Key Usability and Ethical Issues in the NAVI programme (KEN)

Deliverable 3

Products and Services for Personal Navigation – Usability Design

Part I

Usability Design and Evaluation Methods

Version 2.1

Eija Kaasinen (ed.)

VTT Information Technology

Last modified on 20.12.02
KEN_D3_I_21.doc

Version history

Version	Date	Author(s)	Description
0.1	22.05.2001	Eija Kaasinen	First draft
0.2	26.6.2001	Eija Kaasinen Matti Helin Juha Kolari Juha Luoma Rolf Södergård	Contributions from KEN partners
1.0	29.6.2001	Eija Kaasinen Ari Ahonen Veikko Ikonen Sari Lehtola	Chapter 5, Usability evaluation of some commercially available navigation products and services Abstract, Conclusions
1.1	16.7.2001	Ari Ahonen	Language-check by Richard Walker Minor modifications
2.01	10.9.2002	Eija Kaasinen	New chapters: Chapter 5: Related research – literature studies Chapter 7: Case studies with NAVI-projects Chapter 8: Usability guidelines
2.02	16.9.02	Eija Kaasinen	Deliverable divided into Parts I, II and III Chapter 5 Human-Centred Design in a nutshell
2.03	15.10.02	Eija Kaasinen	Small additions to chapter 1 and 3.5.1. Abstract, conclusions
2.04	8.11.02	Eija Kaasinen	Language-check by Richard Walker
2.1	10.12.02	Eija Kaasinen	Tiivistelmä

Contact information

Eija Kaasinen
VTT Information Technology
P.O. Box 1206, FIN-33101 Tampere, Finland
Street Address: Sinitaival 6, Tampere
Tel. +358 3 3163 323, fax +358 3 317 4102
Email: Eija.Kaasinen@vtt.fi
Web: http://www.vtt.fi/tte

Tiivistelmä

KEN-projektin (Käytettävyys ja Eettinen arviointi NAVI-ohjelmassa) ensimmäinen julkaisu, *Basics of Human-Centred Design in Personal Navigation* (D1, julkaistu keväällä 2001) esitteli käytettävyyden suunnittelun ja arvioinnin yleisiä periaatteita sekä erilaisia käytettävyyden suunnittelun ohjeistoja (guidelines). Näitä ohjeita voidaan käyttää perustana myös henkilökohtaisen navigoinnin palveluiden käytettävyyden kehittämiseen.

KEN-projektissa olemme tunnistaneet henkilökohtaisen navigoinnin tuotteiden ja palveluiden kehittämiseen sopivimmat käytettävyysmenetelmät. Projektissa on tutkittu myös erilaisia navigointituotteita ja palveluja ja tämän työn tuloksena on koostettu eri tyyppisten henkilökohtaisen navigoinnin tuotteiden ja palveluiden käytettävyyden suunnitteluohjeita. Tämä raportti syventää D1-raporttia keskittymällä erityisesti henkilökohtaisen navigoinnin tuotteiden ja palveluiden ominaisuuksiin. Samoin kuin teimme raportissa D1, jaamme tässäkin raportissa ohjeet käytettävyyden suunnittelun ja arvioinnin menetelmiin (tämä D3-raportin osa I) sekä käytettävyyden suunnitteluohjeisiin (D3, osa III). Menetelmäosuudessa neuvotaan, miten käyttäjävaatimuksia määritellään, ja miten evaluointi tulee toteuttaa. Suunnitteluohjeissa esitellään ratkaisuja, joiden käytettävyys on osoittautunut hyväksi jo olemassa olevissa tuotteissa.

Käyttäjäkeskeisen suunnittelun ydin on käyttäjävaatimusten tunnistaminen mahdollisimman oikein ja mahdollisimman varhaisessa vaiheessa suunnitteluprosessia. Tässä raportissa kuvaamme menetelmiä, jotka soveltuvat käyttäjävaatimusten määrittelyyn tuotekonseptia tai itse tuotetta suunniteltaessa. Henkilökohtaisen navigoinnin laitteet on tarkoitettu liikkuvalle käyttäjälle, ja laitteet ovat siksi kooltaan pieniä. Tarkastelemme miten laitteiden pieni koko ja mukana kuljetettavuuden vaatimus voidaan ottaa huomioon käyttäjäevaluoinnin suunnittelussa ja toteutuksessa. Ajoneuvolaitteen käyttäjä voi kohdentaa vain osan havaintokyvystään ajoneuvolaitteeseen, sillä käyttäjän on keskityttävä itse ajamiseen. Vastaava tilanne on tyypillinen navigointilaitteille, oli laite sitten suunniteltu ajoneuvonavigointiin tai jalankulkijan käyttöön. Esittelemme tässä raportissa ajoneuvolaitteiden evaluoinnin periaatteita. Menetelmiä voidaan soveltaa myös henkilökohtaisen navigoinnin laitteiden ja palvelujen evaluointiin.

Raportin lopussa kerrataan tiivistetysti käyttäjäkeskeisen suunnittelun (Human-Centred Design) periaatteet. Tätä koostetta voi käyttää muistilistana projektin eri vaiheissa toteutettavista toimenpiteistä.

Abstract

The first deliverable by the KEN project, Deliverable 1, Basics of Human-Centred Design in Personal Navigation, presented general principles for usability design and evaluation as well as general usability design guidelines and standards that can be utilised in designing products and services for personal navigation.

The KEN project has now identified most suitable usability design and evaluation methods for navigation products and services. We have also studied different navigation products and services to identify general usability design guidelines for different types of products for personal navigation. Based on this work, we now enhance the contents of Deliverable 1, concentrating on issues specific to Personal Navigation. As in Deliverable 1, we classify the guidance as methods (this report) and guidelines (Part III of this deliverable). Methods give guidance on defining user requirements and evaluating products and services, whereas guidelines present design solutions that have already proved to be usable in existing products.

Identifying user requirements correctly and as early as possible is one of the key issues in human-centred design. We describe some useful methods to define user requirements in designing a product concept or in designing the product itself. Navigation devices are by nature intended for mobile use and that is why they are often handheld-sized. In this report we describe some specific issues that should be taken into account when evaluating small handheld devices and services for them. When using in-vehicle devices, the user can devote only partial attention to the devices, his/her main attention being directed at driving. The same situation is typical for navigation devices, whether intended for drivers or pedestrians. We describe in this report some specific issues that should be taken into account when evaluating in-vehicle devices. These methods can be utilised in evaluating any kind of navigation devices.

We conclude this report with a description of the very basics of Human-Centred Design approach. This description can be used as a checklist for human-centred design activities in a design project.

Table of Contents

1 Introduction

1.1 Purpose

Key Usability and Ethical Issues in the NAVI programme (KEN project) is one of the horizontal support projects in the Personal Navigation (NAVI) programme of the Ministry of Transport and Communications in Finland. The aim of the KEN project is to ensure that usability and ethical issues are taken into account in the projects of the NAVI programme. Together with the projects, we have identified and solved usability and ethical problems related to personal navigation.

One of the tasks of the KEN project is to give guidance for designing usable navigation products and services. The first deliverable by the KEN project, Deliverable 1, Basics of Human-Centred Design in Personal Navigation, presented general principles for usability design and evaluation as well as general usability design guidelines and standards that can be utilised in designing products and services for personal navigation.

The KEN project has identified the most suitable usability design and evaluation methods for navigation products and services. We have also studied different navigation products and services to identify general usability design guidelines for different types of products for personal navigation. Based on this work, we enhance the contents of Deliverable 1, concentrating on issues specific to Personal Navigation. As in Deliverable 1, we classify the guidance as methods and guidelines. Methods give guidance on defining user requirements and evaluating products and services, whereas guidelines present design solutions that have already proved to be usable in previous products.

This report is targeted at anyone participating in the design of products and services for personal navigation and does not require any background knowledge about personal navigation.

1.2 Scope

This is the second version of Deliverable 3. This version enhances the first version of the deliverable by presenting overviews of related research as literature reviews. We also present the results of our case usability studies with NAVI projects and organisations. This version of deliverable 3 sums up the results of the KEN project as a usability design handbook for navigation products and services, presented in part III of this report. Two case studies are still going on and will be added to this report later. These are: studies of the usage of mobile maps (a case study with the Gimodig project) and studies of location-based search service (a case study with the Navisearch project).

The deliverable has been divided into three parts:

Part I Usability design and evaluation methods

Part II Related research

Part III Case studies and usability guidelines

Part I complements the overview of usability design and evaluation methods presented in KEN deliverable 1 by presenting methods specific to mobile products and services. The contents of Part I were originally published in version 1 of this report, in the summer of 2001.

Part II presents an overview of other published research results on key application or technology fields in personal navigation.

Part III presents the results of our usability design and evaluation case studies. Part of the cases studies deal with commercial products and services but the major part encompasses case studies conducted together with NAVI projects and organisations. The case studies dealing with commercial products were originally published in version 1 of this report, in the summer of 2001.

1.3 Definitions, Acronyms and Abbreviations

GPS	Global Positioning System
GSM	Global System for Mobile Communications
HMI	Human-Machine Interface
PN	Personal Navigation
PNS	Personal Navigation System
WAP	Wireless Application Protocol. A lightweight protocol to provide browser-based services to small mobile devices.

1.4 Overview of Part I

Identifying user requirements correctly and as early as possible is one of the key issues in human-centred design. In Chapter 2 we describe some useful methods to define user requirements in designing a product concept or in designing the product itself. Navigation devices are by nature intended for mobile use and that is why they are often handheld-sized. In Chapter 3 we describe some specific issues that should be taken into account when evaluating small-screen devices. When using in-vehicle devices, the user can devote only partial attention to the devices, his/her main attention being directed at driving. The same situation is typical for navigation devices, whether intended for drivers or pedestrians. In Chapter 4, we describe some specific issues that should be taken into account when evaluating in-vehicle devices. These methods can be utilised in evaluating any kind of navigation devices. Chapter 5 concludes this report by presenting the very basics of human-centred design approach.

2 Methods for Defining User Requirements

Matti Helin
Helsinki University of Technology

2.1 Introduction

The purpose of this chapter is to introduce methods for defining user requirements and to discuss why it is important to find user requirements as early as possible in the product design process.

Defining user requirements is one stage of user-centred design. It is a part of usability engineering that can and should be applied from the very beginning of a project. Defining requirements as early as possible results in a more usable and more useful product and will also shorten the time usability design takes later in the project. There are two kinds of requirements: 1) what users want, and 2) how they want to achieve it.

Defining what users want to do is called conceptual design. In the conceptual design stage the design team first gathers user information by various methods. User requirements for the concept are defined on the basis of this information. After that, several concepts are designed. The concepts are tested and evaluated with users and customers. They are combined with each other and refined. The result is a few concepts that can all be used for further product design. Usually, one of them is selected and that is the concept for the product.

In the product design stage the design team should first ask how users use things and how they will use the product. Usually, the design team interviews and observes the users or uses some other available methods to gather user data. The resulting information is interpreted and the user requirements for the product are defined. After that, a few prototypes and testing should be done to determine whether the requirements are right and how to fulfil them.

This chapter is based on previous research on the subject and on two interviews. The interviewees are HUT researchers specialised in user-centred product design. The first interviewee is a 31-year-old woman who has been working on user-centred product design since 1987. She focuses on usability engineering in product design. The second interviewee is a 27-year-old woman who has been working on user-centred product design since 1994. Her speciality is conceptual design. Both interviewees have written several articles on the subject. All quotations from the interviews are marked with (I1) for the first interview or (I2) for the second interview.

2.2 User-Centred Product Design and User Requirements

User-centred product design means a product design method where from the very beginning of the project there is contact with the users of the product and the product is designed from the point of view of the user. To do user-centred product design, the design team has to have contacts with the users to find out from them how the product should be designed. User requirements are formed from users' behaviour and actions. It is not enough to just guess what users want to do – it must be confirmed from the users.

According to the first interviewee, one of the worst problems in user-centred product design is that designers and marketing personnel discuss among themselves what the user wants and do not actually visit the real users.

“In user-centred product design there are three main principles: Early focus on the users, iterative design and empirical testing. (I1)” Defining the user requirements is the way to focus early on the user. User requirements tell of users' needs and expectations. If the product does not match the user's needs, (s)he is not likely to be satisfied with the product. User requirements depend on users’ likes and dislikes, user environment, users’ experience and generally on users’ behaviour. There are several methods of finding user requirements but the thing in common for all the methods is that the designers have to communicate extensively with the users.

There are basically two types of user requirements: what users want to do and how they want to do it. Finding the user requirements can also be seen in two stages. The conceptual design stage gives the concept of the design and the main idea of what users want and what they want to do with it. The next product design stage is to define how users want to do it.

2.3 Purpose of Defining User Requirements

Finding user requirements is one of the most important parts of designing easy-to-use devices or systems. If the requirements are not met, the consequences could be serious. Not only is it likely that the product will be less usable for the users, but it is also possible that the users do not even need the product and the whole project will be a failure. If user requirements are defined correctly, the project is more likely to be a success.

> There is some research on the subject indicating that successful companies and projects use an above-average amount of time on requirements specification and they have more contact with the users. This also shortens the time needed in later stages. Even if in the beginning they used more time, the time needed in the whole project is shorter. -- -- In my research there is evidence suggesting that if the beginning of a user-centred project has gone astray so that the world of the user has not been understood correctly and the kind of product users want has been misunderstood, not even an enormous amount of usability testing has been enough to fix the problem. (I1.)

User requirements should be defined to design products that users want to use so that problems can be avoided later in design. All in all, defining requirements at the beginning of the project increases the probability of success and user satisfaction, and decreases the time, effort and cost involved in correcting mistakes later on.

2.4 Methods for Defining User Requirements

There are many ways to find out users requirements and the best way is to use several of them. Which methods should be used depends on the user group and the design project and design stage of the product. The following chapters each introduce one method or part of defining user requirements and discuss when and how the methods should be used.

2.4.1 Conceptual Design

Conceptual design is a design stage in which the product is observed at a very abstract level. “The concept stage is where it is considered whether the product is really needed or what people would use it for (I2).” The goal of conceptual design is to gather data on the potential users, what do they want to do, their likes and dislikes and other feelings, and to create several product concepts out of it. Once the concepts have been created, they are evaluated and combined to form even better concepts. “It (the conceptual design) starts by examining the needs of the user. -- -- As many as possible concepts are designed to the scenario level. Then possibly some light prototypes and testing are done. The concepts are evaluated with users and customers. Finally, the concepts are combined and refined. (I2)”

Conceptual design can be done in several different ways. The usual way is to start with a *focus group* to find out what should be the focus of the concept design. Also *diaries* can be used to gather more long-term data. After that come interviews, observing and probably contextual inquiry. The design team brainstorms several concepts out of the gathered user information and then the concepts are tested, evaluated, combined and refined. This is the usual way of doing conceptual design, but as in all other forms of usability engineering, the methods to be used depend entirely on the user group and the context.

2.4.2 Usability Engineering in Product Design

When the conceptual design has been completed, product design starts. Whereas conceptual design tries to answer the question what users want to do, in product design stage it is essential to know how they want to do it? Users want to do things and perform tasks rather than use objects. The goal of defining requirements is to clarify what users really want.

The usual way to design usable systems is to start by interviewing users. Diaries and artefacts can be used for support. After that, users are observed in some usage situation. Instead of normal observing, contextual inquiry can also be used. All gathered information should give a fairly good idea of how users want to perform tasks. From that information user requirements are specified. Requirements should be concretised using scenarios. They should also be tested as soon as possible with users.

2.4.3 Scenarios

A scenario is a description of activity in a given situation. A scenario can be anything from a narrative text to video mock-ups. Scenarios are used to help envision user activity. Scenarios concretise user information from abstract data to an easier-to-understand form. In a way, scenarios are a very light form of prototype. Scenarios can and usually should also be tested with users. Scenarios are useful because they can be applied to all levels of design. They can concretise users' behaviour with the product. They can tell how data is moving through the organisation and applications. They can help to give an understanding almost anything in the design process or in the user's world. Scenarios are usable throughout the whole project. (See Carroll, 1995)

2.4.4 Interviewing

Interviewing is an effective way to get subjective user data. Interviewing is a less time-consuming way to obtain some user data. The data obtained is always subjective and depends on the interviewer, the interviewee and the interviewing situation. One extremely good point in interviewing people is that interviewing increases the ability of understanding the user's work and hence makes it easier later on to observe users while they work. Interviewing is also useful when observing is not possible. It could also be useful to interview people who are not directly users of the product but are connected to it somehow, for example supervisors and managers.

There are several good ways to interview people. How the interview should be done depends as much on the interviewer as on the subject. One good way of interviewing is to do process analysis with the user. In process analysis the interviewer asks the user to walk him or her through the usual work process. Although users usually fail to mention actions, which they do automatically, process analysis gives a good general idea of users' work. Another good interviewing method is the ethnographic interview. In an ethnographic interview the interview questions are refined after each interview. The idea is that the interviewer's knowledge and understanding of the subject grow in each interview and it helps the interviewer to improve the focus of later interviews or observations. A good method in interviewing is also the semi-structural interview. It is an interview where only basic guidelines and focus points are decided in advance. It is flexible and useful especially when the interviewer is not yet familiar with the subject. (See Hackos 1998, p. 135 – 138.)

Interviewing is equally useful at the conceptual stage and the product design stage. Basically the same interviewing methods can be used. The only differences are in focus and questions.

2.4.5 Observing

Observing the user is one of the most efficient ways to gather relatively objective user data. However, the problem with observing is that one has to be observing at the right place and at right time. In the worst case, it would require continuous observing for a long time to gather enough useful data. Observing the user will provide useful data if the observer knows what (s)he is looking for. Efficient observing requires lots of experience and preparation for the situation. Observing is most beneficial when combined with other techniques.

Observing is equally useful in both conceptual design and product design. Observing can reveal what users need and want, or how users tend to approach problems and products. The difference is in the focus of the observing. In conceptual design it is important to observe what users do, why they do it, and what they like or dislike. In product design it is important to observe how the users perform tasks, why they perform them that way, how much they use technological products, what products they use and how they use them.

2.4.6 Contextual Inquiry

Contextual inquiry is a kind of combination of interviewing and observing. The main idea in contextual inquiry is to go where the user works, observe the work and talk to him or her about the work (Beyer 1998, p. 41). The goal is to get the user to talk about his or her work naturally. According to Hackos, contextual inquiry is a philosophy as well as a technique. She gives several good tips on how to conduct a contextual inquiry. Among them she urges the inquirer to talk and ask about the task when user is performing it or, if that is not possible, right after the task. (Hackos 1998, s 130-131.)

Contextual inquiry also provides information useful for both conceptual design and product design. As in observing, the difference is mainly on the focus of the research. In contextual inquiry the interviewer must be even more aware of the differences in focus because the interviewer is an active participant.

2.4.7 Focus Group

In a focus group, users tell stories of real situations on a given topic. Users can also be asked to make collages or some other visualisation of their stories and job beforehand. Stories tell about users' likes and dislikes. They also give some idea of what users do. The goal of a focus group is to find out what the users do and feel and hence to map their behaviour. The focus group is a useful method at the beginning of the conceptual stage. The result is several stories that tell what users do. The information is important in brainstorming the concept ideas. (See Hackos 1998, p. 140 – 142)

2.4.8 Diary

When long-interval user data is needed, a useful way to obtain it is to have the users keep diaries. A diary gives a general idea of what the user does in the long run. It gives information of the user's routines, habits, where (s)he moves and what (s)he does there. It is generally a good idea to interview the user after (s)he has returned the diary.

A diary does not have to be a written record. It can be any kind of record that the user updates when (s)he does something. A diary can, for example, be a map where the user draws where (s)he has been and writes briefly what (s)he did there.

Diaries give useful long-term data. They can be used both at the conceptual design stage and at the product design stage. What diaries are used for depends a lot on the design stage though. In the conceptual stage, diaries are used to gather information on long-term user behaviour. In a way it is a substitute for observing when long-term observing is not possible. In the product design stage, diaries tell what the users do and where they move, which helps the designer to observe the right places and the right things and therefore supports interviews.

2.4.9 Prototyping

Prototyping is a useful way of evaluating design decisions. They can also be used to visualise and concretise requirements and specifications. In the conceptual design stage, prototypes do not try to simulate the product, but the experience and feeling of using the product. In the product design stage, prototypes simulate how users use the product and how they perform tasks. Prototypes can be anything from a navigation map to an interactive prototype or a physical mock-up. What kind of prototypes should be used depends on the design state, design decisions, the users, and what information is being sought.

Lots of information on good and bad points of the product can be gathered using simple paper prototypes. Almost all prototypes should be tested with a few users to see what users think of the design so far. Prototyping should be done as soon as the design decisions are made. Finding mistakes early save lots of time and work later.

2.4.10 Artefact Walk-Through

Artefact walk-through is a method whereby artefacts gathered from users are read through and analysed with a user. Artefacts are usually paper sheets, screenshots and booklets that the user uses or produces in the process of accomplishing a task. Typical artefacts are manuals, notes, input forms, output sheets, etc. These artefacts are then read through and usually analysed with the user. Understanding how artefacts are used by individuals, members of larger organisations or even among organisations helps to understand the users' needs beyond the immediate use of the product and helps to make a product which fits in better with the users' environment. (Hackos 1998, s 138 – 140.)

In the conceptual design stage, artefact walk-through is used to understand the wider environment of the concept. In the product design stage, artefacts are useful to point out where previous designs are not sufficiently usable. Users may have done hand-made artefacts to help them where the product is hard to use. Using artefact walk-through to analyse these artefacts helps to understand why the problems exist.

2.5 Using the Methods

Defining user requirements is a long process. It starts with the conceptual stage, which usually begins from some assignment. It is probably a market opportunity created by new technology or identified by marketing research. The conceptual design starts with focus groups, keeping diaries and interviewing users. After that, users are usually observed in certain situations. Using this data, the design team develops a few concepts.

One of the concepts is selected for further development. The designers interview and observe users again. This time they focus on certain tasks and how the users like to perform them. The design team defines user requirements and tests them with prototypes and scenarios.

2.6 Conclusions

There are several good methods to define user requirements. This chapter introduced the most common and useful ones. None of them is sufficient to define all the requirements correctly. However, an experienced usability design team can define requirements well by using these methods together.

Several times in product design there are decisions to be made. If user requirements are defined early and are kept in mind, it should be possible to make decisions that are more beneficial to the user. Designing a product by user-centred methods from the very beginning is important for the usability and usefulness of the product. It also reduces the cost and effort needed later in the work. Above all, though, it increases the users' satisfaction with the product. One thing that should be kept in mind throughout the whole project is that in the end it is the users who decide whether to use the product not.

3 Methods for evaluating small devices and their services

Juha Kolari and Eija Kaasinen
VTT Information Technology

3.1 Introduction

Unlike PC interfaces and web usability, the use and evaluation of small screen devices is still relatively new and the evaluation methods are being established. This chapter highlights usability issues, which will arise when evaluating applications or services for hand-held devices. It also reports experiences and provides practical advice for carrying out evaluations of small screen interfaces in laboratory tests or field trials. The document relies heavily on the experiences of VTT's WAP research.

We do not present a review of evaluation methods, but rather an introduction to issues that need to be addressed when evaluating applications and services on small screen devices. The physical device itself plays a much larger role than in PC software. The passages take a look at the usability issues through expert evaluation, laboratory trials and field trials.

3.2 Dimensions of usability in the laboratory

In PDAs, the variety of functions is wider than in smaller devices such as cell phones. A wider range of input methods become available. At the larger end of the scale, portable computers already resemble desktop PCs. In general, PDA and palm device usability can be approached in the same way as, say, WAP phones and services. All of these devices are mobile by nature and have limited input methods. At the same time, established guidelines for web or PC usability can be adapted remarkably well for small devices as well.

The limitations that laboratory tests have are especially apparent with mobile devices. The devices are meant to be carried around as a part of the user's personal or professional life. A short test in a laboratory can only reveal observations about the device's ease of use and initial attractiveness to the user. It cannot tell much about usefulness in real life, starting from how the size of the device would suit the user and ending with true integration into the user's everyday life. This is especially true with navigation products and services that are intended for mobile use and that often do not even work indoors (e.g., GPS).

3.2.1 Device vs. application / service

The physical PC device with its keyboard is rarely the main target in usability tests and neither is the operating system. In contrast, the physical hand-held device itself will be under scrutiny even if the evaluation is targeted at a specific application or service. Issues such as size, portability and input methods will be commented on. It cannot be expected that the users themselves will be able to separate problems with the device from those caused by the application/service. All comments should be taken into account, since they affect how the users feel about the application.

The user most likely will not see the software "as is" but filtered through how accessible to the user the device itself is. The users' opinions about the device and service are quite inseparable. Evaluators need to decide carefully where to target the evaluation and what other aspects of the device affect that goal. How to evaluate a single application without evaluating the complete device? How to evaluate the device without getting lost in details?

Navigation devices and services are available as embedded systems (e.g., Garmin, Benefon Esc!), applications on standard platforms (e.g., Magellan) and as a generic browser-based service working on several different platforms (e.g., Sonera Pointer on WAP devices). In embedded systems everything affects everything and these systems should be tested as a whole. However, functions that do not affect navigation can be left out of the test. When evaluating applications on standard platforms, it is important to know the user interface philosophy of the platform. It is also important to test the application both with users who are already familiar with the platform and users who are not. Generic browser-based services should be evaluated on each major device platform of the target market. The service can be designed according to generic design guidelines if it is not possible to test the application on different platforms. However, the most usable services can be achieved by utilising the unique features of different phones and browsers.

3.2.2 Size

In laboratory tests, the size of the device can be evaluated more on an aesthetic scale than from the perspective of usefulness. PDAs are usually meant to be used with two hands, which sets limits on when they can be used. These restrictions may not become apparent in short laboratory tests. Still, the users may have an idea about which size/features/display combination could be best for them, especially if they have previous experience about the matter. In phone/PDA combinations, individual users may prefer either separate devices or one device for everything.

3.2.3 Software

One can set out to evaluate the look and feel of the **complete interface** and the interplay between different applications. Users have a mental model of how things should work. This model is usually adopted from the operating system they know, most likely Windows. The model can also come from somewhere else: maps and paper calendars, for example.

In the evaluation of the whole philosophy of a device, one should be careful about not getting entangled in usability issues of separate applications, thus snowballing the user test and the covered topics. One would not usually perform a usability test on, say, the complete Windows. Despite their small size, systems on mobile devices are becoming more complicated and the topics that are to be covered should be defined carefully.

Different operating systems, such as Epoc and Windows CE, are used with portable devices. The look of the UI can vary even within systems according to device limitations, such as screen size or display. The device manufacturer may add its own special features on top of the operating system. One should assess whether there are assumptions that the user makes about the appearance of the system and whether these assumptions help or hinder the use of the device. Innovative design solutions, which may be helpful in the long run, may be confusing for users when they first try the device.

Another option is to evaluate a **specific application** within the device. This is more straightforward than the evaluation of the complete interface, as the whole way of interacting with the device is not in question. The limits of the evaluation are easier to set. They may not be as easy to keep, however, as was suggested earlier. The device itself may attract a lot of comments that need to be reported if they affect the use of the software. Experienced users of a certain device may be able to separate device features from application features more easily than inexperienced users. Interviewing only experienced users may, however, give a more optimistic view of usability. This is why both experienced and inexperienced users should participate.

There can also be **other conventions**, outside the PC, that the user may have that the software does not match. Especially applications such as map or notepad should support "real-life use" of regular paper maps, especially if hand-held devices are marketed as their replacements. People's use of maps, calendar or making notes vary. It should be noted whether these conventions are supported in the software, not only on a general level, but also for each user.

Questions to cover these issues can be, for instance, the following: *"I see you didn't get quite where you assumed when you chose that option, what did you assume it would do? Why did assume that? Do you use it that way somewhere else?"*

3.2.4 Input method

Along with software, the input method of the device will affect how the user perceives the system. The user will bring to the device his/her assumptions about how certain types of systems should work.

When developing applications, the balance between different input methods should be studied. Once the user is settled with one type of input method, (s)he should not need to constantly switch between keyboard, pen and application buttons. Consistency and a focused approach are needed.

Keyboard

Several PDA devices include a keyboard. Users can comment on the feel, size and general ergonomics of the keyboard. However, usually these issues cannot be changed when designing software. These user comments should be taken into account, however, since ergonomics can have an effect on how the user perceives the software.

In addition to ergonomics, there are at least two issues that are involved with using a keyboard on a PDA device. These are keyboard **conventions from the PC** and **user experience** in using the keyboard for on-screen navigation. Whether or not the software supports the conventions the user expects from a keyboard can have an affect on how the usability of the device/software is perceived.

Conventions from the PC are easily transferred to another device with a similar input method. The keyboard is usually a smaller version of a regular Qwerty keyboard. Thus, the keyboard is likely to attract conventions of use from the PC world.

User experience in using the keyboard can vary greatly. One can use the PC without learning any shortcuts and navigating completely by mouse. Other users try to minimise mouse use by learning to use several shortcuts, preferring some to others.

Questions to cover these issues can be the following: *"How would you do this on a PC? Do you use Tab a lot or some other buttons? What would be the natural way to do this task for you?"*

Control buttons

In addition to the main method of input, the keyboard or a touch screen, the device is often equipped with control buttons for other actions (Figure 1). The buttons can be *physical command buttons*, as in PalmPilot or Nokia Communicator, or a *graphic toolbar* in the display, as in Psion devices. It should be determined what the balance between the buttons and other input methods is. Are the buttons there for on-screen navigation, for most frequently used actions or only for major, non-repetitive actions, such as power on/off? Does the software and its use reflect the intended purpose?

Figure 1.Control buttons for main actions.

If the buttons are programmable and meant for the most frequently used options, then that is exactly what they should house. The actions that the users will need to access should be assessed carefully before deciding on the control button functions. For instance, the smaller the screen, the greater the need to design sequential actions on consecutive screens. This creates more need for the user to retrace his/her steps. Hence, an easy way to get back should always be present.

Questions to cover these issues can be, for instance, the following: *"If you could choose what this button does, what option would you put there? You seemed to be looking for the right option from another place. Is this option in the right place for you, can you find it there? "*

Direct manipulation device

Using a touch screen and a stylus pen is a more direct way to interact with the system. The pen can be used as an amalgamate of the mouse, cursor keys and Tab all in one. Conventions from mouse use usually carry well to touch screen, e.g. double clicks / double taps, selecting text. Pointing feels natural for most users. The stylus can also be used for purposes that may be completely new for the user, such as handwriting recognition. One should evaluate whether using the pen is the best approach for the user in question. The user may or may not feel comfortable with the pen and this can have an affect on how (s)he perceives the whole device.

One should also realise that special characters are often needed for handwriting recognition. If the user has not had any experience with them before the test, it may have a great effect on how well the user can perform any tasks involving writing.

Questions to cover these issues can be the following: *"What feels different when you use this pen, as opposed touching buttons? Which do you prefer?"*

3.3 Expert evaluation

The practical limits of evaluating small screen devices are not as apparent in expert evaluation as they are when users are involved. There is no need to record, film or observe use. Expert trials emphasise the usability of the service and are not an evaluation of the service's usefulness. Expert evaluation checks for usability problems and conformance to style guides and guidelines.

As in laboratory trials, small screen applications and devices can be evaluated with the **actual device** or with PC **emulation**. The pros and cons of these options can be found in the chapter about laboratory trials. Both options can be used in expert evaluation.

Expert evaluations rely on general usability guidelines as well as guidelines written for a particular device platform. In the evaluation, the conformance of the application/service to existing style guides or guidelines for the device platform is controlled. Generally accepted guidelines concerning small screens are being established even as in-house style guides, and guidelines for individual devices are beginning to emerge. General principles will no doubt soon follow from these separate guidelines.

3.4 Laboratory trials

Laboratory trials can address more issues than expert evaluation. In addition to usability, actual users can voice their opinion on how they cope with the interface and how they feel the device could be useful in their lives. However, quite a few small screen devices are mobile by nature, which sets more restrictions on laboratory tests than with PC applications. The true usefulness of mobile devices cannot be established in laboratory conditions. This restriction is especially apparent in mobile navigation.

On the other hand, laboratory trials offer an excellent opportunity to observe the usability of services and applications that are designed for small screens. In the field, seeing what is happening on a small screen is difficult. With video equipment and monitoring devices in the laboratory, several evaluators can observe the use.

3.4.1 Devices

Small screen applications and devices can be evaluated with the **actual device** or with PC **emulation**. The PC emulation can be augmented with a **touch screen**.

The **actual device** should always be preferred to emulation. The interaction with actual device is difficult to simulate with a PC, which the user obviously cannot hold in his/her hand and press the buttons. Using the actual device allows greater mobility as well, although this is restricted if a video is being recorded of the interaction.

PC emulation can be chosen for a few reasons: The device itself is not available or the application is not yet available for the actual device. Using a PC also allows screen capture and possibly logging the user's actions. Accurate videotaping of the device is often better than screen capturing. Logging user actions and response times can be easier with emulation. Also, using emulation or even a paper prototype allows evaluation to take place early in the design process, which is a recommended practice.

Figure 2. Touch screen being used with a PC emulation of a mobile phone.

A **touch screen** can be used with PC emulation if one wants to avoid using a mouse as an input device, since it most likely won't be available for the device. See Figure 2. A touch screen has its complications, however. The users may not be familiar with using it and may have trouble hitting the correct target on the screen. The screen needs to be calibrated for the users as well.

Especially typing with a cell phone is difficult for users, because it requires quick, multiple key presses. In the real device the finger is not lifted from the button to make the presses, but a touch screen cannot register different key presses unless the user lifts the finger from the screen. With a mouse this can be achieved more easily and the haptic feedback from the mouse resembles a real button as well. Pressing the buttons on a cell phone emulator with mouse and cursor is a better alternative than a touch screen.

If, however, the actual device has a touch screen and a pointing device itself, using a touch screen with PC emulation can be useful. In other cases, using a mouse may be preferable.

3.4.2 Recording the laboratory trial

The strength of the laboratory trials lies in the equipment that can be used to observe the use. Users have said that they do not mind an observer peeking over their shoulder and thus not seeing the person who is asking him/her questions. The position can be quite awkward for the interviewer, though, and the view of the screen is not ideal. The situation worsens if two observers want to see what is happening on the screen.

Video. The following example is from the laboratory evaluation of a WAP service. The tests were conducted at the usability laboratory of VTT Information Technology in

Tampere. Two experts administered the test: one interviewed the user, the other mainly observed from a monitor and through a one-way mirror in the adjacent monitoring room.

The test was videotaped with two cameras. The subject sat in a chair and a video camera shot the screen of the cell phone in his hand. The second video camera shot a general view of the subject. The general view was inserted into the main footage of the phone. The images were relayed to a TV set next to the interviewer (Figure 3). The footage was also relayed to the observer in the monitoring room.

Smaller hand-held devices can also be filmed with relative ease and accuracy if the user is sitting down. Users do not tend to move the device too much as long as they are comfortable.

Audio can be recorded with whatever equipment the interviewer is comfortable with, with one exception: An open cell phone connection, for instance for using a WAP-service, causes interference with minidisk players if the recorder is located near the phone. The problem is not as great in the laboratory as it is in the field, because the recorder can be placed farther away from the user, rather than being carried on him/her.

Figure 3. Set-up of the laboratory trial of a WAP service.

3.4.3 PDAs and portable computers

The technical aspects of the evaluation set-up do not differ much for larger hand-held devices. A video camera and monitor are usually needed for a better view of the device and for recording the view from the device screen. There can be some unexpected problems, however. One results from the **larger display**, the other from the **position the device is used in**.

Quite surprisingly, **larger displays** seem to pose a problem rather than a solution, as far as shooting the display is concerned. If one wants to view the whole screen + possible keyboard, the camera needs to zoom out more than with devices with a very small screen. This results in a smaller and more blurred view of the screen. The display properties have an effect on how good a picture is captured on video. With monochrome displays, the backlight should be on. With colour displays, contrast and brightness should be adjusted for best effect.

Hand-held devices are used in **different positions**, depending on the device, the task and the user. Phones can be used in a chair, sitting back, as in Figure 3, which allows for easy camera angles. PDAs with keyboards can be used at a table with both hands on the keyboard. This makes it more difficult to shoot the display, as users tend to hunch over the device.

3.5 Field Trials

Field trials attempt to capture the use of a device or a service as true to real life as possible. Sometimes the use is more or less simulated, in which case the user can be instructed to use the service in a certain way or (s)he does not need to cover the real life costs of using the service.

The field trial can be either short-term or long-term. A **short-term trial** resembles a laboratory test, because it usually takes place within one session. The expert can observe the use all the time and can ask appropriate questions during use. The session is classified as a field trial, because it takes place "in the field", i.e. in the actual context the service is meant to be used in. This is often the way to evaluate navigation services, as indoor laboratory trials do not yield much information about navigation in the field.

In a **long-term field trial** the user uses the device/service from several days to months as (s)he would use it in real life. This means that the actual use can only be observed during interviews. A long-term trial may also include interviews with observation that resemble the short-term trial. Usage information can be gathered on log files for later analysis.

In a field trial, the means of observing use are more limited than in laboratory trials. The means of assessing the context of use are better and observations of the context should be conveyed in reporting. Good use should be made of the interviews by taking photos of how the device is being used.

In both long-term and short-termfield trials it should be assured that the device and application to be tested is reliable enough. Technical problems may severely affect the usage and even make the evaluation fruitless. Having a back-up device in a short-term field trial is a good way to reckon with technical problems and to ensure that the test can be finished properly. In a long-term field trial it should also be assured that the users have proper technical support during the trial. If you have doubts about reliability, it might be a good idea to start with a short-term test.

Figure 4. Taking photos of the context of use is recommended.

Especially in short-term field trials the time of the trial can affect the test. The weather is one remarkableaspect that can disturb the test situation – both the test user and the test device. Changes to the evaluation plan or even postponing the evaluation session may be necessary due to cold, hot, windy or rainy weather.

3.5.1 Users

The background of the users should be tested for experience of the device that applications are tested on. It appears that experience in using computers and the Internet does not help as much in the use of small screen devices (such as WAP phones) as experience in using the device itself.

If the users are given devices especially for testing a service during the field trial, one should take into account that the main motivation for some users may be in having the device and not using the service. Especially younger users can be interested in having a "cool" device.

When devices and services are evaluated outdoors and with mobile user, there are always risks for accidents and device breakdowns. The responsibilities of the evaluators and the end users should be clearly defined on a written contract. The contract should also define the rights of the evaluators to use the audiovisual material recorded during the tests, either for research purposes only or also for publications of the research.

3.5.2 Devices

Field trials offer information that cannot be gathered in laboratory trials. In a field trial, the use of actual devices is critical, although sometimes the use of portable equivalents of the actual device can be considered. Because small devices tend to be designed for mobile use, the use of something more cumbersome than the actual device undermines the purpose of a field trial. True mobile use in a long-term field trial offers a lot of information about how the device/application integrated into the users' daily lives.

3.5.3 Recording the field trial

In a field trial, parts of the use can be observed and recorded by the evaluators. In a short-term field trial the session can be a short walk in the location where the service is intended to be used, in which case all of the use can be observed.

Taking notes during mobile use, e.g. walking, can be difficult. It is recommended to concentrate on the interview and to go through the recording immediately afterwards.

Photographs of typical usage situations are very useful and illustrative in real contexts of use. Filming the field trial use of a mobile, small screen device is usually difficult without interfering with its use. Short video recordings are recommended in short-term field trials in situations where users have complicated usability problems. In this way the situation can be later analysed in details.

Audio recording is most valuable tool to document the field test**.** As discussed in connection with laboratory trials, a cell phone causes interference in the sound of the minidisk recording. It can be very distracting if the user has a carry-on recorder and walks with the phone near the recorder or lifts the phone near it, for instance to enter information. The user can be asked to put the recorder in the front left jacket pocket. If the user carries the phone in his left hand, then the recorder should be in the right jacket pocket.

The low microphone sensitivity setting is enough if the user carries a microphone. Also, traffic noise is less distracting on the recording if the microphone is less sensitive. The users naturally tend to raise their voices if the environment is noisy. However, one should take into account that when you are listening to the tape afterwards you often cannot hear what the interviewer is saying.

Log files store reliable information on the amount and type of usage. Log files and the software needed to analyse the files should be designed in parallel with the application itself. An on-line analysis tool makes it possible to analyse the usage in parallel with the ongoing trial.

3.6 Conclusions

With hand-held devices, the device itself will be under scrutiny even if the evaluation is targeted at specific software. Issues such as size, portability and input methods will be commented on. Such comments should be taken into account, since they affect how the users feel about the application. It cannot be expected that the users themselves will be able to separate problems caused by the device from those caused by the application/service. Evaluators need to decide carefully where to target the evaluation and what other aspects of the device affect that goal.

Along with software, the input method of the device will affect how the user perceives the system. The user will bring to the device his/her assumptions about how certain types of systems should work. They also have mental models and conventions of use that they have learned, usually with the PC. When different hand-held devices become more widespread, users may start to move away from the PC model.

When developing applications, the balance between different input methods should be studied. The user should not be expected to switch between different methods only because the function did not fit the screen.

Especially in field and laboratory trials, records of previous experiences help in setting up future trials. The report should include experiences and examples of how the test set-up could be improved in similar future tests: What choices were successful and why, what arrangements could be improved and how. The topics can include the selection of users, equipment, the software, log file, location, interview, forms, etc. Experiences of the test set-up should be reported for future use.

4 Methods for evaluating in-vehicle systems

Juha Luoma
VTT Building and Transport

4.1 Introduction

The evaluation of in-vehicle systems includes two points of view. Firstly, manufacturers and service providers are interested in the usability of the system as such. In general, this aspect of usability can be evaluated like other services (e.g. heuristic evaluation) and it is more thoroughly discussed elsewhere (Kaasinen et al., 2001). However, this chapter briefly presents evaluations such as user acceptance that are typically carried out when evaluating in-vehicle systems.

Secondly, a special problem with in-vehicle information and communication systems concerns the *interaction* between the driver's primary task (driving) and his/her secondary task (use of in-vehicle system), and the potential safety problems caused by this interaction. Specifically, the European Commission (1998) suggests that the system should be designed to support the driver and should not give rise to potentially hazardous behaviour by the driver or other road users. If an information and communication system is intended for use while driving, the system should be designed in such a way that the allocation of driver attention to the system displays or controls remains compatible with the attentional demand of the driving situation. These recommendations lead us to the questions of attentional demand of a system and workload problems while driving.

There are three types of measurements to examine these issues, namely subjective reports and assessments, performance measures and physiological measures. An overview of most measures will be given, although some of the measures will receive more attention than others. Finally, the use of these measures are discussed in terms of the design and procedure used in the evaluation.

4.2 Subjective reports

4.2.1 Workload measures

Subjective reports (or verbal reports or self-report measures) have always been very appealing, because there is no simple way to provide an objective judgement with respect to experienced mental load. On the other hand, one could argue that the source of the resource demands is hard to introspectively diagnose within a dimensional framework. (De Waard, 1996).

Mental load is usually measured by using one or several rating scales, such as the NASA Task Load Index, TLX (Wickens, 1992), which assesses work load on each of five of 7-point scales (mental demand, physical demand, temporal demand, performance, effort and frustration level). While the original TLX requires a two-pass process with paired comparisons, a modification called the Raw Task Load Index (RTLX) does not require task paired comparison weights. The RTLX is a simple average of the six TLX scales. In addition, the RTLX has been used in various modifications that are designed for car driving (which is not the case with the original TLX). Other frequently used rating scales are the Subjective Workload Assessment Technique, SWAT (Wickens, 1992) and the Modified Cooper-Harper scale, MCH (Wierwille and Casali, 1983).

Self-report scales have several advantages, the major one perhaps being their high face validity. In addition, the ease of application and low costs can be mentioned. The limitations of self-report measures include a possible confusion of mental and physical load in rating, i.e. the operator's inability to distinguish external demands from actual effort or workload experienced. Also mentioned are limitations in the operator's ability to introspect and rate expenditure correctly. (De Waard, 1996).

4.2.2 Other subjective measures

In addition to the workload assessments, evaluations of traffic-related services typically include the following assessments (Katteler et al., 1999):

- User acceptance which is a prerequisite for operational impacts: aiming to estimate users' attitude to and perception of the service (frequency of use, perceived benefits and disadvantages, willingness to pay).
- User's response: the way and the degree the service directly affects users' behaviour. The user response can be in terms of changed route choice, change of travel mode, saved travel time or change in individually travelled distance. At an aggregate level, this provides information on the gain in traffic efficiency.
- Other impacts: the measurement or estimation of the impacts (effects) of the service, particularly those on safety and transport efficiency. These include (a) impacts in terms of driver comfort (e.g. reduction of uncertainty), (b) improved continuity of information in time and in space (service provision without interruptions), (c) a higher degree of being informed and (d) impacts on safety (e.g. increased use of secondary roads, which means decreased safety or timely speed reduction alerted by the information provided).
- Usability of the human-machine interface (HMI) and safety aspects related to this HMI (e.g. easy identification and understanding of the information; detailed consequences of distraction of attention).

In addition, the road authorities are frequently interested in socio-economic evaluation: this evaluation area aims to estimate the social gains or losses as a result of implementing a service in comparison with the existing situation.

4.3 Performance measures

4.3.1 Primary-task measures

Primary-task measures are typically speed or accuracy measures. Evaluations of in-vehicle systems may involve the recording of performance duration with a given task. A slow performance means that there is less time for the use of other controls and perhaps less mental capacity for the primary tasks. Another measure can be the frequency of errors or halting use of the system. Such behaviour can be measured using a video camera.

The recording of eye movements is related to primary-task performance (most tasks are of a visual nature). Visual-search strategy, or the selective attention to relevant visual stimuli, has been shown to be indicative of information needs. The eye-scanning patterns in terms of frequency of fixations and fixation duration are typically recorded. The total dwell time or off-road time per task connected to the in-vehicle system may be the most frequently used measure of visual load.

The main problem with eye point-of-regard analysis is that eye fixations always `fill up' the total time. This is a problem especially in low to moderate workload situations, in which not all fixations are relevant and required for task performance. Moreover, the sensitivity of measures of eye-fixation will be restricted to visual workload, and the measure can be considered diagnostic in that respect. (De Waard, 1996).

Another problem related to eye movement analysis is the difference between looking and perceiving. A fixation does not necessarily imply perception, and perception does not necessarily imply a fixation. (Luoma, 1991).

Eye fixations can be measured using a video camera, by recording cornea reflection superimposed on a video image of the visual field, or by the recording the ElectroOculoGram (EOG). Perhaps the most frequently used technique uses a video camera in the dashboard, which can be either hidden (e.g. Penttinen et al., 2000) or visible (e.g. Wierwille et al., 1988) to the driver, although the techniques both suffer from labour-intense and time-consuming data analysis. The cornea reflection technique is accurate in point-of-regard evaluation, as long as the equipment is calibrated regularly, i.e. every 15 minutes or so. The EOG technique has the disadvantage that an accurate foveal point-of-regard is hard to assess. (De Waard, 1996).

4.3.2 Secondary-task measures

When another task is added to the primary task, secondary-task measures can be taken. According to the multiple-resource theory (Wickens, 1992) the largest sensitivity in secondary-task measures is achieved if the overlap in resources used is high. In other words, in order to perform the secondary task, spare capacity of the same resource should be required. Time sharing is expected to be less efficient if the same resources are used. This large overlap in resources used is at the same time a threat to undisturbed primary-task performance because primary-task intrusion is greatest if two tasks that use the same resources have to be time-shared. The choice for a secondary task is more difficult in tasks approaching everyday performance. Car driving, for instance, is to a large extent automated and mainly a visual task. The value of a secondary auditory digit-addition task is therefore not completely distinct. It is possible that performance on the latter task reflects central resource use. However, the extent to which performance of the primary task makes use of central resources is not clear in advance. The use of secondary tasks in applied environments is more complex than in laboratory experiments, and for this reason caution is required.

Examples of secondary tasks include choice reaction-time tasks (e.g. Alm and Nilsson, 1995; Hancock, 1999), memory-search tasks and mental arithmetic (see Wickens, 1992, for overview). Furthermore, several measures of driving behaviour have been used as indirect indicators of the use of in-vehicle systems (although not explicitly considered as secondary-task measures). The most frequently used measures include speed behaviour (e.g. Verwey, 1996), headway, lateral position, steering wheel frequency (e.g. Alm and Nilsson, 1995).

4.3.3 Expert evaluations

Driver performance and especially safety-related performance while using an in-vehicle system can also be assessed by an instructor. (S)he can either travel with the test driver (Verwey, 1996) or assess the driver's visual behaviour from video tapes (Penttinen et al., 2000). The items may involve behaviour at the control level of driving (braking and course keeping) and various situation-dependent types of behaviour at the manoeuvre level (i.e. indicating interactions with other road users) (e.g. Verwey, 1996). Each item is scored as being 'satisfactory', 'acceptable' or 'unsafe', for example. Verwey (1996) suggests that safety effects in urban traffic can be measured more reliably by subjective ratings than by objective measurements, because subjective estimates can take many changes of typical behavioural patterns into account.

4.4 Physiological measures

Different physiological measures have been found to be differentially sensitive to either global arousal or activation level (e.g. pupil diameter), or to be sensitive to specific stages in information processing (e.g. the evoked cortical brain potential). The advantage of physiological responses is that they do not require an overt response by the operator, and most cognitive tasks do not require overt behaviour. Moreover, most of the measures can be collected continuously, while measurement is nowadays relatively unobtrusive due to miniaturisation. The main disadvantages of physiological measures include the necessity for specialised equipment and technical expertise, and the critical signal-to-noise ratios. (De Waard, 1996).

Physiological measures have also been used in evaluations of in-vehicle systems. However, the detailed presentation of these measures is omitted as the measures are mostly under development and their use has been rather limited. An overview of physiological measures of workload related to in-vehicle systems is given by De Waard (1996).

4.5 Design and procedure of data collection

The above-listed measures are techniques of data collection that are useless if they can not be applied appropriately. More specifically, every evaluation has to include a proper selection of techniques which are used with a proper design, and every evaluation must be carried out with a proper procedure. Furthermore, these call for a theoretical and thorough understanding of driver behaviour. Although this presentation does not aim to be a text book on driver behaviour research, some practical and methodological factors are listed in order to highlight the question.

One of the most important factors affecting the nature of the evaluation is the implementation stage. There are at least two options (Katteler, 1999): (1) a test or demonstration stage with one prototype or a few prototypes, or a very limited service and (2) a full service under market conditions. Systems in a test stage have full control over the test drivers in terms of the selection and number of drivers (although evaluations frequently include only a few drivers, who do not represent the whole driver population). In contrast, for services operating in market conditions, the link to the buyer of the set is substantially looser and there is even the risk of having no link at all.

If the system is in the test stage, the evaluation can focus on performance measures with test drivers, such as recording performance duration, errors, eye movements, etc. This data can be complemented with a (extensive) subjective assessment made by the test drivers or an instructor. If possible, the evaluation should include a comparison of two or more systems, or there should be another type of base line for the evaluation. For example, the use of a system requiring visual attention can be compared with the use of conventional controls, such as a radio. The within-subjects designs are preferred because of their effectiveness.

In market conditions, the evaluations typically focus on subjective measures. Comparisons between different systems are difficult to conduct and data collection is less extensive compared to the test stage. In addition, the evaluation is less effective because it may be difficult to set any base line. However, the measures can be strengthened by evaluating driver acceptance and driver response at the trip level (instead of the person level). More specifically, the power of the findings is substantially higher if the behavioural response is expressed in terms of percentage of trips. For example, route change could be addressed more precisely then. Therefore, the use of a trip log is encouraged. It is acknowledged, however, that the use of a trip log would be more feasible in service test conditions than under service operation conditions, when control over drivers is substantially lower. An extra advantage of using a trip log is that it allows one to include a reference situation in the evaluation design more easily. (Katteler, 1999).

5 Human-Centred Design in a nutshell

Eija Kaasinen
VTT Information Technology

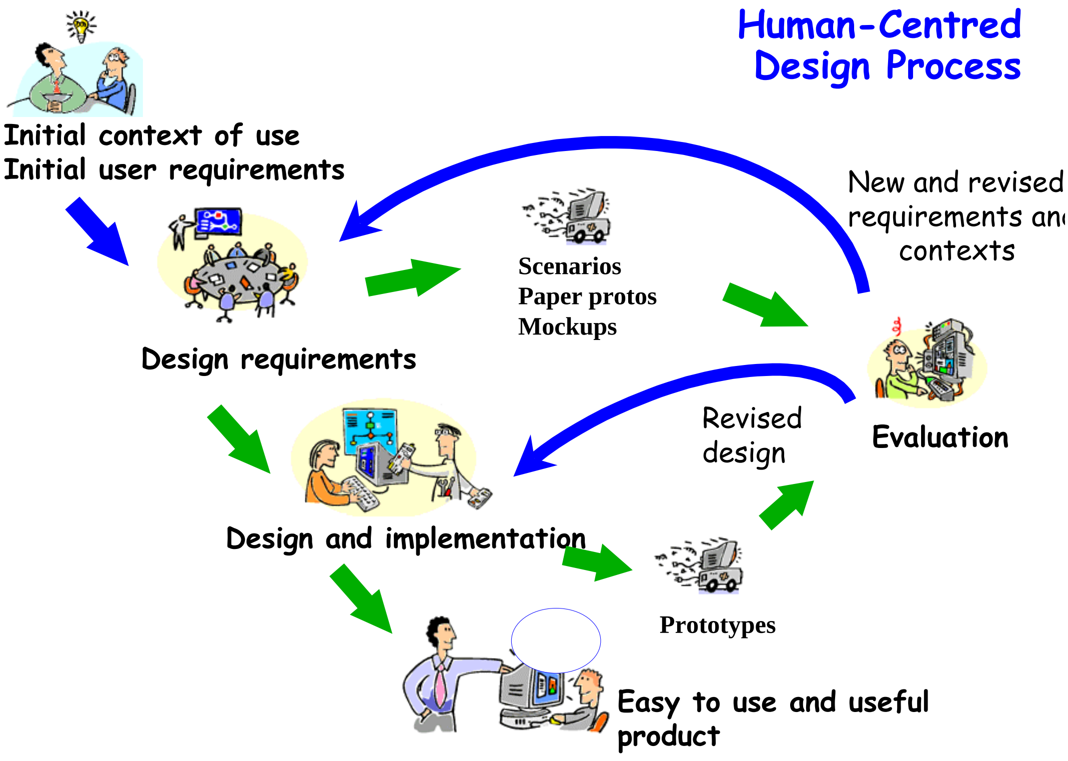


Figure 5 Human-Centred design process

KEN deliverable 1, Basics of Human-Centred Design gives an overview of the methods used in human-centred design. This part I of deliverable 3 complements deliverable 1 by describing usability methods specific to personal navigation products or mobile products in general. This chapter gives an overview of the human-centred design process. The elements of the process and the suggested methods are described in more details either in this report or in KEN deliverable 1.

Elements of human-centred design:

- defining contexts of use
- defining user requirements
- making design solutions
- evaluating the solutions with the users
- iteration based on user feedback

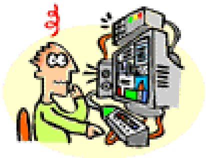
(ISO, 1999)

Basics of human-centred design

- Multidisciplinary design team (technical, application field, industrial design, usability, marketing, maintenance, etc. experts)
- Usability design is an integral part of the design process
- The results of the usability evaluations are analysed together
- The design team decides on necessary changes
- Design rationale tracks user requirements throughout the design
- The learning process also produces design guidelines, in-house or general

Challenges for human-centered design

- Consumers are a very heterogeneous group
- Contexts of use vary a lot
- The context of the mobile user may change during use
- Needs for personalised and adaptive features
- Product development processes are getting faster
- Little or no time to react to the results of the evaluation
- Need for thorough user studies before starting the development process

Initial contexts of use 	**Who will be using the forthcoming product?** **In what kinds of situations?** Experiences of previous products Literature Interviews Observation Scenario building Brainstorms
Initial user requirements 	**In what kinds of real-life tasks should the product support the users? How?** Literature Evaluation of previous products Interviews Observation Contextual inquiry Scenario evaluations Brainstorms Laws, guidelines, standards Definition of usability measures
Illustrating design decisions 	**And what does that mean in practise?** Scenarios Paper prototypes Wizard of Oz Mockups Prototypes
Evaluating 	**Does the design conform to the user requirements?** **Expert evaluation** Checking against usability heuristics or guidelines by usability experts Possible problems identified, corrections suggested **Pluralistic walkthrough** Designers, usability experts and users go through the design together **User evaluation in laboratory** Qualitative evaluation 5 - 8 users will do Can be done with early design illustrations Feedback on usability **Field evaluation** Can be qualitative or quantitative Laboratory evaluation in field environment or long-term evaluation Long-term evaluation requires a stable prototype Feedback both on usability and utility
Iteration based on user feedback 	**What is the impact of the evaluation results to the design?** Analysis of the results by the design team Design decisions new or revised user requirements/contexts of use, new design Maintaining design rationale

6 Conclusions

Identifying user requirements correctly and as early as possible is one of the key issues in human-centred design. We have described some useful methods to define user requirements in designing a product concept or in designing the product itself. Navigation devices are by nature intended for mobile use and that is why they are often handheld-sized. In this report we described some specific issues that should be taken into account when evaluating small handheld devices and services for them. When using in-vehicle devices, the user can devote only partial attention to the devices, his/her main attention being directed at driving. The same situation is typical for navigation devices, whether intended for drivers or pedestrians. We have described in this report some specific issues that should be taken into account when evaluating in-vehicle devices. These methods can be utilised in evaluating any kind of navigation devices.

We concluded this report with a description of the very basics of Human-Centred Design approach.

7 References

Alm, H. and Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. Accidents Analysis and Prevention, 27(5), 707-715.

Beyer, H. & Holtzblatt, K. (1998). Contextual Design. Defining Customer-Centered Systems. San Francisco: Morgan Kaufmann Publishers, Inc.

Carrol, J. M. (1995). Scenario-Based Design. Envisioning Work and Technology in System Development. New York: John Wiley & Sons, Inc.

De Waard, D. (1996). The measurement of drivers' mental workload. Ph.D. thesis, University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Centre.

European Commission. (1998). European statement of principles on human machine interface for in-vehicle information and communication systems.

Hackos, J. T. & Redish J. C. (1998). User and Task Analysis for Interface Design. New York: John Wiley & Sons, Inc.

Hancock, P.A. (1999). The effects of in-vehicle distraction on driver response during a crucial driving manoeuvre. Transportation Human Factors, 1(4), 295-309.

ISO 13407:1999. (1999) Human-Centred Design processes for interactive systems. International Standard. The International Organization for Standardization.

Kaasinen, E., Luoma, J., Penttinen, M., Petäkoski-Hult, T. and Södergård, R. (2001). Basics of human-centred design in personal navigation. KEN Project Deliverable 1. Tampere: VTT Information Technology.

Katteler, H., Escard, J.C., Vazquez, R.G., Graullera, D., Van Hemelrijck, K. and Luoma, J. (1999). Evaluation of RDS-TMC services: recommended actions and instruments (FORCE Document No. 400-W824.D02). Brussels: FORCE project.

Luoma, J. (1991). Perception of highway traffic signs: Interaction of eye movements, recalls and reactions. In, A.G. Gale, I.D. Brown, C.M. Haslegrave, I. Moorhead and S. Taylor (eds.) Vision in Vehicles-III. pp. 325-332. Amsterdam: North-Holland.

Penttinen, M., Luoma, J., Kulmala, R. and Nygård, M. (2000). Links between HMI indices and traffic safety: Initial analyses of driver behaviour while using an RDS-TMC receiver. Paper presented at the International Conference on Traffic and Traffic Psychology (ICTTP), Berne, Switzerland, September 4-7, 2000.

Verwey, W. B. (1996). Evaluating safety effects of in-vehicle information systems (IVIS). A field experiment with traffic congestion information systems (RDS-TMC) and preliminary guidelines for IVIS. TNO Human Factors Research Institute, TNO Report TM-96-CO68. Soesterberg: TNO.

Wickens, C. D. (1992). Engineering Psychology and Human Performance. Second Edition. New York: Harper-Collins.

Wierwille, W.W., Antin, J.F., Dingus, T.A. and Hulse, M.C. (1988). Visual attentional demand of an in-car navigation display system. In, A.G. Gale, M.H. Freeman, C.M. Haslegrave, P. Smith and S.P. Taylor (eds.), Vision in Vehicles-II pp. 307-316. Amsterdam: North-Holland.

Wierwille, W.W. & Casali, J.G. (1983). A validated rating scale for global mental workload measurement application. In, Proceedings of the Human Factors Society 27th Annual Meeting pp. 129-133. Santa Monica, CA: Human Factors Society.