | Threaded Messaging Menu chosen |
|---|---|---|
| 1 | | 6 |
| 2 | | 4 |
| 3 | 6 | |
| 4 | | 2 |
| 5 | 1 | |
| 6 | | 2 |
| 7 | | 5 |
| 8 | | 3 |
| 9 | 1 | |
| 10 | 1 | |
| 11 | | 1 |
| 12 | | 4 |
| 13 | 8 | |
| 14 | 9 | |
| 15 | 6 | |
| 16 | 6 | |
| Average: | 4,75 | 3,375 |

### C.5 *Threaded Messaging Test* Usability Scores

| Participant | Score | Standard Menu chosen | Threaded Messaging Menu chosen |
|---|---|---|---|
| 1 | -50 | -50 | |
| 2 | -100 | | -100 |
| 3 | 50 | | 50 |
| 4 | -100 | -100 | |
| 5 | 100 | | 100 |
| 6 | -100 | -100 | |
| 7 | 50 | | 50 |
| 8 | 100 | | 100 |
| 9 | -50 | | -50 |
| 10 | -100 | -100 | |
| 11 | 50 | | 50 |
| 12 | -50 | | -50 |
| 13 | -50 | -50 | |
| 14 | -100 | -100 | |
| 15 | -100 | -100 | |
| 16 | -100 | -100 | |
| **Average:** | **-34,375** | **-87,5** | **18,75** |

### C.6 *Usability versus Aesthetics Test* Usability and Aesthetics Scores

| Participant | Score Usability | Score Aesthetics | Aesthetic Menu chosen | | Threaded Messaging Menu chosen | |
|---|---|---|---|---|---|---|
| | | | Score Usability | Score Aesthetics | Score Usability | Score Aesthetics |
| 1 | -50 | 100 | | | -50 | 100 |
| 2 | -100 | -50 | | | -100 | -50 |
| 3 | 0 | 100 | 0 | 100 | | |
| 4 | -100 | -100 | | | -100 | -100 |
| 5 | -100 | 100 | -100 | 100 | | |
| 6 | 0 | 50 | | | 0 | 50 |
| 7 | -100 | 100 | | | -100 | 100 |
| 8 | -100 | -100 | | | -100 | -100 |
| 9 | 0 | 0 | 0 | 0 | | |
| 10 | 0 | 100 | 0 | 100 | | |
| 11 | -100 | 100 | | | -100 | 100 |
| 12 | -50 | -100 | | | -50 | -100 |
| 13 | -50 | 100 | -50 | 100 | | |
| 14 | 100 | 100 | 100 | 100 | | |
| 15 | -50 | 100 | -50 | 100 | | |
| 16 | 0 | 50 | 0 | 50 | | |
| **Average** | **-43,75** | **40,625** | **-12,5** | **81,25** | **-75** | **0** |

### C.7    *Usability versus Aesthetics Test* Usability and Aesthetics Grades

| | Aesthetic Menu | | Threaded Messaging Menu | |
|---|---|---|---|---|
| Participant | Usability | Aesthetics | Usability | Aesthetics |
| 1 | 4 | 4 | 5 | 3 |
| 2 | 2 | 3 | 5 | 3 |
| 3 | 4 | 5 | 5 | 2 |
| 4 | 3 | 2 | 5 | 3 |
| 5 | 3 | 5 | 5 | 3 |
| 6 | 3 | 3 | 3 | 3 |
| 7 | 2 | 5 | 3 | 3 |
| 8 | 2 | 3 | 5 | 5 |
| 9 | 3 | 4 | 2 | 3 |
| 10 | 4 | 4 | 4 | 3 |
| 11 | 2 | 4 | 4 | 2 |
| 12 | 3 | 2 | 5 | 3 |
| 13 | 4 | 5 | 4 | 3 |
| 14 | 4 | 5 | 4 | 3 |
| 15 | 3 | 4 | 2 | 3 |
| 16 | 3 | 4 | 4 | 3 |
| Average: | 3,0625 | 3,875 | 4,0625 | 3 |

## C.8    Analytical Hierarchical Process Results (AHP)

| Participant | Design | Price | Functions | Brand | Ease of use |
|---|---|---|---|---|---|
| 1 | 0,218 | 0,101 | 0,095 | 0,096 | 0,489 |
| 2 | 0,105 | 0,152 | 0,235 | 0,046 | 0,462 |
| 3 | 0,318 | 0,126 | 0,144 | 0,318 | 0,093 |
| 4 | 0,133 | 0,083 | 0,369 | 0,076 | 0,338 |
| 5 | 0,116 | 0,24 | 0,112 | 0,045 | 0,487 |
| 6 | 0,128 | 0,45 | 0,089 | 0,067 | 0,266 |
| 7 | 0,072 | 0,455 | 0,216 | 0,067 | 0,19 |
| 8 | 0,246 | 0,137 | 0,068 | 0,167 | 0,383 |
| 9 | 0,359 | 0,086 | 0,079 | 0,079 | 0,395 |
| 10 | 0,11 | 0,177 | 0,184 | 0,052 | 0,477 |
| 11 | 0,205 | 0,069 | 0,145 | 0,069 | 0,512 |
| 12 | 0,222 | 0,176 | 0,084 | 0,145 | 0,374 |
| 13 | 0,393 | 0,113 | 0,198 | 0,183 | 0,113 |
| 14 | 0,116 | 0,159 | 0,265 | 0,104 | 0,356 |
| 15 | 0,082 | 0,101 | 0,367 | 0,082 | 0,367 |
| 16 | 0,095 | 0,085 | 0,486 | 0,085 | 0,25 |
| **Average:** | **0,182375** | **0,169375** | **0,196** | **0,1050625** | **0,347** |