For Happy Users, press 1

Investigating and improving the usability of a touch-tone interface

*Susanne Wedin * Kristina Carlander

LIU-KOGVET-D--04/15--SE
2004-08-24

Masters Thesis in Cognitive Science
Supervisor and Examiner: Pär Carlshamre
Department of Computer and Information Science
Linköpings universitet, Sweden
Abstract

Touch-tone interfaces are today widely used in help-centers and support services. Studies have shown that interfaces like these have many limitations and are therefore hard to design. MVAS is a voicemail interface using touch-tone input for navigation. Today, shortcomings in the interface limit the users’ ability to use the functionality in a satisfying way. This thesis describes a mainly qualitative study which evaluates and tests the interface of MVAS to come up with how the interface should be designed to be easier to use. The results show that the usability of MVAS is poor but the functionality of the same is both impressive and appreciated. The suggested redesign of the system, based on the identified usability problems, considers both the interaction model used in the interface as well as the conformity to the set of heuristics used in the evaluation. The proposed redesign keeps all the functionality in the system intact and also makes the functionality more explicit through improving the usability. A more explicit structure will facilitate usage of a larger portion of the functionality. However, the limitation of the key-pad affects the redesign so the most favorable design is unreachable. If the interaction model is changed or furthered developed to allow speech input the limitations experienced with the current redesign will diminish and a higher degree of usability can be reached.

Keywords:
Acknowledgements

Thanks to all employees at Mobeon for making us feel welcome and for your interest in our work. Especially thank to all who helped us recruit participants to the test and the TUI-team helping us with technical issues regarding MVAS.

Thanks to Andreas Wiencke and Kristian Ryhd for your patience and support during the production of this thesis. Also, thanks to Jamie-Lee Thorpe and Graham Fowler for helping us proof-reading and to Therese Lönnborg for providing great comments and ideas.

Finally we want to thank our supervisors; Magnus Björkman for giving us an inside view of MVAS, and Pär Carlshamre for encouragement and good ideas regarding the production of this thesis.

Kristina and Susanne

Sundsvall, 13th of May, 2004
# Table of Content

1 **Introduction** ................................................................................................. 1  
  1.1 Purpose ........................................................................................................ 2  
  1.2 Delimitations ............................................................................................... 2  
  1.3 Targeted Readers ....................................................................................... 2  
  1.4 Report Overview ........................................................................................ 2  

2 **Background** .................................................................................................. 5  
  2.1 History of Touch-Tone Telephones .............................................................. 5  
  2.2 Touch-Tone Systems .................................................................................. 5  
    2.2.1 Voicemail Applications ...................................................................... 6  
  2.3 M3: The Messaging Solution ...................................................................... 7  
    2.3.1 MVAS: the Telephone User Interface ................................................. 7  

3 **Theoretical Framework** .................................................................................. 9  
  3.1 What is Usability? ....................................................................................... 9  
  3.2 Why is Usability Important ....................................................................... 10  
  3.3 Usability Research ..................................................................................... 10  
    3.3.1 Mental Workload ............................................................................... 10  
    3.3.2 Memory ............................................................................................... 11  
    3.3.3 Mental Models ................................................................................... 11  
    3.3.4 Social Context ................................................................................... 12  
    3.3.5 Design Principles .............................................................................. 13  
    3.3.6 Organization of Menus ...................................................................... 16  
    3.3.7 Help Functions .................................................................................. 18  
  3.4 The Usability Process .................................................................................. 18  
  3.5 System Evaluation and Testing .................................................................. 19  
    3.5.1 Usability Evaluation ......................................................................... 19  
    3.5.2 Usability Testing ............................................................................... 21  
    3.5.3 Combining Evaluation and Testing ................................................... 24  
  3.6 Related Research ....................................................................................... 25  
    3.6.1 Auditory versus Graphical Interfaces .......................................... 25  
    3.6.2 The Phoneshell Study ....................................................................... 26
# TABLE OF CONTENT

3.6.3 Speech Recognition ................................................................. 27
3.6.4 Skip and Scan ................................................................. 27

## 4 Method ................................................................. 29
4.1 Usability Evaluation ................................................................. 29
4.2 Usability Testing ................................................................. 30
  4.2.1 Design ................................................................. 30
  4.2.2 Material ................................................................. 33
  4.2.3 Pretest ................................................................. 37
  4.2.4 Procedure ................................................................. 38
  4.2.5 Apparatus ................................................................. 39
  4.2.6 Classifying the Material ................................................................. 40
  4.2.7 Method for Analysis ................................................................. 42

## 5 Analysis ................................................................. 43
5.1 Evaluation Analysis ................................................................. 43
  5.1.1 Identified Usability Problems ................................................................. 44
5.2 Testing Analysis ................................................................. 51
  5.2.1 Qualitative Analysis ................................................................. 51
  5.2.2 Quantitative Analysis ................................................................. 57
5.3 Summary ................................................................. 63

## 6 Redesign ................................................................. 65
6.1 Example-Changes ................................................................. 65
6.2 Prototypes ................................................................. 68
  6.2.1 Skip and Scan Model ................................................................. 69
  6.2.2 Standard Menu-Style Model ................................................................. 71
  6.2.3 Comparison of the Two Models ................................................................. 73
  6.2.4 Conclusion ................................................................. 75
6.3 Redesign: System Description ................................................................. 76
  6.3.1 Scenario 1 ................................................................. 77
  6.3.2 Scenario 2 ................................................................. 78
  6.3.3 Scenario 3 ................................................................. 80
6.4 Summary of Main Changes ................................................................. 81

## 7 Method Discussion ................................................................. 85
7.1 Design ................................................................. 85
7.2 Usability Evaluation ................................................................. 86
7.3 Usability Testing ................................................................. 86
  7.3.1 Background Questionnaire ................................................................. 86
  7.3.2 Walkthrough ................................................................. 87
  7.3.3 System Settings ................................................................. 88
7.4 Usability study ................................................................. 88
7.5 Generalization ................................................................. 89
# Table of Content

8 Future Work and Conclusion ........................................................................................................ 91
  8.1 Future Work ......................................................................................................................... 91
  8.2 Conclusion ............................................................................................................................. 92

9 References ................................................................................................................................. 93

**Figures**
- Figure 1: Traditional IVR-interface ......................................................................................... 7
- Figure 2: Tree structure – hierarchy. ......................................................................................... 18
- Figure 3: The design of the study .............................................................................................. 31
- Figure 4: Test room settings ....................................................................................................... 38
- Figure 5: The grand mean of the ratings .................................................................................. 60
- Figure 6: The participants’ performance on each scenario ...................................................... 62
- Figure 7: Overview of the key-pad in the skip and scan menu-style ....................................... 70
- Figure 8: Graphical view of the skip and scan menu-style ....................................................... 70
- Figure 9: Graphical view of the Standard menu-style ............................................................. 72

**Tables**
- Table 1: The main usability problems identified in MVAS 10.3.0 ........................................ 64
- Table 2: The main usability problems identified in MVAS 10.3.0 and the proposed changes to correct them ........................................................................................................ 83

**Appendix**
- Appendix A: Instructions .......................................................................................................... 99
- Appendix B: Background Questionnaire .................................................................................. 101
- Appendix C: Walkthrough Heuristics ..................................................................................... 102
- Appendix D: Scenarios .............................................................................................................. 103
- Appendix E: In-depth Interview ............................................................................................... 105
- Appendix F: Ratings .................................................................................................................. 107
- Appendix G: Successful Completion Criteria ........................................................................ 108
- Appendix H: Redesign of the Global Key Pause ..................................................................... 109
1 Introduction

In recent years a new focus has evolved within the human-computer interaction (HCI) community. Errors that occur are no longer seen as the humans fault, but rather as shortcomings in how the computers’ interfaces are designed. The amount of literature written in the area is enormous, but there is still a lot left to do. Many researchers present tools for investigating the usability of telecom products. Nielsen (1993) gives a set of principles on how usability investigations should be conducted. Even though these principles exist, every product is individual, and therefore these principles need their own interpretation and investigation to be applicable to the product of interest.

In 2003 Mobeon AB was founded through the unification of a design center, which until this date was included in the Ericsson concern, and Mobeon, a company which were involved developing software products. Mobeon AB is today developing messaging solutions, such as voicemail applications. Mobeon provide more than 19 million people around the world with an IP-based messaging solution, called M3.

M3 is used mainly through its telephony user interface, MVAS. The ambition behind MVAS is to be a telephony user interface that users are willing to use in their everyday life. With such an ambition it is important to start with a focus on how the system is used today and what bottlenecks that might prevent good usage of the system. It is also important to investigate what actions could be taken to improve the system from a usability point of view. The usability is important to be able to compete on the market and to make the users happy.
INTRODUCTION

1.1 Purpose
The purpose of this thesis is to investigate and improve the usability of the voicemail application, MVAS 10.3.0 produced by Mobeon AB. This is done through usability testing with real end-users and usability-evaluation based on heuristics.

The project includes gaining an understanding of how the users perceive the system and what parts of the system that they find easy or difficult to use. Through this purpose the following problem statements have been formulated:

- How does the end-user perceive the usability of MVAS?
- How should MVAS be designed to be easy to use?

1.2 Delimitations
There exist many versions of MVAS but this thesis is limited to MVAS 10.3.0 and investigates only the voicemail feature of it. Only the retrieval dialogue of the system was included in the test. The system tested is in Swedish which limited the scope of participants to only those with Swedish as their native tongue.

1.3 Targeted Readers
This thesis is mainly written for those interested in usability testing and evaluation of telecom products.

1.4 Report Overview
The Background gives a brief history review and background of touch-tone systems in general and the system investigated in this study. The Theoretical Framework then defines the term usability and describes human characteristics such as mental workload and mental models which play a part of the perceived usability of a system. Thereafter it is described how usability testing and evaluation should be conducted. The chapter ends with a short briefing about related research.

In the Method a review of how the usability was tested and evaluated is given. The data gathered from the testing and evaluation is analyzed in the chapter Analysis. Thereafter the analyzed data is used as a
foundation for the *Redesign*, where the proposed redesign for the system is described and illustrated. The thesis is rounded off with a *Method Discussion* followed by a chapter called *Future Research and Conclusion* which describes how this thesis could be followed up with future research and also the conclusion of the thesis.
2 Background

This chapter reviews some history regarding touch-tone phones and applications. It ends with a brief introduction and description of the application investigated in this study.

2.1 History of Touch-Tone Telephones

The first touch-tone telephone, which used tones in the voice frequency range was installed in Baltimore in 1941. Operators in a central switching office pushed the buttons. This fact made it much too expensive for general use. Despite this, touch-tone phones were still developed because it increased the speed of dialing. By the early 1960s, low-cost transistors and associated circuit components made the introduction of touch-tone into home telephones possible. Extensive human factor-testing determined the position of the buttons to limit errors and to increase dialing speed even further (AT & T Labs-Research, 2004).

2.2 Touch-Tone Systems

A touch-tone system is an application with input from the telephone keypad coupled with speech output from the connected application, for example voicemail.

Today most systems using touch-tone input are developed for commercial use. The idea is to help free customer service centers from simple requirements, and repetitive calls to gain more time for challenging inquiries. Touch-tone systems also extend the hours when customers can get information or help. When designed properly, these
systems are supposed to deliver customer satisfaction and improve overall productivity at lower costs (Claritus, 2004).

Touch-tone systems have been a great success although many users are frustrated by the lengthy and deep menu structures, or by long informational speeches in which it seems that the needed fact is always at the end or omitted (Shneiderman, 1998).

While much is made of peoples dislike for automated touch-tone systems, the fact is that most callers will go for them every time if presented with the prospect of being left on hold to speak to customer services (Claritus, 2004).

Today the development of help system for customer service centers are more focused on a more expensive alternative, speech recognition. Speech recognition allows an application to understand human speech and transfer it into digital or analog signals. This means that both the input and output is spoken. Like touch-tone systems, speech recognition systems need more research within the area of usability to be able to fully satisfy the users’ requirements (Jurafsky et al. 2000, Weinschenk et al. 2000).

2.2.1 Voicemail Applications
A voicemail is a message that is left in the callers’ own voice and is retrieved by the user from a touch-tone phone.

In the late seventies Gordon Matthews began working on the technology that would eventually be called voicemail. He applied for a patent for his voicemail invention in 1979 and sold the first system to 3M the same year. (800 Voice Mail Store, 2004).

Voicemail is today used across the world by millions of users. The systems are getting more sophisticated, containing more enhanced features. These features lead to larger menus, with more options that make the systems harder to handle and more time-consuming for the end-users. The fast development of voicemail systems makes it important to repeatedly test the usability of these systems.
2.3 **M3: The Messaging Solution**

Mobeon’s M3 series is a IP messaging solution with over 19 million voice mail boxes sold in five continents; Europe, Africa, Asia, and North- and South America.

M3 is a unified messaging product. Unified messaging is a concept introduced in the late 1990s in which a single mailbox is provided to combine multiple message types with distant access from multiple device types (Mobeon AB, 2004). For example; in M3 this means that email, fax and voicemail all can be accessed from one single device.

M3 is built on IP messaging technology, which means that the messaging platform is fully based on the internet protocol. All connections to the telecommunication network are converted directly to IP removing the need for proprietary hardware or digital signal processing-devices. IP based messaging minimizes the time that a connection is maintained between two computers, which reduces the load on the network. It also frees up the two computers communicating with each other so that they can accept information from other computers as well (Mobeon AB, 2004).

### 2.3.1 MVAS: the Telephone User Interface

MVAS is the telephone user interface (TUI) of M3. The structure of MVAS can be referred to as a standard Interactive Voice Response (IVR) interface. This can be represented as a tree of nodes with the sound of each node being prompts telling the user what buttons to press to go to other nodes, see Figure 1, (Resnick and Virzi, 1992).

![Figure 1: Traditional IVR-interface.](image-url)
MVAS is structured according to a 3-option model. Throughout the system the menus have three options responding to, for example, keys 1, 2, 3. The rest of the options in the menu are “hidden” under key 0. This means that if a menu contains five options the first three options are presented associated to key 1, 2 and 3 and then the system say “For more options, press 0”. If the user presses 0, the options at key 4 and 5 will be presented.

With MVAS users can send and receive voicemail, listen to e-mails being read by a synthesized voice and print fax messages. The users can also amongst other things record different greetings and change their messaging notifications.

Below, an extract from the dialogue of MVAS 10.3.0 is presented.

1. System: Main menu
   If you want your saved messages now, press 1
   If you want to send a voice message, press 2
   To work with your absence greetings, press 3
   For more options, press 0

2. User: 1 (saved messages)

3. System: To get your voice messages, press 1
   To get your faxes, press 2
   To get your e-mails, press 3

4. User: 1 (voice messages)

5. System: You have “1” saved voice message
   Message from *name of the sender*
   *Time for deposit*
   *Message is played*
   To listen to the message, again press 1
3 Theoretical Framework

This chapter contains theory about usability, including when a usability study should be conducted as well as how. The chapter also includes theory about some human characteristics playing an important part in how a user perceives a system. It ends with a review of some related usability studies on auditory user interfaces.

3.1 What is Usability?

There is no widely acknowledged definition of usability that everyone in the usability community would agree on. Despite this it is generally accepted that the term usability includes one or more of the following four factors; Usefulness, Effectiveness, Learnability, and/or Attitude (Booth, 1989). The factors are briefly explained below.

- Usefulness concerns the degree to which a system enables a user to achieve his or her goals, and the user’s motivation of using the system at all.
- Effectiveness concerns the systems’ ease of use, and how good the system is at doing what it is supposed to do.
- Learnability has to do with the user’s ability to operate the system to some defined level of competence after an amount of training.
- Attitude refers to the user’s perceptions, feelings, and opinions of the system.

The International Organization for Standardization (ISO) defines usability as;
“...the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in particular environments...” (ISO DIS 9241-11, Faulkner, 2000)
In conclusion, what people think of concerning the term usability seems to differ but what it all comes down to in the end can be summarized with this quote made by Dumas and Redish (1999);

“Usability means making users productive and happy.”

3.2 Why is Usability Important

Today products supported by computers are used in all kinds of contexts. People who use these products might not know anything about how the product actually work, the only thing they care about is that it works, in the right way and in the right situation. What the users want is to get the task done, not interacting with the system itself. (Wilson, 1999).

Since the users are not experts in how technical applications work, it is important to make sure that they are able to use these applications successfully. When more and more features and functionality are added to the applications it is important to make sure that the usability is kept intact (Faulkner, 1998). If the users pay money for applications with many features, the designers should strive for making the users able to use all these features.

3.3 Usability Research

When doing usability research there are three components to take into consideration. These are the human, the computer, and the interaction between the two. Below characteristics playing an important part in how the users perceive a system are reviewed. Thereafter it is described how an interface should be designed and organized to be usable.

3.3.1 Mental Workload

Mental workload is the system’s demand placed upon humans. People’s perceptions of their own mental workload are their subjective experiences which are task specific and person specific (Rouse et al., 1993). According to Macdonald (1999) the level of workload is the interaction between the operator and the work task characteristics. It is important to minimize the mental workload put on the users, so that
They fully can focus on accomplishing their task without being distracted by unnecessary problems. Mental workload is known to have a negative effect on the user's stress level. If the user perceives a low sense of control over the system this can increase the stress levels, and hence increase the user's mental workload (Macdonald, 1999).

One way to reduce mental workload has been through automation of the systems. This may reduce the number of discrete tasks to be performed by the user, but not necessarily decrease the mental workload as a consequence (Macdonald, 1999).

### 3.3.2 Memory
Lack of consistency in a system interface is a problem when it comes to the demand the system puts on the users' memory. In auditory interfaces, serial presentation frustrates users and put demands on their attention resources. As soon as a word is spoken the users must try to remember what have been said and still pay attention to what is coming. Although this place a large demand on the users, auditory output in interfaces can be argued to be a good choice since the short-term memory seems to be primarily acoustic (Baddeley, 1966).

Memory research shows that the shorter the prompts are the more likely the users are to remember them. George Miller's classic paper, “The Magical Number Seven-Plus or Minus Two,” identified the limited capacities people have for absorbing information (Miller, 1956). People can rapidly recognize approximately seven chunks of information at a time and can hold these chunks in short-term memory for 15 to 30 seconds. The size of a chunk of information depends on the person's familiarity with the material. The recall theory shows that the more time that passes between the prompts and menus, the more likely users are to forget which key they where supposed to press (Wickens and Hollands, 2000).

### 3.3.3 Mental Models
A mental model can be viewed as a user's internal working model of the system that is less detailed and more concrete than the system itself, as the user understands it (Sternberg, 1999).
The users’ mental models of the system change over time through progress in learning the system. When a mental model of an interactive system is created by the users, it is assumed that it will be used to make inferences about how to carry out tasks in the system. Mental models are also used to understand what to do when something unexpected happens with a system and when encountering unfamiliar systems (Preece et al., 2002).

Users do not believe, or at least do not want to believe, that they make mistakes when they are using a system. Therefore their mental model does not include their own errors. Hence the users’ mental model means free the users of blame. The solution for the interface designer is to completely abandon the idea that the users can make mistakes. Meaning that everything the users do is something they consider to be valid and reasonable. (Cooper and Reiman, 2003)

In a study by Halasz and Moran (1983) two groups were tested concerning mental models. One group was presented to a model of the system while the other group learned the system by themselves. When performing tasks in the system there were no differences between the two groups, at least while the tasks were already known. When trying to conduct new tasks the group which had seen a model performed better. It was argued that the model helped users to construct a better problem space in which they could carry out the problem-solving process necessary for creative solutions (Wilson 1999).

3.3.4 Social Context

When doing usability research it is important to take the individual characteristics as well as the context surrounding the user into consideration. Information products should embody the capability to interact with the user in all environments, in school, at home, at the marketplace etcetera. Usability in such nontraditional usage contexts is likely to prove a harder target to meet than in case of the workplace (Stephanidis and Akoumianakis, 1996). User interfaces have to be suitable for everybody and therefore it is necessary for the information artifacts to capture the variations between users and within individual users as they change over time (Wilson, 1999). Information artifacts
should not only support more effective and efficient user interaction, but also address the individual end-user requirements and expectations (Stephanidis, 2001).

Our world is becoming more and more international. A system may be developed in one or many countries but used by people with different cultural backgrounds and nationalities. A system is expected to work equally well for every person that uses it, no matter background. This might lead to that the system has to be modified for some specific users to be able to fully satisfy their needs. (Sun, 2002)

3.3.5 Design Principles
When it comes to the computers’ part of the interaction with the human, there exists several high-level principles concerning usable interfaces (e.g. Nielsen, 1993). These are;

- Consistency
- Feedback
- Error messages
- Visibility
- Minimize users’ memory load
- Shortcuts/Flexibility

All of these principles contribute to each other. Therefore they should not be separated but instead be considered together as an entirety. The principles will be further described below.

Consistency
Consistency refers to designing interfaces so they have similar operations and use similar elements for achieving similar tasks (Preece et al., 2002). This principle is based on research that shows that people learn faster and can transfer what they learn better when what they see and do is consistent (Teitelbaum and Granda, 1983). The same information should be presented in the same location and it should be formatted in the same way to facilitate recognition. Consistency is not just a question of layout design but also includes considerations on how the system should be structured (Nielsen 1993).
In a graphical interface consistency means that layout, terminology and structure are the same throughout the entire interface, creating a feeling of familiarity. When it comes to touch-tone interfaces the consistency is even more important, because the user only has auditory information to rely on. Consistency in a touch-tone system is created through using terminology that makes sense to the user and is the same throughout the entire system. It is also important that the system’s structure is consistent to make it possible for the users to create a mental model of it.

**Feedback**

Donald Norman (2001) describes feedback as the way to inform the users about what action that has actually been performed. Feedback must be accordingly efficient and appropriate. System feedback should not be expressed in abstract and general terms but rather restate and rephrase the user’s input to indicate what is happening. In the case of a system failure, informational feedback should also be given (Nielsen 1993). The conclusion is that users never should be left in any doubt about the state of the system they are working with (Faulkner, 2000).

Feedback in a graphical interface is fairly easy to create through visual effects, for example; when a button is pressed its appearance can change to give the users feedback that an action has been executed. Feedback in touch-tone systems is more complicated. The users have to understand that something has happened as a result of their actions, but at the same time this feedback has to be carefully considered not to annoy the users. Feedback in form of a visual effect is less interrupting than auditory feedback which the users have to listen to before proceeding.

**Error Messages**

Good error messages should be precise rather than vague or general. They should also be polite, not intimidating, and not put the blame explicitly on the user (Nielsen, 1993). Error messages should indicate the problem and explain how to recover from it (Faulkner, 2000). Ideally, they should be treated as how-to-fix-it messages. Instead of explicating what has happened, they should state the cause of the problem and what the users needs to do to fix it (Preece et al., 2002). The modality of the error message does not matter, the problem lies in formulating the message in a good way.
Visibility
Since interfaces are based on recognition they rely, to a great extent, on the ease of visibility. Exposing too many objects and attributes will result in a relative loss of salience for the ones of interest to the user. Therefore care should be taken to match object visibility as much as possible with the user’s needs (Gilmore, 1991).

Using feedback in the right way can also provide the necessary visibility for user interaction (Preece et al., 2002). A system with good visibility should provide the users with the essential information to make it obvious what functionality the system offers. The term transparency is sometimes mentioned in these contexts. It refers to the systems ability to show the users what is going on inside, not just input and output but also what happens in between (Löwgren and Stolterman, 1998). This matter should be considered in both graphical and auditory, such as touch-tone, interfaces.

Minimize Users’ Memory Load
To minimize the users’ memory load, the system should be governed with a small number of rules that apply throughout the entire user interface (Nielsen, 1993). This makes it easier to remember and predict the system’s behavior. If the system is not governed by rules at all, the users will have to learn every single dialogue element by heart, which will increase their memory load.

The use of global keys is one way to let a few rules govern a complex system (Rosenberg and Morgan, 1984). Global keys make it sufficient for users to learn a few commands in order to manipulate the system in every context. One of the main advantages of global keys is that they support transfer of learning from one context in the system to the other; hence users can use commands that they already know (Nielsen, 1993). The global keys in a graphical interface are often easy to retrieve, because of their constant position. But in a touch-tone interface, the commands are only given at certain times and the users have to remember them until they are needed, which increases their memory load (see 3.3.2).
A way to minimize the users’ memory load in a touch-tone interface is to not present too many options at the same time. It is easier and faster to scan a list of options with the eyes than listening to them (Schmandt, 1993). Also the memory load is decreased if the users are provided with sufficient feedback.

**Shortcuts/Flexibility**

The interface has to support both inexperienced and experienced users (Preece et al., 2002). It should be possible for the experienced user to perform frequently used operations fast, using dialogue shortcuts (Nielsen, 1993). A good way to achieve this is to have a “settings” menu or similar to allow the user to personalize both the system functionality and appearance.

In touch-tone interfaces shortcuts can be referred to when the users are able to interrupt prompts, or use global keys. If the users often select the same options they will eventually learn the options by heart and can make the options without listening to the information every time. In graphical interfaces the users can learn the options by heart but the time to reach the goal will not be much faster because the flow of the system can not be interrupted in the same way as in a touch-tone interface.

### 3.3.6 Organization of Menus

There are many different ways in which menus can be organized in an interface. Different types of menus can be suitable in different systems and contexts.

Menus of any type are always used as a way to present a list of options to the user. The user should not be given too many options at once, in a way to minimize the users’ memory load (see section 3.3.1). Most research indicates that three or four is the optimal number of options to be presented in a menu (Engelbeck and Roberts, 1990).

The presentation sequence of the items in a menu is another important thing to consider when making the menus. If the items in a menu have a natural order, such as days of the week, then the decision is trivial. But in many cases the items have no task-related ordering and then the
designer must choose another way to present the items. (Shneiderman, 1998)

There are four common ways in which the items of a menu can be presented. These are;

- Alphabetic sequence of terms
- Grouping of related items
- Most frequent used items first
- Most important items first

Choosing titles for menus is a complex matter that deserves serious thought (Shneiderman, 1998). A menu name should say something about what it contains and also help guide the user to the right menu. Another complex matter is the phrasing of menu items. This matter has no perfect solution but there are some main points to take into consideration when phrasing menu items. These are; to carefully select terminology that is familiar to the designated user community, to make sure that each item in the menu is clearly distinguished from others, and to review the collection of items to ensure consistency and conciseness. It is also important to not use words that intimidate the users. All this makes the users feel more comfortable and become more successful in their performance.

In touch-tone interfaces there is another thing to consider concerning menu organization. Resnick and Virzi (1992) discuss whether the prompts in touch-tone menus should be presented in key-action order “Press 1 for X” or in action-key order “For X, press 1”. The most recent research indicates that action-key order is preferable (ibid). The users tend to listen after what they want in the menu, ignoring what comes before until they hear the right option. Therefore users will remember the key better if it is presented after the option than if it is presented before.

Menus are often hierarchical, which means that the menus are organized into a tree structure, see figure 2. A menu option may have submenus with more options and each of these options has its own submenus and so on. The hierarchy can become infinitely deep.
According to Cooper and Reimann (2003) abstract hierarchies are often very difficult for users to navigate.

![Figure 2: Tree structure – hierarchy.](image)

Shneiderman (1998) describes several empirical studies which have dealt with the depth-breadth tradeoff, and the evidence is strong that breadth should be preferred over depth. In fact there is reason to encourage designers to limit menu trees to three levels. When the depth goes to four or five levels, there is a good chance of users becoming lost or disoriented.

### 3.3.7 Help Functions

Developers are usually not very interested in writing help texts, they are more interested in developing the system (Dumas and Redish, 1999). They may not realize that the help function is as important as any other part of the system. This means that in general too little effort and time is allocated to developing the help function, which will result in help functions that are insufficient. Developers also tend to use a very technical vocabulary that scares many users after reading just a few lines of text.

The major problem is that users often are reluctant to use help functions. The users do not want to spend time exploring the help functionality, they just want to solve the problem as quickly as possible.

### 3.4 The Usability Process

It is concluded by many researchers (Stephanidis, 2001, Hackos and Redish 1998, Rubin 1994) that the sooner usability is introduced in the system development cycle, the better. When usability research is done late in the cycle only small changes and improvements can be done on the system’s structure and interface. Also, considering an economical
point of view it costs more to make changes late in the development cycle.

Usability research is most powerful and most effective when implemented as a part of an iterative development process. That is; to test, measure, redesign, and then test again (Gould and Lewis, 1985). It is always better to make a good system from the beginning than to improve it with expensive help documentation (Hackos and Redish, 1998).

Studies have shown that programming effort regarding usability, for example developing user interfaces, often exceeds 50% of the total programming effort required for the entire system (Stephanidis, 2001). If that much effort is required it is important to do it right from the beginning.

### 3.5 System Evaluation and Testing

There are many ways to test and evaluate user interfaces in regards to usability. In the following sections some methods how to conduct evaluation and testing will be reviewed.

#### 3.5.1 Usability Evaluation

Usability evaluation, also called expert evaluation, is conducted by researchers who investigate the interface and predict problems users would have when interacting with it. Typically these techniques are relatively easy to learn as well as effective, which makes them appealing. In addition they can be used in any stage of a design project (Preece et al., 2002).

**Heuristic Evaluation**

Heuristic evaluation is a usability evaluation technique developed by Jakob Nielsen and his colleagues in which experts, guided by a set of usability principles known as heuristics, evaluate whether user-interface elements conform to the heuristics or not (Preece et al., 2002).

These heuristics closely resemble the high-level design principles discussed in 3.3.1, but when used in evaluation they are called heuristics. The original set of heuristics used in this kind of evaluations
was derived empirically from an analysis of more than 200 usability problems (Preece et. al., 2002)

Nielsen’s original set of heuristics;

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

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

- **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. The system should support undo and redo.

- **Consistency and standards.** Users should not have to wonder whether different words, situations, or actions mean the same thing. The system should follow platform conventions.

- **Error prevention.** Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

- **Recognition rather than recall.** Make 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.

- **Flexibility and efficiency of use.** Accelerating-features, unseen by the inexperienced user, may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. The system should follow users to tailor frequent actions.

- **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.

- **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.

- **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 find, focused on the user's task, and not be too large.
These core heuristics are too general to be applicable to all systems and therefore there is a strong need for heuristics that are more closely tailored to the specific system in question. Different set of heuristics are needed when evaluating different devices, therefore evaluators must develop their own by tailoring existing heuristics to fit the system. Exactly how many heuristics that are needed and which are the best are arguable and depend on the system. Even heuristics often do not go far enough to help design a system. When several people are working on the same system each one might implement the same heuristic in a different way. To really achieve consistency and meet the heuristics, local rules should be made which tells all the people who work on the system how to handle specific cases. A major difference between heuristics and local rules is that heuristics can conflict with each other while local rules are absolutes for the context in which they are applied, that is the specific system evaluated (Dumas and Redish, 1999).

In heuristic evaluation the experts work with a specific set of heuristics and use the system as if they were typical users noting any problems they encounter. Nielsen (1993) suggests that heuristic evaluation should be conducted by 3-5 experts to be able to identify as many usability problems as possible. The evaluation period is recommended to be 1-2 hours in which each expert independently inspects the system. The expert needs to take at least two passages through the interface, the first to give a feel for the flow of the interaction and the second to allow the evaluator to focus on specific interface elements. If the evaluation is for a functional system, the evaluators need to have some specific user tasks in mind so that the exploration is focused. After the evaluation period the experts come together to discuss their findings and to prioritize the problems they have found and suggest solutions.

### 3.5.2 Usability Testing

Usability testing consists of an analysis of user performance in relation to the proposed system. The test may be conducted by letting users perform tasks using the system, observation, questionnaires, experiments and interviews. Usability testing involves working with users and gathering data that will have to be analyzed (Faulkner, 2000). The focal point in a study of usability should be on what is actually
used, not what is believed will be needed (Hackos and Redish 1998). Therefore to be able to conduct a good usability test the center of attention has to be on how people use the system, rather than how people think they should use it, or what computers can do (Suchman, 1987).

Five basic activities of a usability test described by Faulkner (2000) are;

- Identifying the target group
- Recruiting participants
- Establishing the task
- Carrying out the test
- Reporting the findings

**Identifying the Target Group**
The more closely participants represent actual users, the more useful will the test be (Dumas and Redish, 1999). Company employees should not be used in a usability test. Even if the employees are new to the particular system, they may know too much about similar systems. The usability test will not tell the right things if the employee has an easier time with the system than the external user do. Testing the wrong users can lead to two types of problems. If the participants in the usability test do not have as much experience as the real end-users, more problems than there is actually need to deal with will be found. If the participants used in the test are more experienced than the actual users will be, the real usability problems will not be found.

**Recruiting Participants**
It is not simply to recruit the participants to carry out the task; it will also be necessary to ensure that they are the appropriate user-type with the necessary range of skills. This can be made by letting the possible participants fill in a questionnaire to identify which people to use in the test (Faulkner, 2000).

**Establishing the Task**
Most testing methods will very likely involve some kind of task. These tasks can be for example questionnaires, scenarios and interviewing users. How the task is formulated will depend on what is wanted to establish from the testing. The task will need to be established and then
tested before it is used in a proper test situation. A task that appears to be clear to the producer of the test may not be clear to the user (Faulkner, 2000).

Carrying out the Test
It is important that the users know exactly what will happen and what is expected of them. It must be ensured that the users know that it is the system that is being tested and not themselves. The users should be told what will happen with any recordings, questionnaires and other testing material. They should also be told that they can quit the test at any point if they want to. Once the users are comfortable and know what to expect, the task can be presented and the test can commence. While the users are performing the task they might be observed. When the users have finished the test there should be a debriefing where the users can ask questions (Faulkner, 2000).

Reporting on the Findings
At this stage any findings, including problems, encountered during the test should be listed so that they can be analyzed and examined for possible causes and solutions. Some problems may be ignored because of other contributing factors but each problem should be considered and possible causes and solutions examined (Faulkner, 2000).

Observation
It is useful and interesting to watch users perform a task. However, it is important not to disturb the way in which someone typically works on a task. It is important to remember that people being watched may act in a different way because of the presence of the observer. This is not always conscious and it might not be the users’ intention to deceive the observer. It is called the Hawthorne effect when the results are affected by unrelated factors. It is important to remember that observing users may inevitably alter the way in which they work (Faulkner, 2000).

To avoid that the user feel awkward being observed by a crew of evaluators it is important to balance the numbers of test observers and test participants during a test session. That is; if one user is performing the test, there should be only one or two observers to avoid that the user feels inferior (Breakwell et al. 2000).
Questionnaire
The questionnaire is probably the single most common research tool. The main advantages of the questionnaire are its simplicity, flexibility and low cost as a method of gathering data. There are many different kinds of response formats. They can vary from multiple response items where the user is asked to circle one or more option, rating scales where the answer is marked by a cross on a scale, to ranking formats which asks the user to rank the alternatives from, for example, 1 – 6 (Breakwell et al., 2000).

Scenarios
Scenarios describe the task in a way that takes some of the artificiality out of the test (Dumas and Redish, 1999). Scenarios can be used as a method of gathering information about how users will use the system to deal with their task. A scenario can also reveal problem areas and errors (Faulkner, 2000). A good scenario is formulated in the user’s words, it is short, and it is unambiguous so all users will understand it (Dumas and Redish, 1999).

Interview
Interviews can be thought of as a “conversation with a purpose” (Kahn and Canell, 1957). There are four main types of interviews: open-ended or unstructured, structured, semi-structured, and group interviews (Preece et al., 2002). After having specified the research questions there is a need to formulate them into a form that the users will understand. It is important to avoid leading questions because it places a pressure on the interviewees to give the “right” answer. Even if it is not the real motivation for the research, the user has to be given some notion of why the questions are being asked and must feel that the sequence of questions makes sense (Breakwell et al. 2000).

3.5.3 Combining Evaluation and Testing
Usability evaluation is a good way to identify eventual bottlenecks in the system which then can be quantified with usability testing. It is probably a more sensible approach to combine these two techniques than to do them separately, because each method yields different findings (Faulkner, 2000).
3.6 Related Research

A very few studies concerning usability on telephony user interfaces with touch-tone input have been conducted. Since the shift has gone from touch-tone to speech recognition lately, most studies today are investigating systems with speech input. Below are studies that relate to usability on touch-tone input interfaces or handling of information in auditory interfaces described.

3.6.1 Auditory versus Graphical Interfaces

Voice is very difficult to utilize effectively. Listening to speech is much slower than reading for most users. When reading they can more easily skim and let their eyes wander while speech must be accessed serially. Speech is transient and requires the users' full attention; while a screen full of text remains available until the users are given a chance to look at it. Audio output is lost unless it is attended to at the moment it is spoken (Marx and Schmandt, 1996). Despite these disadvantages with speech, auditory interfaces are important because of their current practicality over the telephone.

A limitation with a telephone based voice mail interface compared to a visual interface is its linear access style of messages. (Schmandt, 1993) Retrieving messages over the phone is more cumbersome than with a graphical user interface. With a visual interface, the user can immediately see what messages are available and access the desired one directly. In a nonvisual environment, however, a system must list the messages serially, and since speech is slow, care must be taken not to overburden the user with a long list of choices. The slow, serial and transient nature of speech makes finding important messages not only frustrating but also time-consuming (Marx and Schmandt, 1996).

A recent empirical study of message retrieval in auditory versus graphical user interfaces (Wolf et al., 1995) revealed that a major problem with the otherwise well-received auditory interface was finding important messages. The interface in this study had both spoken input and output (Marx and Schmandt, 1996). Whereas the eyes can scan a list of several dozen messages in a matter of seconds, the ear may take several minutes to listen to the same list read out loud
THEORETICAL FRAMEWORK

(Schmandt, 1993). Further, the users must rely on short-term memory to recall the spoken items whereas the screen serves as persistent reminder of the options.

A conclusion that can be drawn from this study is that a first important step towards effective message management in an auditory interface is to prioritize and categorize messages and through this help the user to get a good overview of the system (Marx and Schmandt, 1996).

3.6.2 The Phoneshell Study

Schmandt (1993) describes a telephone-based auditory interface, Phoneshell, and also discusses the user experience of this system. Phoneshell uses touch-tone input and speech as output. It provides interactive access to voicemail, electronic mail, and a personal calendar.

The user experience of Phoneshell was investigated through an informal user study with users of the system (Schmandt 1993). Many of the participants had helped develop the system in question and this can jeopardize the validity of the study. The dominant issue in Phoneshell usage is its ability to deliver information concisely and also the minimal demand the system put on the users memory. This includes that the system speaks rapidly, concisely and always allow the user to interrupt its prompts. In Phoneshell the waiting time between prompts can be modified, as well as the times the list should be presented before the system shuts down. During output, the user may skip ahead to the next sentence, repeat the current sentence, or quit.

The speed of Phoneshell interaction was assigned to certain factors. These factors were that the users can increase the rate of both synthesized and digitized speech (Orr et al., 1965). This is a good feature because after some exposure the users find the normal speech rate unseemly slow (Beasley, 1976). Another factor contributing is that all prompts are interruptible, making it possible to skip menu prompts or to abort playback of an uninteresting message. Combined these factors take care of at least some of the frustrations users experience dealing with touch-tone systems (Schmandt 1993). Certainly a key to the good response of Phoneshell is its ability to integrate many functions into a single interface (Schmandt 1993).
3.6.3 Speech Recognition

Marx and Schmandt (1996) describe sequential navigation in a telephone-based messaging service, MailCall, using speech recognition and synthesis. The use of speech recognition, however, raises the users’ expectations of the interaction since it implicitly resembles a human conversation. These heightened expectations can be damning since speech recognizers are far less adept than humans. Knowing what to say in a speech recognition system is a stumbling block for beginners. Yet taking the time to listen to the options can be tiresome for experienced users. Further, speech recognition errors slow down the interaction because it requires the system to perform constant “grounding” between the system and user to ensure that they share a common perception of what is transpiring (Brennan et. al., 1995).

Studies show that many people do not ascribe much competence to a spoken language application but instead expect to be led step-by-step as with an Interactive Voice Response (IVR) system. Marx and Schmandt (1996) showed that for sequential navigation, speech was in fact a disadvantage. The time necessary to say “next” and then wait for the recognizer to respond can be greater than just pushing a touch-tone, especially when the recognizer may misunderstand.

3.6.4 Skip and Scan

Resnick and Virzi (1992) present a new telephone interface style in which callers use explicit commands to accomplish the same skipping and scanning activities as in traditional numbered menus, that is when the users listen to their options and figure out which one to choose. The style presented is called “Skip and Scan” and is argued to give the users more control over the process of listening and recording. The users are only presented to one option at the time, and can choose between selecting the option, listen to the next or previous option. This makes the users more certain of what options that are available, and reduces the users’ memory load. The implicit structure of recorded prompts and information is made explicit and available to users for navigation purposes. Initial evidence indicates that the new style is preferred by users and lets them access information significantly faster than in traditional numbered menus.
THEORETICAL FRAMEWORK

Result from experiments, comparing traditional numbered menus with skip and scan menus, showed that the participants made selections using skip and scan menus more quickly than in traditional, numbered menus, and preferred the skip and scan menus in subjective ratings. The skip and scan style of selecting an option from a menu gives users more control over what prompts they hear.

Skip and scan is a promising telephone user interface style. Through explicit navigation commands, it gives users some of the control they get from shifting their gaze in visual interfaces.
4 Method

This chapter explains the method used in this study. To be able to investigate the usability of MVAS two research methods was combined. These were usability evaluation through heuristics, and usability testing with real end-users. The chapter reviews how the evaluation was conducted and how the test was designed and executed.

4.1 Usability Evaluation

The recommended number of evaluators used in a usability evaluation is 3-5 (Nielsen, 1993). Despite this fact, the lack of resources led to that this evaluation was made with only two evaluators.

The system was first gone through by the evaluators to give an initial understanding of it. Then a set of seven heuristics were formulated on the basis of the initial understanding. The heuristics used were;

- **Shortcuts**: shortcuts should be provided to help users navigate in the system in a fast and efficient way.
- **Feedback**: appropriate and sufficient feedback should be provided to the user when needed.
- **Consistency**: the information should appear in a consequent and logic order. The terminology and structure should also be consistent through the entire system
- **Terminology**: the terminology used should be familiar to the users.
- **Error**: error messages that help users recognize, diagnose and recover from errors should be given when appropriate.
- **Visibility**: the users should be aware of what is happening inside the system. The visibility should be based on recognition rather than recall.
METHOD

- Minimize users’ mental workload; the system should be designed in a way to try to minimize the mental workload and the memory demand put on the user.

These heuristics were chosen from different set of heuristics described by Nielsen (1993), Faulkner (2000) and Norman (2001) because of their applicability on an auditory touch-tone system. With this set of heuristics in mind the system was then gone through iteratively and usability problems and issues found were noted and related to these heuristics. The system was gone through with certain tasks in mind so that the experts would act like real users. The tasks varied from simple to complex that is; from sending a voice message to change the pin or to add a contact list. The two evaluators worked independently until the end of the iterations when they came together and discussed their findings.

To make up for the fact that there were only two evaluators, the evaluation was combined with usability testing. The results from the testing validate the evaluation, which makes the need for more evaluators not as evident as it would have been if only an evaluation had been conducted.

4.2 Usability Testing

The usability testing had both a qualitative and a quantitative part. The qualitative part was the main focus of the study while the quantitative part was conducted to validate the qualitative findings as well as investigating differences between the participants which were not possible to find through the qualitative data. For example if there were any differences in how the participants performed depending on how experienced they were in using touch-tone systems.

4.2.1 Design

A between-group design was chosen because a difference between using the system via a mobile phone comparing to a landline was wanted to be measured and analyzed. This design was chosen to cover the entire problem area when the system today is used via both mobile and landline phone.
The participants (P) were assigned to one of two main groups, the mobile-group and the landline-group. This was done by assigning every second test participant respectively to the mobile (M) and the landline-group (L). Since there were 40 participants in the test it lead to that it was 20 participants in each cell. These two groups were then divided into two further groups depending upon which parts of the testing procedure they were about to take part in. These groups were; participants doing only the interview (I) and participants doing both the walkthrough and interview (WI). According to Shneiderman (1998) eight participants are a valid number to do statistical testing on, and therefore the number of participants is enough for doing statistical measurements.

For example; the first participant was assigned into the MI group (mobile and a testing procedure with interview only) and the second participant was assigned into the LI group (landline and interview only). Participant number three and four were assigned into MWI and LWI, respectively. That is the mobile or landline-group with a testing procedure containing both an interview and walkthrough. See Figure 3.

**Figure 3: The design of the study**

**Participants**
20 men and 20 women in the ages between 17 and 63, with an average age of 30 participated in the test. The men ranged from 20 to 42 and had an average of 31 years old. The women ranged from 17 to 63 and had an average of 28 years old.

**Criteria for Participating**
The criteria for participating in the study were carefully chosen. These criteria were that the participants were aged over 15, understood
spoken Swedish, and had no impairment of hearing. Another criterion was that they had been in contact with at least one touch-tone system before but not with MVAS. The last criterion was that they used a mobile phone at least twice a month.

The criterion of an age limit of 15 years was due to the need of understanding a certain level of spoken Swedish and also due to the fact that the participants had to have been in contact with at least one touch-tone system before (see 4.2.3). These kinds of systems are often used to book tickets or to get support via telephone. A person under 15 years old might not require these kinds of services yet and therefore this age limit was set.

The criterion of no impairment of the participants hearing ability was due to the criticality that the test participants could hear the system properly. Of course this rule was not rigid in the sense that if a participant would have an impairment in hearing he or she was welcome to try if he or she thought that it was possible to participate in the test or not. If this situation would appear the test conductors would judge about the situation and decide whether or not to include the data from the participant in question in the results or not.

As mentioned above one criterion was that the participants in the study had been in contact with at least one touch-tone system before. The reason for this was that it would improve their understanding of the purpose of MVAS and would therefore make it possible for them to discuss the system in further detail than a person who never had used a touch-tone system before. It was however of crucial importance that the participants had not been in contact with MVAS before because this would change their ability to analyze the system without prejudice. The reason for selecting inexperienced participants was that inexperienced users would look at the system without prejudice and the only experienced users of MVAS that were available are the ones employed at the company, who could not be included for validity reasons.

The final criterion, that the participants should be regular users of mobile phones, was included because the participants had to feel
comfortable enough using a mobile phone otherwise this would have affected them in the test situation. 
If any of these criteria was not fulfilled the participant was eliminated from the study

**Recruiting Participants**

Since the company does not have any explicit intended segment for their products, only that their product should suit all phone-operators interested, the intended segment used when recruiting participants to this study was possible end-users for any operator in Sweden. Since the intended segment is so extensive, it is hard to judge if the intended segment was actually going to be reached in the recruitment or not. The criteria used for identifying the participants were therefore developed with this fact in mind.

The recruitment of the participants was done through the employees of the company, among their family members and friends. The employees were told what criteria that had to be fulfilled for a person to be able to participate in the test.

Since the participants were recruited with help from the employees at the company in question an issue was raised if the results would somehow be affected by this. A question about what the participants knew or had heard about MVAS was therefore asked before the test was conducted so that the evaluators could take this into consideration when analyzing the data. However no participants reported that they had any knowledge about how the system worked or had been told any opinions about it in advance.

**4.2.2 Material**

Below all the material used in the testing sessions is described.

**Instructions**

The instructions (Appendix A) were written down in beforehand as they were going to be presented to the participant. They contained information about the purpose of the test, how personal information would be treated and that the participants were allowed to quiet the test at any time. The instructions were given to make the participant aware of why and how the study was going to be conducted. Information
about why there were one observer and one monitor and their respective tasks was also given. The evaluators did not ask the participants to focus particularly on the system’s ease of use or dialogue. This was by purpose; when asking this would have made the participants think about the system in a different way than wanted, this would make them find problems just for the sake of it and that was not the purpose of the test.

**Background Questionnaire**
The background questionnaire (Appendix B) contained questions about the participants’ age, gender, and their background concerning use of mobile phones and touch-tone systems. Also a question about their view on new technology was asked. These questions were asked to certify that the participant criteria were fulfilled before the test was conducted. The purpose of the questions was also to be able to divide the participants in different groups depending on their background characteristics for the sake of the quantitative analysis.

**Walkthrough Heuristics**
The walkthrough heuristics (Appendix C) contained information about what topics to rise during the walkthrough and how the monitor should act if the walkthrough came to a complete standstill. The walkthrough was included in the study because it was thought to give an idea about what features the users really wanted in a voicemail application which was one of the problem statements of the study.

The walkthrough was conducted in the beginning of the test. This was to ensure that the participants described a system they wanted instead of the one in which they were going to perform tasks in this study. Also it was important that the walkthrough had been conducted before the interview took place to be able to, if necessary, go back to the walkthrough and make changes if the participant felt he or she had something more to add.

**System Settings and Scenarios**
The system test, in which the participant performed scenarios (Appendix D) through using the system, was the major part of the study. The system test was included in order to give the user an idea of
how the system works and hence to be able to talk about the system in the interview.

The system’s settings were adjusted to the test to be more similar to the systems used by the company’s customers and to fit the purpose of the study. The changes consisted of that the features to read e-mail and print faxes were excluded, and that auto-play of new messages was active.

The scenarios contained four separate tasks which the participants were asked to perform. These were; Scenario 1; to activate the account, Scenario 2; to answer on a new voice message, Scenario 3; to send a voice message, and Scenario 4; to change the personal greeting and search for a command to end the call. The scenarios in the study represented the main tasks the system is used for and also represented one task each under each of the three options in the “main menu” as well as the activation of the account.

To decide which tasks to include in the scenarios informal discussions were held with four users of voice mail applications. They concluded that voice mail is most frequently used for listening to or send voice messages. More seldom the system will be used for changing the pin or to record a new greeting. Therefore it seemed like the scenarios supported the tasks that are the most common ones.

The participants were given the scenarios to read by themselves. The reason for this was that the scenarios contained information that had to be used while the scenarios were performed, such as telephone numbers and the like. The participants were only given one scenario at a time as an attempt to try to minimize the stress and workload that the participants felt (Rubin, 1994) and also to clarify which scenario that they were going to perform at each time.

The recordings of voice messages along with other input that the participants had to make during the scenarios were formulated in beforehand so they would be the same for every participant and also feel as natural as possible (Shneiderman, 1998). However since many different participants participated in the test, these recordings can
impossibly feel natural for all of them and this might affect the results. All participants were also asked to use a particular name in the recordings to make the time consumed for this matter identical for all participants. Also, this fact that they used another name than their own can affect the outcome and make some participants feel unnatural in the situation.

Scenario 4; concerning changing the greeting and searching for a command to end the call, was afterwards divided into two different parts (Scenario 4.1 and 4.2). The result of the performance in this scenario as a whole would not say anything about how easy or difficult it was to change the greeting if it at the same time included the time for searching a command. If it is easy for the user to change greeting it does not mean that the system is easy to navigate. It was not possible to divide Scenario 4 into two different scenarios in advance, because it is hard to formulate a scenario where the participant is supposed to navigate in the system without doing a specific task other than that. Therefore the time and log of both tasks were logged separately and afterwards treated as two different scenarios.

The scenarios were ordered in a way which seemed to be the most natural way to perform the different tasks (Shneiderman, 1998). The first thing the participants did was to activate the account, thereafter they listened to their new message. Then they sent a voice message and finally they changed their greeting.

**In-depth Interview**

The in-depth interview (Appendix E) was included to get an opportunity to ask the participants more questions in detail about their opinions of the system and other cognitive aspects that were considered interesting. In an interview it is important that the questions are ordered to make sense to the participant (Dumas and Redish, 1999). This means that they should follow a logical order and that the format of the interview fit the situation.

The interview was semi-structured and contained questions about the overall impression of the system, how the navigation in the system was perceived, how the participants experienced the system’s dialogue, and
questions about how the participants imagined the system in their head, their mental model. Questions about if it was difficult or easy to use the system with a mobile phone or a land-line phone was also asked depending on which group the participant was in.

As a part of the interview the participants were given a ratings sheet (Appendix F) which asked them to estimate different issues regarding the system. This was because some quantitative data was wanted to be gathered about the participants’ subjective opinions. The ratings were made up by a 10 cm long Visual Analogue Scale. The Visual Analogue Scales (VAS) is used to measure a variety of subjective phenomena. A VAS is a straight line with anchors at the end which are labeled as the extreme boundaries of the sensation, feeling, or response to be measured. It enables the rater to make fine discriminations without the constraints of direct quantitative terms (Wewers and Lowe, 1990). For further details on exactly how these scales were used and formulated in this study, see Appendix F.

4.2.3 Pretest

Three pretests were conducted to test the test itself including all the materials used. Two men and one woman, aged 12-30 were tested. All the participants in the pretest got the instructions read out to them and completed the background questionnaire. One of the participants did the walkthrough; all of them did the scenarios and a brief or in-depth interview. The pretests lead to changes in almost all the parts of the study. These changes were that some information about the purpose of the study was added in the instructions. Also some clarifications of the tasks in the scenarios and questions in the background questionnaire were made. The instructions were changed because the test monitor perceived it hard to read them out loud naturally and at the same time keep eye contact with the participant.

One of the pretests made it obvious that a system in English would be hard for some participants to use. The language of the system was therefore changed from English to Swedish. The same pretest lead to that a lower age limit of 15 years was set. This age limit was made due to that one of the pretest participants did not understand how to use the
METHOD

system. The participants had not been in contact with a touch-tone system before.

It was found during the pretests that the participants felt freer to speak if the interviewer did not write down the answers during the interview. Therefore it was decided that there should be one observer taking all the notes and one monitor that could concentrate on listening to what the participant had to say. This was indicated by situations when the participant stopped talking as soon as the monitor started to write and waited for the monitor to finish before continuing.

4.2.4 Procedure

The test took place in a small room with a desk, a whiteboard, and three chairs. One test monitor and one observer served as test leaders (Rubin 1994). The test monitor were in charge of the test and also responsible for all interaction with the participant. The observer observed the participant’s actions and also took notes of what happened. The test monitor was seated near the participant and the observer a bit further behind the test monitor. See Figure 4. The test monitor read the instructions out loud to the participant and then the participant was asked to fill out the questionnaire.

![Participant, Monitor, Observer](image)

Figure 4: Test room settings.

If the participant was to perform the walkthrough procedure the test monitor gave the instructions for the walkthrough, otherwise the system test with the scenarios was started straight away.

If the participant in the walkthrough procedure did not have any questions, he or she was given a starting point: “You receive a notification via a text message, SMS, saying that you have two new messages in your inbox, what do you want to do?” The different steps the participant wanted to happen were written down on the
whiteboard. The walkthrough were ended when the test monitor considered all the necessary subjects covered.

When the walkthrough was completed the test monitor asked the participant a question about if he or she had heard anything about the system in advance. This question was asked all participants, no matter which procedure they belonged to. After this the monitor gave the participant the instructions for the system test. The participant was first given scenario 1. The participant was asked to read it carefully and start whenever he or she wanted to. After the participants had performed all the scenarios, the participant was asked to estimate the time they thought to be reasonable on each task. This was made to validate the time limits used in the successful completion criteria (see section 4.3) but also to see if the participants’ performance was better or worse than they thought was reasonable. Lastly the interview took place.

Control Procedures
In an attempt to make the test session as similar as possible for everyone participating in the test there were some control procedures. Since 40 test sessions was going to be conducted it was decided that the two evaluators were going to take turns in being test monitor and observer. When the test session contained a walkthrough, the sessions were held by the same person every time and it was also done following certain rules, see Appendix C. The testing sessions that did not contain the walkthrough were held by the other evaluator. The instructions were written down so that the evaluators, no matter who were acting test monitor could inform the participant in a similar manner. The interview was also held on the basis of a template so that the questions would be similar.

4.2.5 Apparatus

Telephones
The participants used a mobile phone, Ericsson R320s, or a land-line phone, Telia Response. Key presses and time used for the scenario was logged using the display and features of the mobile. With the landline phone the log was written down by hand by the test monitor because the telephone had no display. It was considered that if the monitor in
the instructions told the participant that the activities would be logged, the participant would not be disturbed by that.

The landline phone supported speed-dial, just as the mobile did, which was necessary so that the procedure using landline/mobile phone was as similar as possible. Since the landline phone did not record the time of the call, the evaluators used a mobile phone to clock the call. Since different ways of taking the time of the landline call and mobile call were used there might be a difference in time between these two groups. The difference was measured to a maximum of one second and therefore not seen to be large enough to have an impact on the result. Further the time-difference would be the same for everyone in the landline and mobile phone group respectively and if needed it was possible to compensate for it.

**Recording Equipment**
To record the participants’ opinions during the walkthrough and interview a minidisc with a microphone was used. It was decided to not film the participants on videotape. Often a video camera makes participants more nervous and therefore they do not behave natural in the test situation, and having the interview on videotape would not have added anything of importance.

**Walkthrough Equipment**
The walkthrough was written down on a whiteboard with a whiteboard pen. The whiteboard gave a necessary overview and served as a good tool to get confirmation from the participant that the monitor had understood everything correctly.

To document the walkthrough without having to copy it by hand, a photograph of the whiteboard was taken using a digital camera. This made it possible to reproduce the walkthrough in matter of seconds, which saved time for the evaluators and also made sure that nothing of the walkthrough was lost in the transition.

**4.2.6 Classifying the Material**
When all the test sessions were completed the timing and the logs from each participant on each scenario was put together. Each participant’s scenarios were then classified according to the successful completion
METHOD

criteria (Appendix G). These criteria were used to get a good measurement of the participants’ performance, including both the time and the error rate. Since errors are a sign of bad usability, both errors and time was wanted to be taken into account. If only time is included in the performance, the participants can make a lot of errors and still have a good performance.

The successful completion criteria were formulated on basis of informal inquiries held with thirteen users of voice mail systems, where the reasonable time doing each one of the activities included in the four scenarios was asked for. Also the researchers took time when they performed the scenarios in the system in question. The performance of the scenarios took longer time, even for experienced users, than the thirteen users estimated. Therefore the times used in the criteria were increased to correlate with the system used in the test.

Thirteen users might be enough to give a true picture about what people in general think is a reasonable time to spend on a certain task in a voicemail system. But since these time estimations were going to be manipulated by the researchers to take inexperienced users into consideration, thirteen opinions were seen as enough to give a rough picture. The thirteen users’ estimated times can also be questioned since it in general is hard to estimate time correctly.

One big advantage of asking the participants in the test to estimate reasonable time was that they had used the system and therefore could have that in mind when estimating the times. This might give their estimations a better reliability than the thirteen users asked in advance.

Except from the obvious classification of the participants, that is if they belonged to the landline or mobile phone group and which testing procedure they did, the participants were also classified in to different groups depending on their age, gender, how experienced they were in using touch-tone systems and if they used voicemail or not in their daily life.
4.2.7 Method for Analysis
The in-dept interview and walkthrough recordings were listened to and written down word by word, disregarding pauses, slip of the tongue etcetera. This can have led to that some comments and issues in the material have not been analyzed in the same detail as others. Some important information can therefore have been missed, or disregarded. Also, since the researchers have specific pre-knowledge and ideas of what to find which is really hard to disregard, the analysis of the material can have been affected.

The qualitative analysis of both the walkthrough and interview material were done through a hermeneutical approach. The characteristic of the hermeneutical approach is the interchange between the entirety and the parts of the material (Ödman in Gustavsson, 2004). The interpretation is seen as a never ending process which both goes upwards (the entirety) and downwards (the parts) in the hermeneutical circle. The entirety and the parts are affecting each other continuously. In a hermeneutical approach the researchers can use their pre-knowledge actively, as long as the researchers give a true picture about how their own prejudice might have an affect on the analysis.

In this study the hermeneutical approach means that the material was gone through iteratively, with a focus on finding common attitudes, opinions and impressions. These were summarized before the material was gone through again proving or rejecting the issues found earlier. The second iteration also lead to that the summary became more extensive. The iteration was stopped when the main issues was thought to be found and when the last iteration had not give anything new of interest. After five iterations the findings that were considered as true as well as the ones rejected but still interesting to discuss were put together.

Hypotheses about the quantitative part of the material were also formulated. The data were entered into a statistical analysis program and the hypotheses were tested with statistical tests (ANOVA).
5 Analysis

This chapter presents the analysis from both the evaluation, and the testing point of view. The analysis of the usability testing consists both of a qualitative and a quantitative part. The chapter ends with a table and a list where the most important findings are summarized.

5.1 Evaluation Analysis

The set of heuristics used in the evaluation is described below (see 4.1). Thereafter the analysis of the system based on this particular set is presented.

- **Shortcuts**: shortcuts should be provided to help users navigate in the system in a fast and efficient way.
- **Feedback**: appropriate and sufficient feedback should be given to the user when needed.
- **Consistency**: the information should appear in a consequent and logical order. The terminology and structure should be consistent through the entire system.
- **Terminology**: the terminology used should be familiar to the users.
- **Error**: error messages to help users recognize, diagnose and recover from errors should be given when appropriate.
- **Visibility**: the users should be aware of what is happening in the system. The visibility should be based on recognition rather than recall.
- **Minimize users’ mental workload**: the system should be designed in a way to try and minimize the mental workload and the memory demand put on the user.
5.1.1 Identified Usability Problems

Short-Cuts
One problem found in the system is the location and formulation of the global keys. Global keys are the term used to refer to the generic commands used in the system. This means that the same command can be used in all different kinds of menus and still have the same function (Rosenberg and Morgan, 1984). The global keys in this system are not presented until a menu is re-prompted, presented two times in a row. After the re-prompt three seconds pass before the global keys are presented. This location of the global keys leaves a lot to wish for. The bad location of the global keys can be associated with bad visibility of the system (Gilmore, 1991). When the users do not know what features the system offers, how can they possibly use them?

A good way to improve the system regarding this matter is to have the global keys in the beginning of the system, in an introduction to the system or presented in the main menu. Then the users would be familiar to the global keys. Putting the global keys only as an introduction would increase the mental workload of the users (Schmandt, 1993) and therefore this is not a strategy sufficient on its own. It is good to present the global keys at re-prompts so the users get reminded if they have forgotten them. All re-prompts throughout the entire system should include at least some of the global keys that are seen as important in that situation. This would enhance the visibility of the system (Gilmore, 1991).

When the users hear the global keys they are often lost or frustrated. The system presents them as following; “I suppose you know that you can press #1 to get back to the main menu...”. This formulation is unfortunate because it assumes that the users have heard the global keys already and also that they remember them (Shneiderman, 1998). Since it impossible for the users to have heard the commands before this prompt, it is wrong to assume such a thing. Further the formulation decreases the visibility of the system since it makes the users believe that they have missed some information that is not present.
A better way to inform the users of the global keys is with a simple statement like; “To return to the main menu, press *1”. This would add visibility to the system and also prevent the users from feeling intimidated (Shneiderman, 1998).

**Feedback**
Some problems identified in the system can be referred to the heuristic on feedback. The system seldom provides feedback which is a shortage. The user should be given feedback when data has been entered as well as when the system status has been changed (Norman, 2001). As it is now, the system gives no feedback on recordings and the feedback that is given when the system status has been changed is insufficient. For example; when the users record a personal greeting, the greeting is activated automatically but the users are not informed about it. Instead the system gives the users an option “to stop using it (the greeting)” and this might confuse the users even more. They do not know which greeting will be de-activated, is it the new one they recorded or the old one? This problem not only refers to bad feedback but it also increases the users’ mental workload (Schmandt, 1993) and makes the system hard to understand; hence reducing the overall visibility of the system.

Ways to improve the system’s feedback would be to give the user feedback on recordings, saying; “greeting recorded”. This would enhance the visibility of the system. If the system also says “this greeting is now active” the system would be even more visible. Giving the user feedback in this way also decreases the mental demand put on the users (Schmandt, 1993).

Another bottleneck in the system due to insufficient feedback is the navigation. When talking about the navigation the directions used in this thesis are upwards and downwards; as in a hierarchy. Upwards means going to a previous menu and downwards means making a selection. The navigation upwards in the menus is performed through the global keys. But the system provides the user with information such as “if you have finished working with this greeting, press *2” which leads to that the users do not understand that *2 is a global key that navigates them upwards to the previous menu. This makes the users confused and it also gives the users a feeling of being trapped in the
menus. They do not realize that there is a way out. This insufficient feedback gives the system poor visibility (Preece, 2002).

A way to improve the system in regard to this matter is if the system instead would say “if you want to return to the previous menu, press *2”. This would explicitly tell the users that they are navigating upwards which would enhance the visibility of the system (Nielsen, 1993).

If the system status has been changed, the system should describe the exact changes to the user, not anything else (Nielsen, 1993). Another problem identified that also can be referred to feedback occurs when the user has listened to the new messages. He or she is then told how many saved messages there are in the inbox. The problem is that the system tells the user the number of messages that were in the inbox before the new messages were saved, and therefore this number is incorrect. This might confuse the users if they are told that “You have no saved messages” but they know they just listened to some messages that were automatically saved.

If this problem is impossible to solve in regards to technical issues, the prompt should not be included. It is better that the system does not tell anything, than tells something that is not true (Norman, 2001).

**Consistency**

Several bottlenecks identified can be referred to the heuristic on consistency. One of them is regarding the presentation of the prompts. The prompts are in general presented in action-key order, that is; the action is described before the key is presented. This is however not consistent in the entire system.

Other inconsistencies occur when the menu options are presented. In some menus, listen is 1 and use is 2, but in other use is 1 and listen is 2 which can confuse the users. Even the global keys are sometimes inconsequently used and numbered. The key to return to the main menu is one example. This option is *1, but in some menus the option to go back to the main menu with key 1 is given. It is confusing to use different commands for the same action.
The ordering in which the options are presented is also inconsistent. In general the menus are presented in numerical order; that is 1, 2, and 3. If, for example, 3 is tied to a certain option, the key 3 is skipped in menus where this option do not occur and that makes the order, for example 1, 2, and 4. Despite this, there are some inconsistencies in the order the options are presented which are confusing to the users. For example in one menu the options are read out in order; 1, 3, and 2.

Further the system is inconsistent when it comes to entering data and recording. When the users are asked to enter their pin the first time, the system asks them to finish by pressing #. However when the system asks the user to re-enter the pin; it does not ask them to finish by pressing #.

There are also inconsistencies in the terminology used. In some menus the system says “If you are satisfied, press 1” and in other “to use this X, press 1”. Both these options serve the same purpose, and therefore these options should be phrased in the same way.

All these inconsistencies make it harder for the users to learn the system (Teitelbaum and Granda, 1983). If the options are inconsequent they can not guess what button to press but instead have to listen to all the prompts before they can select an option. Consequently this leads to an increased mental workload for the users which should be avoided.

**Terminology**

There are also some problems regarding the terminology in the system. The system tested was in Swedish. Therefore all terms and language discussed below are translations made from the Swedish version into English.

Some terms used in the system, can not possibly be understood by a user without any previous knowledge. These are the “standard scheme” referring to the time format and the “personal assistant” referring to the forwarding of calls. Difficult or unfamiliar terminology decreases the visibility of the system, making it harder to understand what the option is about and hence increasing the users’ mental workload (Shneiderman, 1998).
ANALYSIS

Ways of improving the system when it comes to terminology is to use easy and familiar language that makes the user understand the functions (Shneiderman, 1998). In this case, the phrasing “personal assistant” could be replaced with “number”. The prompt would then be; “To choose a number that you want your unanswered calls to be forwarded to, press 2” This will give the user the necessary information without confusing them with a “personal assistant”. A good way to improve the understandability regarding the term “standard scheme”, would be to explain what a standard scheme is. Instead of the prompt “To just use the standard scheme, press *3” the prompt could be “To use the standard scheme, with workdays Monday to Friday from 8.00 to 16.00, press *3”. This change also excludes phrasings as “just” that are unnecessary and makes the users’ feel depreciated and should therefore be avoided. This change also makes the prompts longer but that is necessary to make the users understand their options.

Another reason for changing the terminology is to make it more explicit to the users what action they should take. One situation when this is needed is when the user is activating an account. The system then asks the user to record his or her name. For some people it might not be obvious that it means to actually say their name, so the system can record it. Therefore the word say or speak is a better option to use than record (especially in Swedish). Record can mean a lot of different things such as pressing keys as a way to enter the name. This can also be connected to the visibility of the system. Good phrasing and terminology makes the system more visible (Shneiderman, 1998).

Error

One usability problem identified in the system is the absence and lack of error messages. Since the feedback in the system is also lacking it is hard for the user to know when they have made a mistake. If the user presses a key that is not an option in one menu the system says; “# not a valid key in this menu”. However; when a user is asked to enter the pin and finish by pressing #, if the user makes a mistake and presses * instead of #, the system does not provide an error message or any feedback that the wrong key has been pressed. This means the users do not realize that an error has occurred until they have entered the pin twice and the system gives feedback, such as; “Sorry, lets start from the
even then, the users do not get any feedback to help them understand what went wrong. A good way to solve this problem would be to have “* is not a valid key” then the users would realize their errors and also be able to recover from it as soon as possible (Faulkner, 2000).

Making the opposite mistake, that is confusing * with #, when trying to use the global keys can lead to disturbing consequences. If the user presses # and then 2; the information under option 2 is played immediately, canceling out the error message “# is not a valid option in this menu”. This leads to that when the users tries to, for example, go back to the previous menu and presses #2 instead of *2 the users get into option number 2 in that menu. This is really confusing and since the users often do not know the menus by heart it can take quite a long time before they realize that they are in the wrong menu. Even when they realize this; they might not understand how they got there, and the next time the user tries to press *2 the same error will occur (Faulkner, 2000).

A system should help the users avoid making mistakes, instead of assuming that errors will occur (Preece et al. 2002). Therefore the formulation of the prompt “If you make a mistake, press *3” is unfortunate and should be avoided. Instead, the system should provide better feedback and error messages that gives the users a better explanation about what have happened and how to go on. Good error messages go hand in hand with sufficient feedback and both of these lead to good visibility of the system

**Visibility**

Almost all usability problems described earlier in this chapter can be referred to the heuristic on visibility. However, there are even more identified problems due to the loss of visibility in the system. One example is when a user wants to create a new contact list, the information about how to do it is poorly formulated. The system asks the user to “Enter the number of the contact list you want to work with. You can either create a new list or work with an already existing list”. It is not clear enough how to create a new list. The user is supposed to enter a number of a list that does not yet exist to be able to create a new one, but the information from the system is not sufficient enough to
make the user understand this. This makes it difficult for the user to be able to predict how the system works and how to act, and hence the visibility of the system is decreased (Gilmore, 1991).

If the system explained how to create a new contact list better, for example through an added sentence which says “To create a new contact list, enter the number you want to give the list, and a new list will be created if no existing list is associated with this number”. This would make it easier for the user to know exactly what to do and also to make him or her feel more secure about what is happening in the system (Löwgren and Stolterman, 1998).

Another example of low visibility is the formulation of the greetings in the system. The system gives the user the opportunity to use different greetings for different occasions. The greetings have names such as; “greeting for all calls”, “extended absence greeting”, “greeting for no answer” and “when your phone is busy”. This feature is good but it is hard to understand what the different names of the greetings means because of the terminology used (Shneiderman, 1998). It is also difficult to understand which of these greetings that can be active at the same time.

A way to try and solve this problem would be to cluster the greetings which can be used at the same time together in a new menu. This would add visibility in the sense that the user would know which greeting belong together. This should be complemented with naming the greetings so that they better explain what they stand for (Shneiderman, 1998).

An additional problem creating bad visibility in the system is when the user is supposed to enter a number or make a recording and end the entry with #. The system presents this by saying “Enter your pin, finish with the #-key. If you make a mistake, press *3 and try again. Your pin has to be 4 digits”. This prompt is too long to just tell the user to enter the pin. The talking after “finish with the #-key” should be removed and placed in an error-message. This would increase the visibility and make the system easier to follow and understand (Gilmore, 1991).
Minimize the Users’ Mental Workload

The system is not governed by a sufficient distinctive set of rules and therefore the user has to learn how to act in every dialogue separately which creates problems (Nielsen, 1993). As a conclusion it can be stated that all problems and bottlenecks identified in the system have a negative effect on the mental workload of the user. Therefore all the problems identified above must be taken care of to be able to make the user understand how to use the system and feel comfortable while using it.

5.2 Testing Analysis

Below the analysis made of the usability testing is presented. The analysis of the testing has both a qualitative and quantitative nature. The qualitative analysis is based on the participants’ opinions and comments while the quantitative analysis is based on statistical calculations made on the gathered data during the test sessions. These analyses are seen as complementary to give a complete picture of the situation.

5.2.1 Qualitative Analysis

The findings in the qualitative analysis, including both the walkthrough and the interviews, are divided into two parts, one about the functionality of the system and one section about the organization and structure of the system, which includes more of the participants’ subjective experiences.

Functionality

The participants stated that the functionality of the system was good. This can be shown through this quote; “Good features, it gives me options to choose. I don’t have to use them but I can use them. It was a good thing to record your name, it’s like a header. You know directly who it is and then you can skip forward if you want to.” P4.

The feature that makes it possible for a user to send a voicemail from within the system was appreciated. Then the users do not have to call the one that has left a message to be able to respond, instead they can reply straight away from within their own voicemail system; "I thought that the idea was really good. That you could call in and leave messages to each other..."
Another feature that the participants found surprising but appreciated was the fact that they hear the name of the person the message is from when they receive a message and to who’s number it is associated with when they send one. The participants saw the advantage that this might help them to make sure that the message is sent to the right number. One participant expressed it like this; “I got a bit surprised when I was about to send a message and pressed some keys and then it said “Susanne”. Then it might be more difficult to send the message to the wrong number.” P8.

Concerning the vocal functionality of the system, the participants thought that the voice was more human and personal than touch-tone system voices are in general. The participants also experienced the voice to have a reasonable pace. This is one opinion from a participant; “Pretty nice, more personal than you’re used to. The voice felt more human than usual, like a real person” P23.

The participants asked for a way to prioritize both the message they send and the messages they receive depending on who the message is from. This would help them decide which messages to attend to first. They also wanted to have the option to play the messages in their inbox in their own preferred order. Some participants wanted their messages to be played in first-in-first-out order and some participants wanted them in last-in-first-out order. One participant expressed it as follows; “I want to hear the most important message first, but that is really hard to judge about. The reason for this is that; it is the person that sent me the message that gets the option to do it (prioritizing the message). Or maybe I have prioritized them in some way my self, like I want messages from this sender to be read out first. I think that is the optimal solution. Otherwise I want the messages that have been lying in my inbox the longest time to be read out first. (last-in-first-out order)” P23.

Another function that the participants asked for was that they wanted to be able to change their greeting depending on both what they are doing at the moment and who is calling them. This can be shown through this quote; “On the other hand, you can have alternative greetings either if you are, like in my case if you are in a lesson then you can turn on that you are in a lesson. I should be able to record them beforehand and then alter between them.” P4.
Organization and Structure
One major finding was that the participants blamed themselves and not the system when they were unable to perform a task. This can be due to people’s general attitude to technology. The users take for granted that the product or system is working properly and in the best possible way and therefore when something goes wrong it is their own fault and not the system’s. This quote; “First it was when I was going to reply, one could reply directly when you heard the message, right? But I didn’t, I tried to answer from some other place, then I listen to the message again” P29 clearly states that the participant thought that the system worked in one way, the “proper way”, but when it did not do this, the user blamed himself or herself for the mistake.

As the quote above expressed, the participants thought it was too difficult to reply to a new message and that it took too long time. They believed that it was possible to answer when they had heard the message for the first time, and got surprised when they had to listen to the message again to be able to answer it. One participant expressed it like this; “Is it impossible to answer directly? That’s a loss in the system. It’s all about communication; you shouldn’t have to pick something up from the drawer to do that” P10.

One thing playing a part in that the participants blamed themselves for the errors that occurred is that the system contained too much information. The system was “talking too much” P12. If there is a lot of information the user can have difficulties deciding what to attend to and what not. Further the information given is not sufficient or clear enough to give the user the necessary picture for rejecting or choosing one option. Both these factors contribute to increasing the users’ mental workload (Schmandt, 1993). All this together leads to users not being able to focus on the task properly and hence this leads to that they doubt their own capability or performance at an earlier stage than they doubt the systems.

Many of the participants expressed it more precisely saying they thought it was too many options in the system. They said that: “There were seventy-eight thousand options to choose from. I didn’t get it, it was too complicated” P28 and “One should be able to put more options under the same
One contributing factor to participants thinking there were too many options was that the participants did not seem to understand the functionality of the key 0, “more options”. Even though the participants expressed that they did understand that there were three options in each menu, and then “for more options, press 0” they were not sure what to get if they pressed 0. The participants had several suggestions on what to find, such as “An escape function. Someplace to, well, then maybe you can get back to the main menu or end the call” P36 and “If I didn’t understand 1, 2, 3 then 0 will give me an explanation” P5. The participants pressed 0 to get help or to navigate but instead they got more options.

**Navigation**

Since there was too much information and options and the participants misinterpreted the meaning of 0, the participants had problems navigating between the menus in the system. The participants expressed this through the following quotes; “There was no way to go back. Normally you can go down and then return, but I didn’t get that feeling” P2 and “It wasn’t a problem to get into the menus, but it was harder to get out. I’m either impatient or that you don’t get information about the general stuff (global keys).” P26. The quotes show that the participants thought it was difficult or even impossible to navigate up, out from the menus. As one participant put it; “A way to go back didn’t exist” P2. A reason for this could be that they misinterpreted 0 to be a navigation key. Another reason is the fact that they did not realize that there were global keys for navigating in the system. As one of the participants said; “No, I didn’t know if they worked in the other (menus)” P28. Since they did not get out of the menus many participants expressed that they did not know where in the system they were; “It was too many options, I didn’t know where I was” P27.

Another response from the participants which can validate both that it was too much talking and too many options in the system was that they did not realize that they pressed # instead of *. When the participants pressed # they heard “# is not a valid option in this menu” This did not make the participants realize their mistake, instead they kept trying to press #. The participants did not get sufficient feedback to make them realize their mistake so that they could do anything about it. Also users
perform actions that they think are reasonable and therefore they do not think about them as mistakes even if they fail to perform a task (Cooper and Reimann, 2003). A common way to react when a mistake happens is to just keep acting in the “normal” way, not realizing that something is wrong and therefore changing their behavior. This is exactly what the participants did.

The problems navigating within the system can also depend on inconsistency in the system (Teitelbaum and Granda, 1983), lack of feedback (Faulkner, 2000) and problems in creating a mental model (Nielsen, 1993). These three factors will be described in further detail below.

Inconsistency
One contributing factor to participants reporting it hard to navigate can be that the system is inconsistent. In some menus options are placed inconsistent; “listen” is placed on 1 and “use” on 2 but in other menus they are placed in the opposite order, that is; “use” on 1 and “listen” on 2. This can make users confused and therefore they may not be able no navigate in a satisfactory way. One participant expressed it like this; “To listen, press 2, then to listen, press 1. I thought about it, now I’m gonna listen and then I’m gonna press 2, no they said 1. I might have pressed 2 just on reflex, strange…” P26. This quote shows that the participants not only reflected over the inconsistent placement of the options, but also that this inconsistency was confusing and would have led to them pressing the wrong key if they had not listened properly.

Another inconsistency in the system that the participants reflected over is that it was hard to know when they had to end their inputs with # or not. One example of this is when the users enter their pin code. The system tells the users to enter their pin and then finish with the #-key. When the users had done this, the system said: “Enter your pin code again”. The fact that the system says that they should finish by pressing # the first time but not the second, confused the participants. The following quotes express this; “Sometimes in the beginning it felt like you had to take a chance if it should be # or not.” P30 and “You don’t know if you should finish by pressing # or not” P14. This leads to participants guessing when to use # or not, which hence leads to feeling insecure about how the system works and have problems creating a mental model of it.
ANALYSIS

Lack of Feedback
The participants expressed that they did not know where in the system they were. This can be shown through the quote; “It was too many options, I didn’t know where I was” P27. A reason for this can be the fact that the system did not provide necessary feedback. When the participants re-recorded their absence greeting in the system many of them were unsure about which of the recorded greetings was active when they hung up. This was expressed by a participant as follows; “It would have been good with something that confirmed the choice when I was changing the greeting. Now any could have been active” P24. The participants got confused by the fact that the system never told them which greeting that was active but instead told them that they could “stop use” a greeting.

The participants had big problems navigating out of the menu, especially when they were working with their absence greetings. The system never told the participants how to get out from the menu, instead the system asked if they had finished with their task. When the participants answered the system that they had finished, they got out of the menu, but the lack of feedback meant that the participants did not realize this. One of the participants stated: “It’s good that they say “if you have finished, press *2” but you don’t understand that this goes up in the menus” P26. This statement was made after the participant had been given information from the test monitor about what really had happened when they pressed this key. The participants thought they were still in the same menu and did not know what to do. One participant said “if you pressed the wrong key, what are you supposed to do then? Hang up or how do you get back to the main menu; it might have been an option for that but I didn’t hear it” P14.

Problems Creating a Mental Model
Earlier it has been described that the participants thought it was difficult to know where they were in the system. The fact that they did not know where they were, lead to participants having difficulties in creating a mental model of the system (Preece et al., 2002). One participant said: “It was levels, but I didn’t really feel them” P30.

The participants that had created a mental model felt like it was destroyed when they started to navigate down in the menus. This was described by a participant with the quote: “At first, but it was ruined when
you started to go back, press 0” P29. The participants wanted to have a reference point to which they could return if they did not know where they were, they thought that a good point would be the main menu; “No matter where I am I want to get back to an already known place, the main menu.” P8 or “I went back step by step one at the time but I would have wanted one (option) which lead right back to the main menu, if you want to start all over” P30.

Even though many of the participants said that their mental model was destroyed they still had some sense of feeling about in which direction they had navigated. The participants saw the structure of the system as either a hierarchy with several levels, or as navigating forwards, into new rooms. The participants expressed the navigation as this; “The menu is from left to right and the options are from top to bottom” P22 and “Forwards, like into a new room” P14.

5.2.2 Quantitative Analysis
The results from the quantitative analysis are presented below. These results are complementary to the qualitative results and in some cases also validate the same. The way the analysis has been conducted is that hypotheses about predicted results have been formulated and then tested through ANOVA-tests.

Landline or Mobile
The users reported that using a mobile phone did not disturb them in the test session and that they did not miss any information using it, in respect to that they had to remove the mobile telephone from the ear every time they made a selection. Therefore one hypothesis formulated on the basis of this was that there was no difference in how the users performed depending on if the test was done with a mobile or a landline phone.

A significant difference (p<.05) was however found in scenario1 and scenario3 in how the participants performed in concern to testing the system via mobile or landline telephone. Therefore this hypothesis was partly rejected. The participants that used a landline phone performed better than participants using a mobile phone in these two scenarios. Even if the difference in time between the two groups are compensated with one second added to the participants in the landline group (See
4.2.5 for further details on this matter) the difference is still significant (p<.05). Further tests show that these differences can not be explained with differences in gender, age, procedure, or previous experience of touch-tone systems and voicemail applications.

Also a difference in how the two groups were classified depending on their performance were found (p<.05). The difference was in regard to the classification in scenario3, where the landline group was classified higher than the mobile. This means that on scenario3 the landline group performed better in total, both in time and in error-rate but in scenario1 the difference between the two groups performance was only in time. Therefore this means that the landline group in some scenarios performed faster but not necessarily better.

There was however no significant difference between how the participants rated the system in regard to if they were using a mobile phone or a landline phone. That no difference was found in this matter indicates that they did not get disturbed using a mobile phone to perform the test. However the significant difference found in time between the groups shows that there was in fact a difference even though it did not affect the participants mentally.

**Age and Gender**

The hypothesis was formulated regarding differences in background characteristics were that no difference with respect to the participants’ age or gender would occur in the performance or the ratings.

This hypothesis was found to be partially true. No significant difference in how the participants performed in time or how their performance got classified was found.

There was on the other hand a significant difference on rating2 and rating4 (p<.05) in how men and women rated the system. Men rated these significantly higher than women. The two ratings were about how much control the users felt over the system and how secure they were of that they had made the right selections. In these ratings the participants had to measure how good they were at handling the
system, therefore these ratings can be seen as more subjective than the others (see Appendix F).

The result shows that even if no difference in the performance between the genders was found, men rated that they had better control and felt more secure about their selections than the women did. One alternative explanation could be that there was a difference between the genders in how experienced they were in using touch-tone systems. This explanation was tested and rejected and therefore this can not be seen as an alternative explanation to the differences found.

There is no obvious explanation to this difference. But traditionally men are seen as more secure about themselves and have higher self-confidence than women do. Therefore the men in the study may have thought that the difficulties they experienced were due to the system, while the women blamed themselves and therefore this difference occurred. This is however just one possible explanation.

**Experience of Touch-Tone**

One hypothesis about inexperienced and experienced participants was also formulated on the basis of the test sessions. It seemed to be no difference in how the participants performed or rated the system depending on how experienced they were with touch-tone systems.

To examine this hypothesis the participants were classified into two distinct categories depending on their estimated usage. These were; participants that had used touch-tone systems fewer than 20 times and participants that had used touch-tone systems more than 20 times. These categories were given labels of inexperienced and experienced user respectively.

With these two categories in mind the hypothesis was proved to be true. There was no significant difference found in how the participants performed, in time or how their performance was classified. There was neither a difference in how they rated the system.

**Experience of Voicemail**

Another hypothesis formulated on the basis of the test sessions was that no difference between how the participants performed in time or rated the
ANALYSIS

system depending on if they use voicemail in their everyday life or not would occur.

The hypothesis was proved to be false. A significant difference in scenario1 and scenario3 (p<.05) was found in how the users performed in regards to if they use voicemail or not. The participants that use voicemail performed better than participants that do not use voicemail. Also a significant difference (p<.05) was found in how the participants rated their experienced control over the system, rating2. Participants using voicemail rated a higher experience of control than users who do not use voicemail.

These differences show that experience of using voicemail affects the users’ ability to perform in and feel control over a system that they never have been in contact with before. This indicates that the system follows standards and therefore a positive transfer effect can be seen between other voicemail systems and this particular voicemail system. Since this system closely resembles other voicemail system the learnability is good and therefore users with experience of voicemail performed better and felt more in control over the system than participants with no experience of voicemail.

Ratings
The five ratings given to the participants gave the following results;

![Figure 5: The grand mean of the ratings.](image)

The ratings range from 0-10, with 0 as the most negative point and 10 as the most positive point in regards to the question. A tendency to give the ratings a positive estimation can be seen. Also a tendency that the ratings are getting more and more positive can be seen even though the
ratings were not sequenced to get this result. To see the exact formulation of the ratings see Appendix F.

**Classifications of the Scenarios**
The classification frequencies of the participants on different scenarios show how difficult or easy the different scenarios were to perform. Since the scenarios contained different tasks it is impossible to compare them with each other. It is however possible to judge about what task that was most difficult etcetera. On each scenario the participants’ performance was classified according on their error rate and time consumed to perform the scenario. The classification had 5 levels.

Below the distribution of the classification on each scenario is presented;

- **5**, the log is correct and time consumed is less than 2 minutes.
- **4**, the log contains 1-2 errors and time consumed is 2-4 minutes.
- **3**, the log contains more than 3 errors and time consumed is more than 4 minutes.
- **2**, the participant used the system but was unable to perform the task.
- **1**, the participant was unable to use the system.

![Scenario1 - activating the account](image1)
![Scenario2 - answer on a new voice message](image2)
![Scenario3 - send a new voice message](image3)
Scenario 4.1: record a new personal greeting

Scenario 4.2: end the call with *9

Figure 6: The participants’ performance on each scenario.

The graphs above show that there was a big difference in how the participants performed in the different scenarios. The graphs show that scenario 2 and 4.2 turned out to be the most difficult scenarios to perform. In these scenarios 55% and 63% of the participants were not able to perform the tasks, classification 2. In the other scenarios it was only 2-18% that was not able to perform the tasks. The reason for this can be that scenario 2 and 4.2 forced the participants to navigate between the levels of the system in a higher extent than in the other scenarios. Scenario 2 itself did not force the participants to navigate between different levels of the system but since this scenario seemed to contain a difficult task which meant that the participants went into the wrong menus, they were forced to navigate up and down in the system.

Scenario 4.2 can also be seen as more complex than the other scenarios because of the explicit demand to navigate up in menus. Therefore the big difference in time reflects the poor performance on this task. The large amount of participants that were not able to perform the task, classification 2, can not be explained by this factor since the time consumed is not involved in this classification (see Appendix G).

*Estimating Reasonable Time*

The participants were also asked to estimate the reasonable time to spend on each task involved in scenario 1-4. The actual time the participants spent on each task seemed to be about the same as their estimated reasonable time to spend on each task. A 5 - 15% difference
was tolerated since time is difficult to estimate. However there was one exception to this, scenario2. This scenario involved answering a new voice message. 70% of the participants spent longer time on this task than they estimated as reasonable. The participants estimated that a maximum of 2 minutes was a reasonable time to spend on this task, but only 12 of the 40 users had an actual time under 2 minutes. This proves that this task was too hard to perform in this system and that it is not reasonable that it takes this long time.

5.3 Summary
To summarize the main findings from these both methods, usability evaluation and testing, it can be concluded that the system has many shortcomings that can be referred to heuristics in the area as well as to the participants opinions and comments. The shortcomings can often be related to the bad organization of the system, this includes heuristics such as short-cuts, feedback, visibility as well as the participants opinions that it was hard to navigate in the system and that there was too much information and too many options. On the other side the usability testing also generated many positive comments about the impressive functionality of the system and that plays an important part in the amount of effort the general users are willing to spend on using and learning the system. Below a summary of the quantitative findings and the identified usability problems found through the qualitative analysis are presented.

Quantitative Findings
- There is a difference in time using landline versus mobile
- There is a difference in performance and rating depending on experience of voicemail
- There is a difference in ratings due to gender
- There is no difference in performance due to age or gender
- There is no difference in performance or ratings depending on previous experience of touch-tone
### Identified Usability Problems

<table>
<thead>
<tr>
<th>Area</th>
<th>The problem</th>
<th>Reference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-cuts</td>
<td>“I suppose you know that you can press *1”. The system should not suppose things about the user.</td>
<td>E - Short-cuts</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>There is no feedback on users’ input. The users do not know which greeting is active when they have recorded one. Incorrect feedback; number of saved messages The users misinterpreted *2</td>
<td>E - Feedback</td>
<td>T - Lack of Feedback, P24</td>
</tr>
<tr>
<td>Consistency</td>
<td>Both action-key and key-action order Same options on different keys Sometimes #, sometimes not Not * in front of all global keys The terms satisfied -1 and use -1 are used for the same purpose.</td>
<td>E - Consistency</td>
<td>T - Inconsistency, P26,30</td>
</tr>
<tr>
<td>Terminology</td>
<td>“Standard scheme”, ”Personal assistant”</td>
<td>E - Terminology</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>The error messages are not sufficient The users blame themselves for their mistakes Confusion between * and #</td>
<td>E - Error</td>
<td>T - Organization &amp; Structure</td>
</tr>
<tr>
<td>Visibility</td>
<td>Formulations of greetings Hard to create a contact list</td>
<td>E - Visibility/Terminology</td>
<td>E - Visibility</td>
</tr>
<tr>
<td>Other</td>
<td>Too much information/options The users misinterpreted 0 “To stop using it, press 2” “If you make a mistake, press *3” – users do not make mistakes, it is the system’s fault. It takes too long time and is too difficult to respond on a new message. Too much talking after 8 and # No way to go back</td>
<td>T - Organization and Structure, P28,6 T - Organization and Structure, P36, 5 E - Errors</td>
<td>T - Organization and Structure, P10</td>
</tr>
</tbody>
</table>

Table 1: The main usability problems identified in MVAS 10.3.0
6 Redesign

This chapter presents the redesign of the system. First the way in which the proposals for redesign were done is illustrated with some example-changes. Thereafter two developed prototypes regarding the interaction style is presented and discussed. The chapter ends with the final redesign proposal of MVAS 10.3.0.

6.1 Example-Changes

Below an extract from the system will be presented with the proposed changes in bold. The changes have explanations to illustrate and explain why the changes are done. The flow used in the scenarios represent the flow the user most often are presented to, this means that some mistakes are included. These mistakes are included to present some problems that the user often experience.

The typical system flow when a user tries to reply on a message:

1. System: Welcome to messaging services
   You have “1” new message
   Message from “Susanne”
   “Time for deposit”
   “The message is played”
   Message saved
   You have no saved messages
   Main menu
   If you want your saved messages now, press 1
   If you want to send a voice message, press 2

The system should not give feedback that is false, instead of saying “no messages” it would be better to either say nothing or say the correct number, in this case “You have “1” saved message”.

It is unnecessary to say “now”, it gives the users a feeling that now is the only chance they will get of listening to their messages. Words unnecessary or loaded with subjective judgments should be avoided.
2. User: 2 (send a voice message)

3. System:  To what number?
   Type a phone number or a contact list
   number and
   then press the # key

4. User: *Waits for 3 seconds*

5. System: You are addressing your message
   Type the phone number or a contact list
   and finish with #

6. User: *waits for 3 seconds*

7. System: I suppose you know that you can
   Press *1 to listen to the main menu
   And press *9 or hang up to finish
   the call.

8. User: *1 (to main menu)

9. System: Main menu
   If you want your saved messages
   now, press 1

The system should not assume
things. Instead the formulation
should be; “To return to the
main menu, press *1…”

The options should be
formulated in action-key order
throughout the entire system.
That is; the system should say
“To return to the main menu,
press *1”.

The typical system flow when a user changes
the absence greeting for all calls:

1. System: You have no new messages
   No saved messages
   Main menu
   To send a voice message, press 2
   To work with your absence greetings,
   press 3
   To work with your mailbox settings,
   press 4
   For other choices, press 0

2. User: 3 (work with the absence greeting)

3. System: You are using your personal absence
   greeting for all calls
   To work with your personal absence
   greeting for all calls, press 1
   For the absence greeting for no

The information “you are using
your personal absence greeting
for all calls” is not enough
feedback for the user. Instead of
understanding that this greeting
is active, the user will think that
he or she is working with this
greeting right now. This
becomes especially clear when
the user returns to this menu
after having worked with the
greeting. The user will think it is
a loop. Further it is too much of
the same information at the
same time.
answer, press 2
When your phone is busy, press 3
For more choices, press 0

4. User: 1 (personal absence greeting for all calls)

5. System: *This absence greeting is active*
   To listen to the absence greeting, press 1
   To stop using it, press 2
   To make a new recording, press 3
   For more choices, press 0

6. User: 3 (make a new recording)

7. System: You may begin,
   finish by pressing #

8. User: *records greeting*
   *#*

9. System: To listen to the absence greeting,
   press 1
   To stop using it, press 2
   To make a new recording, press 3
   For more choices, press 0

10. User: *waits for 3 seconds*

11. System: You are working with an active
    absence greeting
    To listen to it, press 1
    To stop using it, press 2
    To make a new recording, press 3
    To choose between the prerecorded
    absence greetings, press 5
    To erase this greeting, press 6
    If you have finished working
    with this greeting, press *2

12. User: *2 (finished working with the absence
    greeting)

13. System: *You are using your personal absence
greeting for all calls*
    To work with your personal absence
    greeting for all calls, press 1
    For the absence greeting for no answers,
    press 2
    When your phone is busy,

   It is good to provide feedback about what greeting is active. This feedback should be given every time the user does something to change the status of the system. As it is now it is only presented once; the first time a greeting is activated.

   "To stop using it" is an unfortunate formulation. Instead the users should be told that the greeting is active, and that they can de-activate it through pressing 2.

   To keep the consistency of the system the rule of having only three options at a time should be kept throughout the entire system. The re-prompt should provide help, not overload the user with options.

   To provide a better visibility and feedback to the user this prompt should also say that the greeting is active, or that changes are saved. It is also important to tell the user that this command will take them back to the previous menu. Further, in this case it would be better to present * 1 instead of *2.
press 3
For more choices, press 0

14. User: *waits for 3 seconds*

15. System: You are working with your personal absence greetings

16. User: *2 (finished working with the absence greeting)*

17. System: main menu
To send a voice message, press 2
To work with your absence greetings, press 3
To change your mailbox settings, press 4
For other choices, press 0

18. User: *waits for 3 seconds*

19. System: You are listening to the main menu
To send a voice message, press 2
To record or change your absence greetings, press 3
To work with your mailbox settings, such as; changing your pin or contact lists, press 4
If you need more time, press 8
If you want to log into another mailbox, press *5
If you want to make an outbound call, press *8
If you want to end this call, press *9 or hang up
If you need more time, press 8

20. User: *9 (end the call)*

21. System: Good Bye!

The system should not tell the users what they are doing, but instead provide options of what they can do and feedback about what they have done. Instead of saying that “You are working with your personal absence greetings” the system should say which greeting that is active and what options the users have.

It is good to give further explanation about the options in the re-prompts.

The users will be confused if the same global key is presented two times in the same menu. All unnecessary information should be omitted to minimize the users' mental workload. Also, all global keys should be presented in the same way throughout the entire, with a * in front of them.

6.2 Prototypes

To be sure that the whole problem area had been covered, two different kinds of models were developed. The two models were then compared to find strengths and weaknesses with the different kinds of interaction techniques. The different techniques chosen were the skip and scan-model (see section 3.7.4) and the standard menu-style-model, which is
used in the system today. These techniques were chosen because they are widely separated in nature and therefore would generate many examples of how two different techniques would affect the system so that the best possible solution could be identified.

The two different interaction techniques are explained and illustrated with scenarios below.

6.2.1 *Skip and Scan Model*

Skip and scan provides only one option at a time to the user. The user chooses whether to navigate to the previous option - 7, to the next option - 9, or to select the current option - #. The # can be seen as an Ok-button, or Enter. The * can be seen as an Undo-button or an Escape-function. If the user wants to navigate back in the menus, or want to undo an action the *-key is used. 7 and 9 are chosen to navigate backward and forward because of their placement on the key-pad. In western countries keys placed to the left are seen as referring to back, while keys placed to the right are seen as referring forward. Because of cultural differences, these keys should be carefully considered so that they match the context they are used in (Sun, 2002).

The model has only four different commands for the user to remember. If these four commands are remembered the user will be able to fully use the system. In addition to these; there are global keys that function as shortcuts for experienced users. The global keys are clustered together, from 1 to 6. These are separated from the navigation keys on the key-pad, see Figure 7 (Löwgren and Stolterman, 1998).

The system provides feedback when an action has been performed or the state of the system has been changed by the user. If the user listens to all options in a menu without selecting one, the menu will be re-prompted. Before the re-prompt, some global keys about how to return to the main menu and how to get help will be presented.
Figure 7: Example of how the short-cuts in the skip and scan menu-style could be distributed on the keypad based on frequency of usage.

Figure 8: Graphical view of the skip and scan menu-style.

Below two scenarios using the skip and scan-model are described. In the scenarios extracts from when a user activates the account and re-records an absence greeting are presented.

**Scenario 1, extract from activating the account**

1. System: Your name is now active.
   To be able to use your mailbox you have to record a greeting. This greeting will be played every time you do not answer a call. Record the greeting after the tone, when you have finished, press #.
   “pip*”
2. User: “Hi, I can’t take your call right now, but you can leave a message after the beep” #

3. System: Greeting recorded
   To go on, press #.
   If you want to listen to the recording: press 9.
   If you want to re-record the greeting, press 7.

**Scenario 2, extract from changing a greeting**

1. System: To send a voice message, press #
   *waits for 1 sec*
   To hear the next option, press 9
   To hear the previous option, press 7

2. User: 9 (next option)

3. System: To work with your absence greetings, press #

4. User: # (work with absence greetings)

5. System: Your greeting for no answer is active.
   If you want to work with your greeting for no answer, press #

6. User: 9 (next option)

7. System: To work with the extended absence greeting, press #

8. User: # (work with extended absence greeting)

**6.2.2 Standard Menu-Style Model**

The model associates the options in the menus to the numbers on the key-pad. The options are arranged in menus, with three to four options in every menu. If there are more options than four they are presented under key 0 – more options. The options are arranged according to frequency, most frequent used items first (Shneiderman, 1998).

The global keys are played in re-prompts and under the help function, key 0. The global keys exist of two buttons, * and then a number. The * is necessary to avoid that the command is confused with other numbers such as pin and phone numbers, and also to not confuse options in menus with the global keys.
The system provides feedback when an action has been performed or the state of the system has been changed by the user. If the user listens to all options in a menu without selecting one, the menu will be re-prompted.

![Figure 9: Graphical view of the Standard menu-style](image)

Below two scenarios using the standard menu style-model are described. In the scenarios extracts from when a user activates the account and re-records an absence greeting are presented. In a real life situation the user would probably make selections sooner, but to give a better picture of how the model works, all prompts that are listen to are fully read out.

**Scenario 1, extract from activating the account**

1. **System:**
   Your name is now active.
   To be able to use your mailbox you have to record a greeting.
   This greeting will be played every time you do not answer a call.
   Record the greeting after the tone, when you have finished, press #.
   *pip*

2. **User:**
   “Hi, I can’t take your call right now, but you can leave a message after the beep” #
3. System: Greeting recorded
   To listen to the greeting, press 1
   To re-record it, press 2
   To go on, press #.

**Scenario 2, extract from changing a greeting**

1. System: If you want to send a voice message, press 2
   If you want to work with your absence greetings, press 3

2. User: 3 (work with absence greetings)

3. System: Your greeting for no answer is active.
   To work with your greeting for no answer, press 1
   To work with your extended absence greeting, press 2

4. User: 2 (work with extended absence greeting)

**6.2.3 Comparison of the Two Models**

The two models both have advantages and disadvantages. To decide which one of the models that should be used in the redesign of the system, several factors have been considered. The factors are; how to make the system consistent, the learnability and visibility of the system, if the system decreases mental workload and memory load, and if the system has potential for scalability. The latter includes questions such as; if it is easy to add new functionality, and if a transition to speech-recognition will be smooth.

**Consistency**

In the standard model it is difficult to keep the consistency throughout the whole system in concern to the options placement and order. However in the skip and scan model it is difficult to keep it consistent as an entirety (Teitelbaum and Granda, 1983). The scenarios above showed that it was a problem to perform the log-in session using a clean skip and scan-interaction. Activating the account is a task that a user only performs one time, but it also gives the user practice in how the system’s navigation works. Therefore it is important that the navigation in the log-in session is as similar to the rest of the system as possible. But if the rest of the system is not consistent, which is a problem in the standard menu-style model, it does not matter how
consistent the log-in session is with the rest, the system will still be difficult to use (Nielsen, 1993).

**Learnability**
The skip and scan-model is easier to learn since it is always the same keys to press (Schmandt, 1993). But the standard model can also be learned quite fast because it provides a better mapping between the actions and the key; it is possible to learn paths to specific options by heart to be able to use the system faster.

**Visibility**
In the standard model all options will be presented before the user has to make the decision. In the skip and scan-model only one choice is presented at a time. This gives the user a worse overview of the existing choices than the standard model does (Gilmore, 1991). But since the skip and scan-model’s structure gives a better mental model in regards to navigation because the same keys are used in every situation, the visibility with skip and scan is considered to be better than with the standard menu-style.

**Workload**
The skip and scan-model reduces both mental workload and memory load since it makes it easier to create a good mental model of the system, and because the navigation works in the same manner all the time (Nielsen, 1993). Also, the skip and scan-model do not put any demand on the user to remember specific commands as the standard menu-styles do. The standard model is however faster to use in regards to the number of keys the user has to press, especially for experienced users, and this reduces the workload (Schmandt, 1993). On the other hand, since the consistency is difficult to keep with this standard model, the workload will be increased (Nielsen, 1993).

**Scalability**
It is important to take the future into consideration when redesigning an interface. If the new system only will last for a few months before another redesign is needed it is probably a bad alternative, both financially and from a usability aspect. The system should support technical progress and make changes easy to implement instead of preventing such development. Also, when new functionality is
constantly added to a system it is important that the users still recognize the structure and are able to use the system. Therefore a redesign of a system must handle all these cases. A big problem today with touch-tone systems using the standard menu-style interaction is that the 12 buttons on the key pad are too few to handle all functionality the system provides. This problem will be solved by using the skip and scan-model since this model does not map the options with a specific number but instead gives the user the option to answer yes or no if that specific option is wanted.

Today many touch-tone interfaces are replaced with speech-recognition interfaces. The difference between an interface using skip and scan and an interface handling speech recognition is smaller than with the standard menu style. Therefore the transition from touch-tone to speech recognition will be easier to handle for the users in a skip and scan-interface because the change will not affect the interaction style and will therefore not be as obvious as with the standard menu-style.

6.2.4 Conclusion
When all the pros and cons above are weighted together it is concluded that for MVAS the standard menu-style model should be used. The reason for this is that the standard interaction style is used by millions of users of this system today and it is considered to be of a bigger importance to make an already used system more usable than to change it to a totally new one which the users have to learn. This decision is made on the basis of other similar situations where the standard version is kept although “better” versions from a usability point of view exist. One example of this is the computer keyboard where the QWERTY-layout is kept because it is standard, even though it is not as fast or as ergonomic to use as the Dvorak-layout. The skip and scan model is however seen as the future alternative. A change of MVAS to this interaction style is seen as the next big step toward a usable system that supports adding of new functionality which is a major advantage for systems like this that keeps evolving.
6.3 Redesign: System Description

When doing a redesign there is one important thing to take into consideration. Should the redesign apply to novice or experienced users or both? In this redesign decisions were made with the experienced user in mind. Since the learnability of the system is considered to be good inexperienced users will learn the commands quickly and then this order will only annoy them the first few times.

The structure of the old system is kept intact; that is the standard menu-style is used as interaction technique. This menu style has however been modified to some extent from the original one. The modifications refer to the structure of only presenting 3 options in every menu and then give the remaining options under key 0. In the redesign a decision was made that when the menu only contains 4 options all these should be presented straight away (see 3.3.6). Also when the menu contains a #-key or a global key these are presented as a forth option in a menu before key 0 is presented. If both # and a global key are presented in the same menu, the global key will be found under 0.

In the redesign the depth of the system was tried to be kept to a minimum (see 3.3.6). In some cases it was however decided to be a better option to rearrange options into new sub menus instead of having too many options presented in one menu (see Table 2, Visibility).

In the redesign the users are presented with a menu directly after they have listened to a new message. Since 55% of the users in the test sessions could not answer on a new message (see section 5.2.2) this was seen as a good solution to recover from this problem. This will also make the systems functionality more explicit.

The prompts in the redesign are always presented in action-key order which is shown preferable to key-action order (see 3.3.6)

Studies have shown that users experience it hard to navigate in abstract hierarchies (Cooper and Reimann, 2003) and therefore the formulations
of the global keys used for navigation have been made more explicit so the user understands how they are navigating.

In the redesign the presentation of global keys were changed so they were included in all re-prompts and also in some menus. They were however decided not to be put in an introduction to the system. This was due to the fact that users seldom want to listen to some help or introduction but instead want to get started to use the system straight away (see 2.2). Therefore the global keys of interest for the situation are presented in the re-prompts of the menus. In the redesign, feedback and error messages are included to a larger extent. These messages are kept as short as possible without losing the sufficient information, in a attempt to minimize the time the users get interrupted by them (see 3.3.5)

Also, language that was unnecessary, intimidating to the user or difficult to understand was excluded in the redesign (see 3.3.6).

This was just some of the main changes made. Below the system description of the redesign follows to illustrate more changes made and how these affect the flow in the system. The description contains three of the four scenarios used in the test sessions; Scenario 1, 2 and 4. Scenario 3 is excluded because this scenario is included in Scenario 2 and no major changes have been done regarding this task.

6.3.1 Scenario 1

The user activates the account.

1. System: Welcome to messaging services
   You have received a pin, now it is time to change it for security reasons.
   Enter your new pin, finish with #.

2. User: 111 #

3. System Sorry, your pin has to be “4” digits.
   Please enter your pin again, finish with #.

   (If the user’s pin contains a * an error message will be played; “Sorry, * is not a valid option.
   Please try again, finish with #)
4. User: 1111#

3. System: To confirm your pin, enter it again and finish with #.

4. User: 1111 #

5. System: Your pin has been changed.
   To be able to use your mailbox you have to record your name.
   This name will be used every time you send a message.
   Record your name after the tone, finish with #.
   *(Here record would be replaced with “tala in” in the Swedish version)*
   *pip*

6. User: Sarah #

7. System: Name recorded.
   To listen to the name, press 1.
   To re-record it, press 2.
   To go on, press #.

8. User: # (go on)

9. System: Your name is now active.
   To be able to use your mailbox you have to record a greeting.
   This greeting will be played every time you do not answer a call.
   Record the greeting after the tone, when you have finished, press #.
   *pip*

10. User: *records greeting* #

11. System: Greeting recorded
   To listen to the greeting, press 1.
   To re-record it, press 2.
   To go on, press #

12. User: # (go on)

13. System: This greeting is now active
    You have “1” new message

### 6.3.2 Scenario 2
The user listens to a new message, replies on it and then erases it. After
that the user listens to the second message.

1. System: Good afternoon
   You have “2” new messages
   Message from “Susanne”
   *Time for deposit*
*The message is played*
To listen to the message again, press 1
To respond, press 2
To erase it, press 3
To save it and go on, press #

2. User: 2 (respond)
3. System: To send a voice message to the sender, press 1
   To call the sender, press 2
   *(If the user waits the re-prompt is added with: to go back to the previous menu, press *2)*
4. User: 1 (send a voice message to the sender)
5. System: Record your message after the tone, when you have finished, press #
   *pip*
6. User: *records message*
   # (finished)
7. System: Message recorded
   To listen to the message, press 1
   To re-record it, press 2
   To send the message, press #
   For more options, press 0
8. User: # (send the message)
9. System: Message sent
   Back to the message you were listening to
   To listen again, press 1
   To respond, press 2
   To erase it, press 3
   To save it and go on, press #
10. User: 3 (erase)
11. System: Message erased
    Message from “Kristina”
    *Time for deposit*
    *The message is played*
    To listen again, press 1
12. User: # (go on)
13. System: That was the last new message
    “1” saved message
    Main menu
    If you want to work with your saved messages, press 1
18. User: *hangs up*
6.3.3 Scenario 3
The user makes a new recording of the personal absence greeting for no answer.

1. System: Good afternoon
   You have “no” new messages
   “1” saved message
   Main menu
   To work with your saved messages, press 1
   To send a voice message, press 2
   To work with your absence greetings, press 3
   To work with your mailbox settings, press 4

2. User: 3 (work with the absence greeting)

3. System: Your greeting for no answer is active
   (If other greetings like “Out of hours” are active, it will be told here)
   To work with your greeting for no answer, press 1
   To work with your extended absence greeting, press 2
   To work with other absence greetings, press 3
   For more options, press 0

4. User: 1 (greeting for all calls)

5. System: To listen to the current greeting, press 1
   To re-record it, press 2
   To de-activate this greeting, press 3
   To return to the main menu, press *1
   For more options, press 0

6. User: 2 (make a new recording)

7. System: Record the greeting after the tone, when you have finished, press #.
   *pip*

8. User: *records greeting*
   *#*

9. System: Greeting recorded and active
   To listen to the greeting, press 1
   To re-record it, press 2
   To de-activate this greeting, press 3
   To return to the main menu, press *1
   For more options, press 0

12. User: *1 (main menu)

17. System: main menu
To work with your saved messages, press 1
To send a voice message, press 2
To work with your absence greetings, press 3
To change your mailbox settings, press 4

18. User: *waits for 3 seconds*

19. System: You are listening to the main menu
  To send a voice message, press 2
  To record or change your absence greetings, press 3
  To work with your mailbox settings, such as changing your pin or contact lists, press 4
  If you want help, press *0
  If you want to log in to another mailbox, press *5
  If you need more time, press *7
  If you want to call another number, press *8
  If you want to end this call, press *9 or hang up

20. User: *9 (end the call)*

21. System: Goodbye!

### 6.4 Summary of Main Changes

The table below shows the major usability problems in the study and the changes made to improve these problems.

<table>
<thead>
<tr>
<th>Area</th>
<th>The problem</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-cuts</td>
<td>“I suppose you know that you can press *1”. The system should not suppose things about the user.</td>
<td>The formulation is changed to; “To return to the main menu, press *1…”</td>
</tr>
<tr>
<td>Feedback</td>
<td>There is no feedback on users’ input. The users do not know which greeting is active when they have recorded one. Incorrect feedback; number of saved messages The users misinterpreted *2</td>
<td>Short but sufficient feedback is given every time the users enters data or changes the system status in some way. Feedback is given when a greeting is recorded and activated. The user is told the correct number of saved messages. The formulation of *2 has been changed to more clearly state that *2 is the key that leads back to the previous menu. The prompt now reads “To return to the previous menu, press *2”. This key has sometimes been replaced with *1 in situations where this command are seen as more helpful.</td>
</tr>
<tr>
<td>Consistency</td>
<td>Both action-key and key-action order</td>
<td>The options are presented in action-key order throughout the entire system.</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Same options on different keys</td>
<td></td>
<td>The options are consistent and ordered in the same way where possible. The numbers used in a menu are also ordered logically; that is it is never ordered 1,3,2 but instead 1,2,3.</td>
</tr>
<tr>
<td>Sometimes #, sometimes not</td>
<td></td>
<td>Entering of data is always ended with #. # is also the button used to “go on” between menus. The #-key can therefore be seen as an OK button or as enter which can help the users remember when to press # or not.</td>
</tr>
<tr>
<td>Not * in front of all global keys</td>
<td></td>
<td>All global keys have *. See Appendix H.</td>
</tr>
<tr>
<td>The terms satisfied -1 and use -1 are used for the same purpose.</td>
<td></td>
<td>The same term is now used for the same option throughout the entire system.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Terminology</th>
<th>“Standard scheme”, “Personal assistant”</th>
<th>Terms like this has been deleted and replaced with more intuitive terms or explained more fully. Hence, when “standard scheme” is mentioned the user gets an explanation of what it means.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Error</th>
<th>The error messages are not sufficient</th>
<th>Error messages are given when an error occurs. These explain what error that has occurred so that the user can recover from it. See Scenario1, prompt 3.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The users blame themselves for their mistakes</td>
<td>The system is easier to handle so that no or fewer mistakes will occur.</td>
</tr>
<tr>
<td></td>
<td>Confusion between * and #</td>
<td>An error message is given if the wrong key is pressed so that the users understands why an error has occurred.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Visibility</th>
<th>Formulations of greetings</th>
<th>The formulation and structure of greetings are changed to be more intuitive and easier to handle. The terms used are changed but also the amount of different greetings that can be recorded. The greetings still cover as many cases as before. See Scenario3, prompt 3.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hard to create a contact list</td>
<td>A solution for this problem is not included in the system description of the redesign, but there are many ways to improve it. See section 5.1.2-Visibility. Which solution that is the best has to be investigated through further research.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th>Too much information/options</th>
<th>Unnecessary words, such as now or just, have been deleted and the structure of the options has gone through some minor changes. It is hard to do anything about this problem without losing the functionality of the system.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The users misinterpreted 0</td>
<td>The use of 0 has been minimized. It is difficult to change the prompt to be more explaining without annoying the user with too much talking.</td>
</tr>
<tr>
<td>Problem</td>
<td>Proposed Changes</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>“To stop using it, press 2”</td>
<td>The user gets feedback that the greeting is active and then an option to “de-active” the greeting is given.</td>
<td></td>
</tr>
<tr>
<td>“If you make a mistake, press *3” – users do not make mistakes, it is the system’s fault.</td>
<td>This prompt has been replaced with better error messages if the users make a mistake. This is better than to assume the fact that the users do make mistakes. See Scenario1, prompt 3.</td>
<td></td>
</tr>
<tr>
<td>It takes too long time and is too difficult to respond on a new message.</td>
<td>A menu has been added after every message, even in the auto-play setting to help the users respond, save and delete their messages. See Scenario2, prompt 1.</td>
<td></td>
</tr>
<tr>
<td>Too much talking after 8 and #</td>
<td>The unnecessary words have been deleted or removed to the re-prompt to make this prompt as short as possible after the 8 (now *7) or # is mentioned.</td>
<td></td>
</tr>
<tr>
<td>No way to go back</td>
<td>The global keys are presented more frequently to help the users navigate.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: The main usability problems identified in MVAS 10.3.0 and the proposed changes to correct the
7 Method Discussion

This chapter contains a discussion about the method and material used in the study.

7.1 Design

The study contained both quantitative and qualitative data. The reason for this was that a quantitative analysis was thought to come up with interesting results that could not been found in a qualitative study only.

In hindsight, the quantitative analysis did not contribute as much to the results as wanted, probably because the nature of the test. However, some interesting results were found that affected the interpretation of the qualitative results and therefore the quantitative analysis contributed to the entirety of the study.

To include 40 participants in a qualitative study can be seen as a lot. Despite this a decision to include all participants in both the quantitative and the qualitative data gathering of the study was made. To just collect quantitative data from the participants and not let them express what they thought about the system in an in-depth interview would leave the participants frustrated and not satisfied with the situation.

As it turned out, it was hard for the participants to understand the purpose of some parts of the study, especially the walkthrough. In this way having many participants was an advantage so that at least some participants would understand the purpose and give information and opinions of interest.
METHOD DISCUSSION

The disadvantage of including 40 participants in both the qualitative and quantitative part of the study is that it was time consuming. If the study had been limited in some way, for example to 20 participants, it would have been possible to spend more time on redesign and a testing of the redesign. This was however not prioritized in this study since it was the first time a usability test and evaluation was done on MVAS. It was seen as most important to do a deep study and well thought through redesign to provide a good foundation to stand on from when doing further tests and evaluations.

7.2 Usability Evaluation
Regarding the evaluation conducted it is possible that different or more usability problems would have been found if the evaluation had been conducted with more than two evaluators. The evaluators’ lack of experience of evaluating such complex and large systems also makes it possible for problems to pass by undiscovered.

Also, another set of heuristics used in the evaluation would have resulted in different types of usability problems found which also would have affected the results and redesign in a different manner.

7.3 Usability Testing
When the pretest was conducted the focus was on how the material was formulated and perceived by the participants. This focus is necessary but more resources should have been put on how to analyze the gathered data and see what the data can tell. The lack of this focus was obvious when it comes to certain parts of the test. These will be discussed below.

7.3.1 Background Questionnaire
The background questionnaire contained questions that in hindsight can be discussed. For example, the questionnaire asked the participants what tasks they use their mobile phone for. If they use it for calling, receiving calls, send or receive SMS, and/or send and receive voice messages. This question did not give any necessary information about the participants except if they use voicemail or not. Therefore a question if they use voicemail or not and how often they use it would have
served the same purpose. This question was necessary to be able to investigate if there were any difference due to the fact if the participants use voicemail or not.

Another question was about how many times the participant had used a touch-tone system before. This question investigated how experienced to touch-tone systems the participants were. The alternatives given to the participant concerning this question are discussable. In hindsight these alternatives are seen as not distinctive enough. It would probably have been better to let them estimate how often per day/week/month they use a touch-tone system instead of how many times they have used one. This might have been more distinctive and hence give more clear statistical results on how their experience affects the overall performance.

7.3.2 Walkthrough

The participants seemed to have problems understanding the purpose of the walkthrough. In the walkthrough the users were asked to think about a voicemail system and describe how they would want it to be if they could decide without any limitations. This seemed to be hard to grip for the users. They often understood the task itself and described a voicemail system, although they often failed in describing how their optimal a voicemail system should be organized. Instead they described an existing voicemail system, often the one they used today.

This lead to that many walkthroughs got a different focus than wanted. The focus was on how voicemail systems worked today and if they worked good or bad. The users said that they thought that the voicemail systems worked satisfying, and some said that they had not thought about it and therefore could not judge about it. The reason for this that they thought the system worked satisfying might be that they did not know how it could work in any other way and therefore this was to easiest answer to give.
7.3.3 System Settings
Some settings were made to the system used in the study to make it suitable to the test situation and the purpose of the study. One setting was that e-mail and fax was excluded so that the system only featured voicemail, the main application for the study. These features were excluded not to distract the users and to avoid making the system any larger. Another setting was that the system used had “auto-play of new messages”. This means that the new messages were played immediately when the user entered the system. This setting was made because it was thought to be the setting that most users would prefer. The walkthrough which raised this question also proved that this was the case. Most users wanted to hear their new messages straight away without any unnecessary information before or having to make any selections to reach them.

Even though the setting with “auto-play” was preferred by most of the participants, this led to other issues. When “auto-play” is active in the system, the new message is read and then the users are transferred to the main menu. From here the users have to go down to a sub menu, number 1 – listen to saved messages, to be able to for example respond or erase the message. If “auto-play” is not active this sequence is changed. Then the users do not hear the message straight way when they log in but are instead transferred to the main menu directly and have to go down to the submenu to be able to work with it. This sequence might have affected the participants’ bad performance on scenario2, when they had to respond to a new message and then delete it. If the “auto-play” had been deactivated the way to respond to a message might have been more intuitive, but not as appreciated.

7.4 Usability study
When doing usability evaluation and testing, it is easy to locate shortcomings regarding to the set of heuristics used. The question that has to be raised concerning this is if correcting all shortcomings really is an appropriate goal to strive for. For example, regarding the consistency; is it really worth striving for having consistency throughout the entire system if it means that the ordering and structure of some menus will be inappropriate due to this? The heuristics should
not be interpreted literally but instead used as guidance to help solve usability problems within the appropriate context of use.

Identifying shortcomings in a system is an easy task, but correcting these shortcomings without creating new usability problems is much harder. Therefore all corrections have to be carefully considered and weighted back and forth so that the best possible solution is chosen. The solution might be the best one even though it might not cover up all the aspects of the identified usability problems.

7.5 Generalization

The intended segment of this study is really wide and it is hard to judge about if the intended segment was actually reached in the recruitment of participants or not. But since all participants tested had been in contact with a touch-tone system in advance and many of them also mentioned a voicemail application as an example of this, many of the participants were within the intended segment.

Consequently a generalization of the results can be done to all end-users of Swedish telephone-operators because of the unambiguous leanings seen in the material. But it is important to remember that the design solutions presented in the chapter “Proposals for Redesign” apply mainly to the context used for this application, not for all voicemail systems.
8 Future Work and Conclusion

This chapter contains suggestions for future work and the conclusion of this thesis.

8.1 Future Work

There is still a lot to do regarding the usability of MVAS. This thesis describes a study of the general usability of the system but further research of the system will be needed to go deeper into the structure of the system to explore all aspects of it.

Even though the proposals for redesign are seen to solve most of the usability problems further user studies need to be performed. These studies should investigate how the changes affect how the users’ perceive the usability of the new improved system and if even more changes need to be done. Only when this is done a way which for sure will improve the system can be found.

Also the skip and scan-model should be implemented and tested to see whether or not it should be considered to use in the future. If this is the case; usability studies are needed to decide in which way this model best can be used. If extensive testing is done from the start, the system will have a better usability and fewer changes will be needed when the system is implemented in a final version.

Since the system in this study was limited to the voicemail feature, more research is needed to investigate the usability on the remaining parts of the system, such as the fax, e-mail and video features. It is important that these features work in a satisfying way independently as well as in harmony with the rest of the system. It is important that user
studies are conducted all the time, to create a continuity of the usability perspective as the system evolves.

Concerning touch-tone systems in general there is a lack of research that is surprising. These systems have been used since the 50’s and yet there is very little research done on how users perceive them. Research needs to be done both to come up with new interaction models and how the existing models are perceived by the users.

8.2 Conclusion

This thesis describes a mainly qualitative study where the interface of MVAS is evaluated and tested to come up with how the interface should be designed to be easier to use. The results show that the usability of MVAS is poor but the functionality of the same is both impressive and appreciated. The redesign of the interface should be further analyzed to see if even more changes are needed. This study was rather extensive and created a broad foundation for further usability testing. It is however important to keep investigating the usability of the system, but these studies do not need to be as broad as this one. It is more important to conduct them continuously.
9 References


REFERENCES


Miller, G. A. *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review. Vol 63. 1956*


References


REFERENCES


REFERENCES

URL References
AT & T Labs-Research
URL: http://www.research.att.com/history/64touch.html,

Claritus
URL: www.claritus.co.uk./dtmf.html ,

Mobeon AB
URL: http://www.mobeon.com

Techstuff.ca
URL: http://www.techstuff.ca/archives

800 Voice Mail Store
URL: http://www.800voicemailstore.com/voice-mail-history.htm
Accessed; March, 2004
Appendix A: Instructions

Generellt


Det är viktigt att poängtera att det inte är du som testas utan systemet och alla personuppgifter som du lämnar kommer att behandlas konfidentiellt, dvs. ditt namn kommer inte nämnas någonstans.

Vi spelar in ljudet med den här minidiscen. Jag är försöksledare och kommer att ansvara för försöket och Kristina/Susanne är observatör dvs. hon kommer att föra anteckningar av vad som händer under försöket.

Om du känner att du av någon anledning inte kan genomföra försöket så får du självklart avbryta utan att behöva ge någon motivering.

Skulle något hända så att du behöver hjälp är det bara att fråga mig, men försök att klara dig på egen hand så långt det är möjligt.

Testet beräknas ta ca en timme.

***Ge bakgrundskäten!!***

***Kolla om de svarat att de inte använder mobil. Låt de då göra systemtestet med fast telefon.***

Walkthrough


***Gör Walkthroughn!!***

System-test

Du ska nu få använda röstmeddelandesystemet. Har du hört talas om systemet förut? Vad isf?? Några detaljer?
APPENDIX

Du ska få göra fyra scenarion, alla involverar att använda systemet.

*** Vid mobile***Efter varje scenario vill jag att du ger telefonen till mig så att jag kan skriva av tid och knapprtryckningar. Det är för att vi ska veta vad som händer i systemet.

*** Vid Landline*** Jag kommer att skriva ner vilka knappar du trycker på medan du gör scenariet. Det är för att vi ska veta vad som händer i systemet.

Har du några frågor??


***Ge 1:a scenariot.***

Du kan börja när du vill.

***Efter varje scenario skriva upp tid, knapprtryckningar.***
Appendix B: Background Questionnaire

1. Kön
   □ Man   □ Kvinna

2. Ålder....................

3. Är du en person som gärna vill ha det nyaste när det gäller tekniska produkter? (Markera ditt svar med ett kryss)
   Nej, ___________________________ Ja, ofta
   aldrig

4. Har du mobiltelefon?
   □ Ja   □ Nej

   Om Ja,
   4.1 hur ofta använder du mobiltelefon?
   □ mer än 1 ggr/dag   □ 1 ggr/dag   □ någon gång per vecka
   □ någon gång per månad   □ mer sällan

   4.2 Till vad använder du mobilen? (kryssa i alla passande alternativ)
   □ ringa   □ bli nådd   □ skicka sms   □ skicka/ta emot röstmeddelanden

   4.3 När använder du mobilen? (kryssa i alla passande alternativ)
   □ i jobbet   □ privat

   □ Ja   □ Nej
   Om Ja, vilka? _____________________________

   5.1. Hur många gånger har du använt tonvalssystem?
   □ 1-5 ggr   □ 6-20 ggr   □ mer än 20 ggr
Appendix C: Walkthrough Heuristics

Detta är några tumregler för hur försöksledaren ska agera under walkthroughn;

Om walkthroughn stannar upp ska försöksledaren fråga ” vad vill du göra nu?”
Om försöksdeltagaren är passiv, säger ”jag vet inte” eller ”vad menar du?” ska
försöksledaren fråga; ”Vill du x eller inte?” X ska ersättas med nästa objekt på listan
nedan som inte har gåtts igenom.

Om försöksdeltagaren svarar ja, ska försöksledaren fråga “Hur vill du x?”, T. ex. ”Vill du
ringa eller hur vill du göra?” Om deltagaren svarar nej ska ledaren fråga; “ Vad vill du
göra nu/med y då?”

Om walkthroughg inte kommer någonvart bör testledaren leda in samtalat på om
deltagaren använder voicemail och ifall till vad. Därifrån ska ledaren sedan återigen
försöka leda samtalen in på hur deltagaren vill att dennes röstbrevlåda ska se ut.

Ämnen att gå igenom

**Funktioner**
- Vilka **funktioner** vill du ha?

**Lyssna**
- **Hur** ska meddelandet preseteras?
  - På en lista?
  - Vad presenteras först?
  - Vilka **detaljer bör presenteras**?

**Spara**
- **När**
  - Inbox = sparade meddelanden?
  - När ett meddelande lästs upp, vill ha alternativ att spara
    eller ska det sparer automatiskt?
  - Ordningsföljd i inboxen, sist inkomna först?

**Ta bort**
- **När**?
  - Fråga- säker på meddelandet ska tas bort?
  - Komma åt redan borttagna meddelanden igen?

**Hälsningsfras**
- Något man **använder**? En gång/ ändra flera gånger?
  - Olika hälsningar vid olika tillfällen?
Appendix D: Scenarios

Scenario 1 – aktivera ditt konto

Följ instruktionerna för att aktivera ditt konto.

Ring din röstbrevlåda genom att trycka på knapparna

Byt pinkod till: 1111.

Ditt namn är: Kim

Din hälsningsfras ska lyda: ”Jag kan inte svara just nu men lämna gärna ett meddelande efter pipet”

Avsluta samtalet genom att lägga på luren när du hör ”Du har 1 nytt meddelande”.

Scenario 2 – lyssna och svara på meddelande

Ring in till din röstbrevlåda och lyssna på dina nya meddelanden.

Svara på meddelandet från Susanne.
Svara genom att skicka ett meddelande till avsändaren där du säger frasen: ”Hej, det är Kim. Jag kan inte ses ikväll, hör av dig!”


Scenario 3 – skicka meddelande

Du ska nu spela in och skicka ett röstmeddelande.

Skicka röstmeddelandet till nummer: 73 325 85 07
Meddelandet ska lyda ”Hej, det är Kim, jag kommer med tåget kl 4 ikväll. Ses då!”

När du spelat in meddelandet, lyssna på det för att försäkra dig om att det blev bra.

Skicka meddelandet och avsluta sedan samtalet genom att lägga på luren.
Scenario 4 – ändra hälsningsfras

Du ska nu ändra din personliga hälsningsfras som de som ringer till dig hör när du inte kan svara. Hälsningsfrasen ska gälla för alla inkommande samtal.

Hälsningsfrasen ska lyda: "Hej det är Kim. Jag kan inte svara just nu, men säg ditt namn efter pipet så ringer jag upp."

Lyssna på din inspelning och se till att det är den som gäller för alla samtal innan du avslutar genom att använda det kommando som avslutar samtalet. (inte genom att lägga på luren)
Appendix E: In-depth Interview

Djupintervju

Nu tänkte jag ställa några frågor om vad du tyckte om systemet.

* = skattningar

Systemet

*******Hur upplevde du systemet i sin helhet?*******

Var det något som du upplevde som bra med systemet?

Var det någonting som var irriterande med systemet?

Har du något förslag på förbättring?

*******Upplevde du att det var du hade kontroll över systemet?*******

Förstod du all information som systemet gav?

Vad tyckte du om menyerna? Hur informationen var uppdelad i menyer?

Något som var bra/respektive dålig vad du kan minnas?

Tänkte du på att menyerna var uppbyggda med tre val (1,2,3) och sedan ”för övriga val, tryck 0”?

Vad tyckte du om det?

Tänkte du på att det fanns knapptryckningar som har samma funktion i alla menyer? (för paus, tryck 8)

Förstod du att de gällde överallt?

Använde du dem någon gång?

Vad tycker du om dem?

Navigering

*******Hur tyckte du det var att navigera i systemet?*******

Var det lätt eller svårt att göra val i menyerna?
Hur tänkte du när du skulle göra val?

********Hur säker kände du dig på att du gjorde ”rätt” val?********

Tryckte du fel någon gång i menyerna?
- Minns du varför du tryckte fel?

********Var det lätt eller svårt att komma rätt igen?********
- Hur gjorde du för att rätta till misstag?

Systemets attityd

Vad tyckte du om rösten?
- Skulle du föredra ”en röst” istället för en ”upphackad”?
- Varför?

Vad tyckte du om dialogen / Hur kände du att du blev bemött? Artigt/oartigt, vänligt/ovänligt, bra/dåligt?

Mobil/fast

********Var det lätt eller svårt att använda knapparna på telefonen? ********

Missade du någon information när du tog bort telefonen från örat?

Påverkade det dig? På vilket sätt?

Mental modell

Visualiserade du dig systemets struktur i huvudet?
- Kan du beskriva? (ge ev. papper och penna)
- Hur tänkte du dig att du navigerade? – uppåt, neråt, sidledes?


I vilken ordning vill du att de ska ligga?

Om de ligger i inboxen, är de sparade eller finns de bara där?

Tänker du dig dem som i en lista/ser du en ”email-inkorg” framför dig?

Något annat som du tänkte på? /vill tillägga?
Appendix F: Ratings

Markera ditt svar med ett kryss.

1. Hur upplevde du systemet i sin helhet?
   Dåligt ................................................ Bra

2. Upplevde du att du hade kontroll över systemet eller att systemet kontrollerade dig?
   Systemet .................................................. Jag
   Kontrollerade ......................................... Kontrollerade

3. Hur tyckte du det var att navigera i systemet?
   Svårt .................................................... Lätt

4. Hur säker kände du dig på att du gjorde rätt val?
   Inte alls ................................................... Mycket

5. Hur var det att använda knapparna på telefonen?
   Svårt .................................................... Lätt
Appendix G: Successful Completion Criteria

This is how the scenarios were classified;

A key press is a something that leads to a selection, for example 1, *2, 1111#. A key press is not necessary one digit.
An error is defined as a key press that is not included in the sequence of key pressed that leads to a successful completion of the task in question.

5 = the log is exactly the same as the correct or alternative log and time consumed is less than 2 minutes.

4 = the log includes 1-2 errors, but the task is performed correctly or time consumed is 2-4 minutes.

3 = the participant got lost in the menus, e.g. more than 3 errors and/or time consumed is more than 4 minutes.

2 = the participant used the system but was not able to perform the task.

1 = the participant was not able to perform the task.
Appendix H: Redesign of the Global Key Pause

To be able to keep the global keys consistent, the command Pause – 8 was changed to *7.

*7 already had a function; to “Listen to the previous message” in menu 1- Listen to saved messages. Since this functionality wanted to be kept intact, the following table and text will describe how the changes will be structured.

<table>
<thead>
<tr>
<th>Function</th>
<th>Default key</th>
<th>Altered key</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pause</td>
<td>8</td>
<td>*7</td>
<td></td>
</tr>
<tr>
<td>Listen to previous message</td>
<td>*7</td>
<td>7</td>
<td>When the user enters the menu after listening to a message, the option to listen to the previous message is given. The system says: “To listen to the previous message, press 7”.</td>
</tr>
<tr>
<td>Rewind 3 sec.</td>
<td>7</td>
<td>7</td>
<td>While the user listens to a message, 7 means to rewind 3 seconds. This will be told in the option *0 – Help. The system will say: “To rewind a few seconds while listening to the message, press 7”.</td>
</tr>
</tbody>
</table>

To summarize; when a user listens to a message, the key 7 means to rewind a few seconds. However, when the user enters the menu after the message is played, the key 7 means to go back to the previous message. This command is working the same way as a CD-player. Before the message has started to play, 7 means previous message, and while the message is playing, 7 means to rewind.

Key 9, that is to go forward or fast forwarding a message will work in the exact same way.
Title: For Happy Users, press 1-Investigating and improving the usability of a touch-tone interface.

Title: Utvärdering och förbättring av användbarheten i ett tonvalsgränssnitt.

Authors: Susanne Wedin & Kristina Carlander

Abstract

Touch-tone interfaces are today widely used in help-centers and support services. Studies have shown that interfaces like these have many limitations and are therefore hard to design. MVAS is a voicemail interface using touch-tone input for navigation. Today, shortcomings in the interface limit the users’ ability to use the functionality in a satisfying way. This thesis describes a mainly qualitative study which evaluates and tests the interface of MVAS to come up with how the interface should be designed to be easier to use. The results show that the usability of MVAS is poor but the functionality of the same is both impressive and appreciated. The suggested redesign of the system, based on the identified usability problems, considers both the interaction model used in the interface as well as the conformity to the set of heuristics used in the evaluation. The proposed redesign keeps all the functionality in the system intact and also makes the functionality more explicit through improving the usability. A more explicit structure will facilitate usage of a larger portion of the functionality. However, the limitation of the key-pad affects the redesign so the most favorable design is unreachable. If the interaction model is changed or furthered developed to allow speech input the limitations experienced with the current redesign will diminish and a higher degree of usability can be reached.

Keywords