Methods and techniques for field-based usability testing of mobile geo-applications

Ioannis Delikostidis
March, 2007

Methods and techniques for field-based usability testing of mobile geo-applications

by

Ioannis Delikostidis

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: (fill in the name of the specialisation)

Thesis Assessment Board

Chairman: Prof. Dr. M. J. Kraak
External examiner: Ms Dr. A. Dilo
Supervisor: Ms. Dr. J. E. Stoter
Second supervisor: Drs. E. L. Poppe

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS

Disclaimer

Abstract

In the recent few years, we have been witnessing a revolutionary change in the ways that geographic information can be utilized. The rapid growth of the World Wide Web, as a medium of acquisition and dissemination of geographic information, triggered a transition from stand-alone to distributed Geographic Information Systems. Concurrently, the evolution of wireless communications and positioning technologies together with the emerge of increasingly powerful and affordable mobile devices, such as the mobile phones and pocket PCs, led to the creation of the Mobile GI systems. Mobile GIS opened a whole new world of opportunities for the development of innovative and useful applications that can provide amongst others location-based information to mobile users. Vehicle and pedestrian navigation systems and mobile tourist guides are examples of mobile geo-applications, which often use interactive maps as interfaces.

Despite of the great potential that mobile geo-applications show, there are also several limitations in their use, related to the mobility of the users and the restrictions of the mobile devices, such as the small size and low resolution of the display and the lack of a proper keyboard, Additionally, the different contexts of use and the various user profiles and requirements make the improvement of usability of mobile geo-applications applications, a difficult task.

Usability testing is one of the evaluation methods that can be applied in order to assess the usability of a mobile geo-application in a specific context of use. However, the already established methodologies for usability testing for desktop environments are not always applicable to mobile geo-applications. There is a strong need for further, improved research methodologies that can be applied to different kinds of mobile geo-applications, and especially for field-based usability testing, which can uncover problems under real contexts of use.

The objective of this research is to build a methodology for field-based usability testing of mobile geo-applications. In order to do this, it first investigates the functionalities, limitations and special characteristics of mobile geo-applications. Then it studies the different methods and techniques for usability testing, and especially these that can be applied for field-based testing.

Based on these studies, a methodology for usability testing in the field is proposed, which could be used for the evaluation of UWSM2 project prototypes when they are implemented. UWSM2 goals and use cases are used in order to build this methodology which compares three different combinations of usability testing methods and techniques, on a selected geo-application similar to UWSM2 ones. Eighteen participants are selected to carry out these tests, which determine the most usable methodology. In order to support this research's methodology and allow the execution of the surveys by only one researcher, a unique technical solution is developed. The output of the surveys is used to further adjust and improve the selected methodology and the technical solutions that are used.

Acknowledgements

I would like to express my gratitude to:

My research supervisors, Ms. Dr. Jantier E. Stoter and Mr. Drs. Eddie J. Poppe for their continuous assistance, very valuable feedback and guidance

Mr. Dr. Corne P. J. M. van Elzakker, the motivator of this research and my initial supervisor, for enriching my knowledge and raising my interest towards usability testing

State Scholarships Foundation (I.K.Y.) and my employer in Greece for funding my studies

My parents Eleni and Savvas, my sister Sofia and my entire family for their infinite agape, encouragement and support during this research

Diana Kusumastuti, for giving me the extra strength needed to carry out this research, and continuing to beautify my days, even from thousands of kilometres away

My ITC colleagues, who voluntarily participated in the field-based surveys of my research, for their enthusiasm and priceless contribution

Areti, for her substantial support and advising during the early stages of this research

Rosemarijn Looije, for her kind help during the literature review stage of this research

Table of contents

List of figures

List of tables

1. Introduction

1.1. Background

As a tool for visualization of spatially referenced data, maps are meant to make the user comprehend the relationships between the spatial entities. Important information concerning distances between features, directions, size of areas, patterns of phenomena or groups of features as well as relations between them can be available to the user through maps (Kraak & Ormeling, 2003). The great value of maps in the process of communicating geographic data, providing awareness and synopsis to the users is undoubtedly recognized; identifying them as the most operative media available (Van Elzakker, 2004).

Maps as means of interconnection of the spatial data and the user are also an integral tool of the GIS, visualizing the spatial data in a meaningful and usable way. But who is actually this "user"? Although the distinction between the user and the producer of the maps is becoming vaguer as a consequence of the cartographic visualization's evolution, s/he is the one that use his/her cognitive and visual abilities in order to answer his/her geographic questions. The user can uses maps produced by others, or by themselves, with the help of GIS tools (Van Elzakker, 2004).

Analysis, querying and customization of geo-data can be efficiently done inside GIS while map representation can be dynamically changeable (Wiegand, 2006). Electronic map displays can represent much more information than paper maps, since the maps can be adjusted according to the specific user needs, using up-to-date geo-data. The GIS user can select features, modify their characteristics, observe the results of processing and change the way a geographically referenced space is being imaged. The capability of instant and accurate obtainment of information like the coordinates of cartographic features, the distances between them, and the calculation of area sizes, together with zooming, panning and special functions like buffering and spatial querying, help manipulate and explore the maps in an interactive way not possible with paper maps (Davis, 2003). However, paper maps are still used as they have some noticeable advantages over map displays; including the inexpensiveness of production, and the easiness of manipulation, folding and storing. Additionally, they don't need computer hardware, software and power sources in order to be used, thus offering great portability and immediate accessibility (Reilly et al., 2006).

Evolution of Internet and positioning technologies together with the availability and a price drop of mobile devices resulted in the creation of a new type of GI systems: the mobile GIS, belonging to the broader family of distributed GIS which also includes internet GIS. Mobile GIS brought a whole new variety of possibilities in the ways in which geo-data can be utilized and disseminated (Reichenbacher, 2001). We use the term "mobile geo-applications" in this research, referring to all these software applications that run on mobile devices, such as Personal Digital Assistants (PDAs), mobile phones or smartphones and embedded systems (Fraunholz & Jung, 2002) and utilize a geographic information element, such as map databases or/and a location specific element, such as white pages, together with information regarding the location of the mobile user. These applications can have a mobile or wireless data connection to a remote server, as mobile devices generally have this capability. (Peng & Tsou, 2003). Location Based Services (LBS), and telegeoinformatics (Karimi, 2004), including car navigation systems, location-based weather and traffic information, and mobile tourist guides are categories characterizing these applications (Figure 1.1).

Figure 1.1 Examples of mobile geo-applications: TomTom car/bicycle/pedestrian navigation system running in a mobile phone, some screenshots of the different pages of its interface, and Culture Around The Corner mobile LBS (Source: URL_1.1, URL_1.2, URL_1.3)

Field GIS is also a main category of mobile geo-applications, according to Shamsi (2005), including applications that are used by mobile workers in the field in order to confirm, add, manipulate and on-line update geo-data in remote geo-databases. ArcPad from ESRI (figure 1.2) is an example of field GIS, allowing access of databases, map representations, and Global Positioning System (GPS) integration (Ng'ang'a, 2005).

Figure 1.2 Field GIS applications: ArcPAd from ESRI (Source: URL_1.4)

Interactive maps are very often used as a type of interface in mobile geo-applications, and not only as a way to present geo-information like paper maps did in the past. There are also big differences compared to the use of map displays in desktop GI systems, since in mobile geo-applications the restrictions of the mobile devices, including the small screen, low resolution and color count, minor CPU power and restricted data input abilities combined with the advantage of mobility of the user, require different types of geospatial data handling and visualization. It has to be mentioned here that laptops are not included in the mobile devices, as their portability, a basic parameter determining mobile devices, is much lower than PDAs' and mobile/smart phones'.

Despite of the great potential that mobile geo-applications show, the mobility of the users and the restrictions of the mobile devices, the different contexts of use and the various user profiles and requirements together with the limitations of mobile devices / clients, make the improvement of usability of mobile geo-applications applications, a difficult task.

Usability testing, as part of a user-centered design and development, described in Chapter 3, is in the need of additional research and improvement (Rao et al., 2006). There is a strong need for further, improved research methodologies that can be applied to different kinds of mobile geo-applications. Usability testing of mobile applications, while it comprises a constriction in their implementation process, is at the same time a very important element for supporting the acceptability of new applications or services (Marcus & Gasperini, 2006, inside Streefkerk et al., 2006).

Mobile geo-applications strongly benefit from field evaluations in real contexts of use, as they are aimed at being used in real-world situations. While laboratory testing has also several benefits, such as the easiness of the process and the more controllable conditions, field-based testing can investigate user-application interaction issues that are often invisible in the laboratory's artificial environment (Isomiru, Kuuti & Väinämö, 2004). In this research, we conduct the field-based testing part of the prototypes evaluation. This is less investigated than laboratory testing.

The usability testing should efficiently detect the weak points of the implemented pre-production prototypes as well as the final products, giving hints for further improvements through user surveys (Fig. 1.3).

Figure 1.3 Usability testing for a mobile geo-application in the field

1.2. Problem definition

While a lot of research has been done recently concerning the development of mobile geo-applications, less attention has been given to the acceptability of the end users of these applications in particular contexts of use. The development of mobile applications or services is mostly supply-driven, as these applications are not designed with the limitations of mobile devices in mind (Reichenbacher, 2004). This leads to products with reduced usability, as they cannot efficiently address significant problems, such as the complexity of visualization, resulting in cognitive overload (Kjeldskof et al., 2005).

To assess the usability of map displays in mobile geo-applications in satisfying context-based usage requirements, little usability testing research has been executed and only a few usability testing surveys have been performed until now, mostly executed in the laboratory. The aim of these methodologies is the evaluation of prototypes and final products in the laboratory or / and in the field, but usually the role of the specific context of use is underestimated or not well defined.

1.3. Motivation for the research

As mobile geo-applications are emerging fast, and there is a strong interest for implementing new applications that can promote high end-user acceptability, it is important to develop proper usability testing frameworks that can improve the process of design, testing and evaluation. Usability testing methods and techniques that have been used in other disciplines, such as in software engineering, desktop GIS or electronic consumer products could be adapted efficiently and used together with methods presently applied to mobile applications in order to build the proposed frameworks. As mobile geo-applications running on mobile devices seem to integrate characteristics from all the above fields, a combination of these methods could be useful, something that will be investigated in this research.

A usability testing methodology for the evaluation of mobile geo-applications will be proposed and tested. The aim of this methodology is to find an appropriate combination of usability testing methods and techniques, using new technical solutions, in order to improve the current approaches. The results of this research may be applied in a larger research project on "Usable (and well scaled) mobile maps for consumers" (UWSM2, see URL_1.5). The goal of this project is to design and implement a mobile geo-application that will be able to provide the user with an adaptive map interface, in terms of dynamically changeable information and different contexts and conditions of use. The final product will be a combination of a tourist mapping and online parking information system. Utilizing modern high bandwidth mobile networking technologies, like Universal Mobile Telecommunications System (UMTS), the UWSM2 application will be able to receive up-to-date geo-information for the concerned geographic area from a mobile server, in order to create properly detailed map visualizations, modified in terms of user requirements and preferences.

Generalization is playing a very important role in this dynamically changing mapping context, although the user is not actually aware of the related processes running underneath but of the visual results. It helps reducing the amount of data sent from the server in each display refresh, as well as enhancing the map visualization, thus supporting the users to more easily understand the particular map by eliminating excessive information. The map scale and information content can be adjusted according to the selected context while at the same time the zooming and panning functions can be executed smoothly and without interruptions (URL_1.5).

In USWM2 project, several position-related, context-aware use cases, such as navigation to a point of interest in special circumstances or requirement of information regarding possible parking places are studied, in order to implement testing prototypes of the system. The contribution of the present research to this project will be the establishment of the possible methodology for the usability testing of these prototypes that are going to be developed, first as simulations in desktop environment and later as real-world applications. Effectiveness and efficiency of the map representation at different scales and user satisfaction through the use of this mobile geo-application have to be assessed as part of the usability testing in this research.

1.4. Research objectives

The core of the research is the development of a usability testing methodology that can be used in the actual prototype evaluation of UWSM2. Field testing with representative users in real contexts of use is the main focus area. UWSM2 has as major objective the implementation of techniques for vario-scaled map representations that can meet the needs of particular contexts of use and user groups' expectations, utilizing amongst others several studies regarding non-uniform scaling, generalization and progressive data transfer.

Research that has been already done regarding mobile geo-applications' usability testing, together with solutions that have not yet applied in this field, will be investigated in order to be used for the development of the methodology that will be proposed inside this research. The feedback from the field surveys that the proposed methodology will conduct will be then used to improve it. The usability testing will evaluate the map zooming and panning functions of a mobile geo-application similar to UWSM2, assessing the visualization results from the end-users' scope, in real conditions, and in simulated real usage scenarios. This main objective can be split into several sub-objectives:

To study different methods and techniques that have been applied for the usability testing of mobile geo-applications, as well as for applications from different disciplines, and especially those that can be applied in the field. To investigate what are the advantages and disadvantages of each of them in terms of resources and usability.

To build a methodology for field-based usability testing of mobile-geo-applications based on a combination of methods and the utilization of unique technical solutions. Different combinations of methods and techniques will be tested during user surveys in order to investigate the most appropriate, which then could be used for the usability testing of the UWSM2 prototypes.

To assess the results of the user surveys in the field, determining the usability of each combination of methods and techniques of usability testing in the predetermined context of use in order to adjust and improve this methodology.

1.5. Research questions

What are the investigated problems in already done research regarding usability testing for mobile geo-applications?

What are the different methods and techniques for usability testing and what are the advantages and disadvantages of each of them?

Which methods and techniques or which combinations of different methods and techniques would be the most appropriate for field-based usability testing of mobile geo-applications and particularly for the assessment of mobile map zooming, as a fundamental part of UWSM2?

What unique technical solutions could be developed and applied in order to overcome current problems of usability testing in the field?

Which is the usability testing methodology that is found to be the most efficient for the evaluation of the mapping scalability functions of the prototypes of UWSM2 in the field and why? How could this methodology be adapted in order to be used for the usability testing of mobile geo-applications in different contexts?

1.6. Thesis structure

As already stated, the main aim of this research is to develop a usability testing methodology for mobile geo-applications that can be applied to the prototypes evaluation of the project UWSM2. To do so, five stages are established in order to conduct this research. These stages contribute to the development of seven report chapters that present the outcome of this work.

In chapter 2, there is a literature review regarding mobile geo-applications, their issues, functions and limitations, together with their goals in specific use and user frameworks. Through the example of a vehicle navigation system, the specific context-based questions that the users can pose to the system are presented and the roles of context-awareness and adaptivity are investigated.

A more deep analysis and evaluation of the existing usability testing methods and techniques in general and for mobile geo-applications in particular is conducted right after, in chapter 3, where the different possible methods' advantages and disadvantages are investigated. Different examples of usability testing in implemented mobile geo-applications are now studied and compared, and conclusions are drawn from them, concerning their weak and strong points. This output, together with the investigation of data collection methods that have not yet applied to usability testing for map displays, will contribute to the construction of the proposed usability testing methodology that will be presented in chapter 4.

This methodology includes user testing surveys in the field, as it proves that field-based testing is more underdeveloped than the laboratory one. The methodology is comparing different combinations of methods and techniques for field-based usability testing in specific contexts. To support this, a unique technical solution that was implemented during this research is used in order to overcome the shortcomings of previous testing approaches and decrease the human resources needed. In order to build a methodology that can be used for the usability testing of the UWSM2 project prototypes, which are not yet available, a commercial mobile geo-application is selected, based on criteria regarding its similarity to UWSM2 expected prototypes. The usability testing addresses the fundamental variables efficiency, effectiveness and user satisfaction. Map scalability is investigated through the user testing, as the core research subject of UWSM2. Finally, the processes of selection of participants, formation of testing groups and setting of the field surveys are described, together with the user tasks and field-based realization.

The next step is the user surveys, utilizing the proposed methodology, presented in chapter 5. The way that these surveys were realized is presented, exposing the problems and limitations, and the improvements and adjustments that were made in order to overcome them. Analysis of the results is also done in this chapter, not in order to asses the usability of the map of the map zoomability of the selected mobile geo-application, or the users' performance alone, but the usability of the methodology itself in investigating usability problems. Comparisons between the different methods and techniques that were used, determination of the most appropriate combination of them, and validity issues are investigated. After that, the methodology is adjusted and improved according to the outcome of the analysis of the user surveys.

Finally, in chapter 6, the conclusions for the whole research work are drawn; the answers to the initial research questions are given according to the outcome of the previous parts, and the satisfaction of the objectives is examined. The general issues regarding the realization of the research are discussed, and recommendations for future extension and improvement of the work are given. The logical structure of the thesis is represented in Figure 1.4.

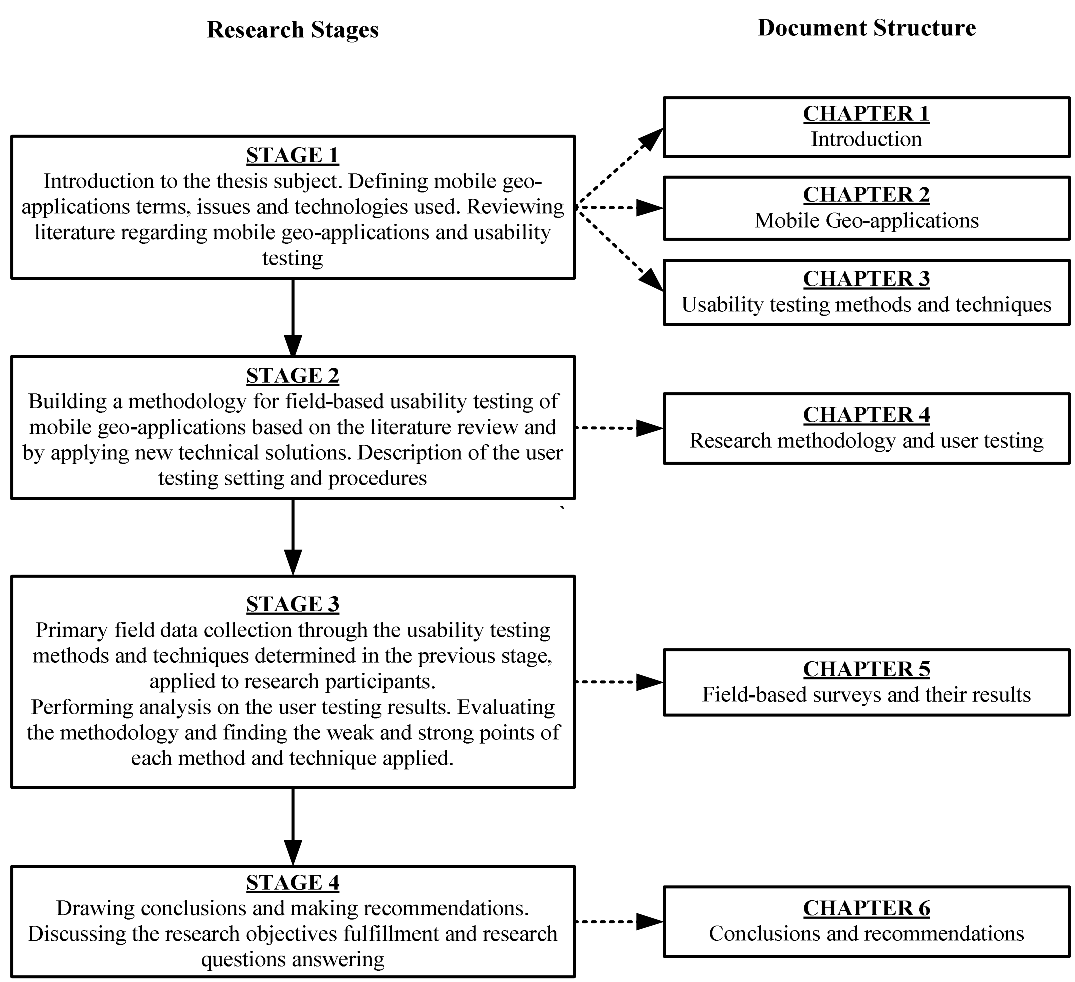


Figure 1.4 Adopted method and structure

2. Mobile geo-applications

2.1. Introduction

In order to start investigating the answers of the research questions initially formulated, a closer look to the mobile geo-applications should be taken. What are the reasons that this technology emerged? What are the components and functionalities of these applications and what user needs and requirements they try to cover? What are the limitations and advantages of these applications? Why the context of use is important in mobile environments? What are characteristic examples of such applications? These are some of the essential questions that this chapter is dealing with.

Chapter 2 is actually a literature review regarding mobile geo-application issues, functionalities, and limitations.

The chapter starts with a closer look at the reasons that contributed into the emergence of mobile GIS and mobile geo-applications, the parts that these applications consist of, as well as their functionalities and the needs that they try to cover. The importance of maps in mobile geo-applications is described, and examples of mapping applications are given. After that, an example of possible questions and tasks that the users of mobile geo-applications could have is given through the vehicle mobile navigation system case and the importance of context-awareness and adaptivity in mobile mapping is stated.

Information systems technology evolution has very much influenced the progress of GIS, as computer hardware is a main part. The first GI Systems were developed based on mainframe, then on desktop and finally on distributed computer systems. Distributed GIS, utilizing internet and wireless communication technologies, enabled the inter-change of geospatial data between heterogeneous remote systems and more efficient use of the current geo-data bases.

While internet-based GIS are referred as internet GIS, the wireless network-based are referred as mobile GIS. Distributed GIS took advantage of the rapid growth of inexpensive high speed internet connections as well as the introduction of modern web-capable desktop and mobile computers and devices. The necessity for the implementation of Internet-based Distributed GI Systems and services originated from three issues: First the rapid increase of the size, complexity and heterogeneity of spatial data files, which cannot be handled properly from stand alone desktop GI Systems. Second, the requirement for adaptable GI sub-units that can construct dedicated software implementations. And third, the need of implementing mobile applications that can utilize location-related information (Peng & Tsou, 2003).

Mobile GIS, as a sub-category of the distributed GI Systems, are related to wireless and mobile device environments. Personal Digital Assistants (PDAs), mobile (cellular) phones and Personal Media Players (PMPs) are some of these. According to Livingston (2004), mobile devices are the devices that are “small enough to fit comfortably into a purse, pocket or holster, so you can conveniently keep it with you at all times”. Therefore, tablet and laptop PCs are excluded from this category, as they do not match with Livingston’s concept while they are also regarded as portable rather than mobile according to the IEEE 2002 definition (Anderson & Blackwood, 2004).

Mobile geo-applications, as applications of Location Based Computing (LBC) are the software applications that run on these mobile devices, involving geographic information together with the determination of the location of the mobile user. Their aim is to provide the users with real-time and location-related information and services. Mobile Location-Based Services (MLBS) are examples of these applications, including, amongst others, route navigation, tourist information and emergency services. Through MLBS a personal profile of each mobile user can be created, utilizing a combination of data from different sources, like the GPS receiver data of the mobile device and several remote geo-spatial databases, together with the preferences of the user that can be asked from the system prior to use. These preferences can then personalize the system and adapt it to the particular user requirements. The personal user profile can then based on the surrounding environment and the conditions in the place were the user is, such as the traffic conditions, the weather conditions and forecasts, the near-by shopping centers, entertainment points and so on. In response, the system can give information to the user, as for example route navigation with turn-by-turn guidance, points of interest proximity notification, and location-related commercial services notification such us online parking payment possibilities.

Interoperability of mobile applications based on location is a fundamental part of LBC, including the exchange of huge amounts of geospatial data in a non-constant environment. As the mobile user moves through space and time, s/he is going through changes in the surrounding environment, related to the time of the day (day / night), weather conditions and place of action together with changes in the surroundings. The answers to the questions that the user poses to the system can also change accordingly. Thus the context of use is dynamically changing. The LBC is able to serve the user with the information that is inquired "anytime" and "anywhere", something that usually requires the combination of data from different sources, such as topographic and cultural Points Of Interest (POI) databases, involving at the same time connectivity, speed and reliability issues of the mobile / wireless network.

Mobile applications' interoperability is managed by location interoperability standards, as is the Geography Markup Language (GML), addressed by the OpenGIS Consortium in 2002 (Zadorozhny & Chrysanthis, 2004). Mobile geo-applications are able to bring to the mobile user geospatial information in the forms of text, image, voice, and video in the field. The required data can be transferred to the device from web servers using, amongst others, Wireless Markup Language (WML) (Li & Maguire, 2003). Maps in mobile geo-applications can be a big part of Graphical User Interfaces (GUIs), therefore Human Computer Interaction (HCI) methods and techniques can be applied to cartographic visualizations in this case (Sarjakoski et al., 2004). Explanation on HCI issues will be given in chapter 3.

Mobile geo-applications have emerged in order to meet the demands of two main user categories: everyday users that want a system that can support their activities and answer their location-related questions and field workers (Peng & Tsou, 2003). In both categories, the capability of serving geographic and other types of data related to the current location of the mobile user is the main aim. Since mobility is playing a continuously increasing role in everyday activities, which involve traveling and moving to unknown places and areas, technological support to their spatio-temporal decision-making procedures is becoming more and more important (Reichenbacher & Töllner, 2004). Vehicle navigation systems are examples of mobile geo-applications that supply location-aware geo-information to a specific category of users. These systems will be described in the following section.

2.2. Elements, functionalities and limitations of mobile geo-applications

As already mentioned, mobile geo-applications as applications of LBC differ from traditional GI systems in the way they utilize the advantages of two technologies, that of Informatics and wireless networking. In GIS the usual case is a desktop application with sometimes additional client-server functionality, whereas in mobile geo-applications several different contributors participate. Additionally, in the second case the requirements are much more than that of GIS, such as (Lopez, 2004):

- High performance
- Scalability
- Reliability
- Concurrency
- Mobility
- Openness
- Security
- Interoperability

A typical mobile LBS infrastructure consists of the mobile device, the wireless network, the communication gateway, the Application server, the Content server (Geodata server) and in the case of an application for mobile phones without positioning capabilities, such an integrated GPS receiver, a Mobile Position server. The example of a system like this can be seen in Figure 2.1.

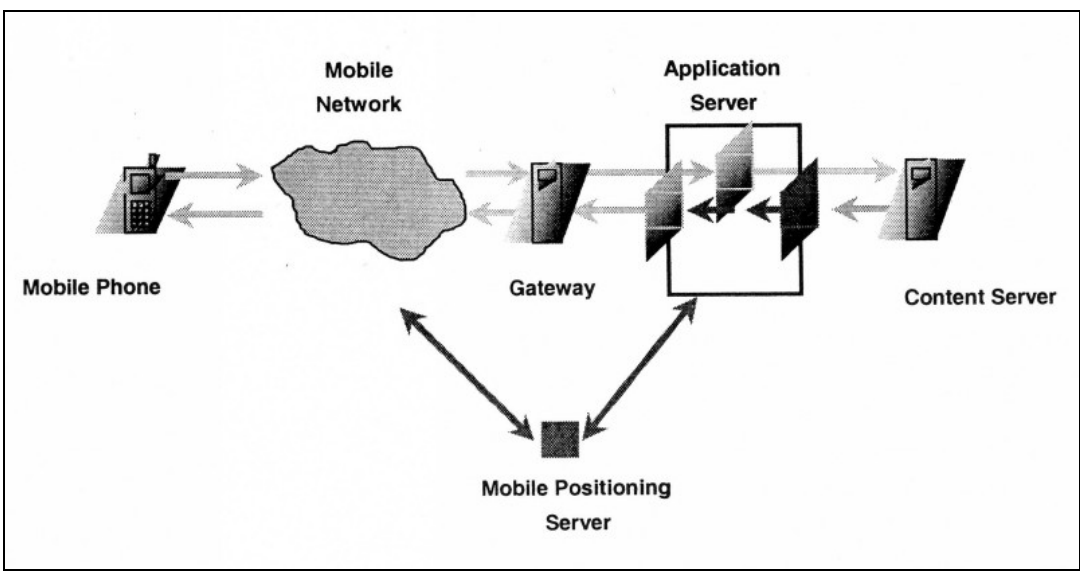


Figure 2.1 Typical mobile LBS infrastructure (Source: Lopez, 2004)

There are two types of positioning assessment in mobile geo-applications. The first is network-based, where several techniques are used to estimate the location of the mobile device, while the second is device-based, where integrated GPS and A-GPS technologies are used (Li et al., 2005; Ramm et al., 2006).

Maps as means of communicating geo-spatial information to the user play a significant role in mobile geo-applications. They help the user to understand geospatial concepts such as orientation, directions, distances, areas, and patterns, faster and easier (Nagi, 2004; Düpmeier et al., 2005). The graphical interface of maps can give a more understandable abstraction of the surrounding real environment, something that is much more difficult when the interface is textual or audible. Classic examples are car navigation systems, which almost always have a mapping interface, used along with voice directions and informative text and symbols. The maps are essential parts of mobile guidance systems, helping the navigational process and at the same allowing the access of different sources of information (Düpmeier et al., 2005).

There are several constraints concerning the use of maps in mobile devices, with the small size of the display being the most difficult to surpass. Other constraints have to do with the resolution of the screen, the maximum number of available colours for the visualization of data, the low processing power, small memory space, limited power source and wireless network bandwidth limitations. A lot of research has been executed recently in order to find proper solutions for the better cartographic visualization in mobile devices. Nagi (2004) developed a pedestrian tourist navigation mobile geo-application, using Scalable Vector Graphics (SVG) language and tested different types of visualizations assessing their effectiveness. He investigated, amongst others, different generalization, symbol placement, localization and features size techniques (Figure 2.2).

Elias, Paelke & Kuhnt (2005) did a research on visualization of landmarks and their impact on map perception. They tested different levels of abstraction of symbols according to their type and proposed a design guideline for new applications (Figure 2.3). Heidmann, Hermann & Peissner (2003) explored the possibilities of use of simplification and abstraction visualization techniques together with interactive labelling and hiding of information inside tool tips. The outcome of the research was used for the development of the SaiMotion LBS prototype (Figure 2.4).

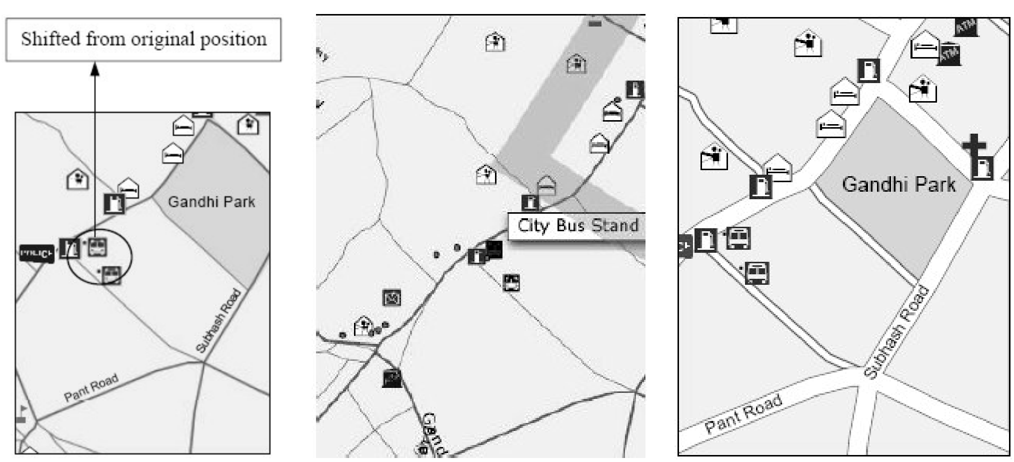


Figure 2.2 From left to right: Shifting for prevention of overlapping of symbols, tool tip option for enhancement of readability, and feature width variation investigation (Source: Nagi, 2004)

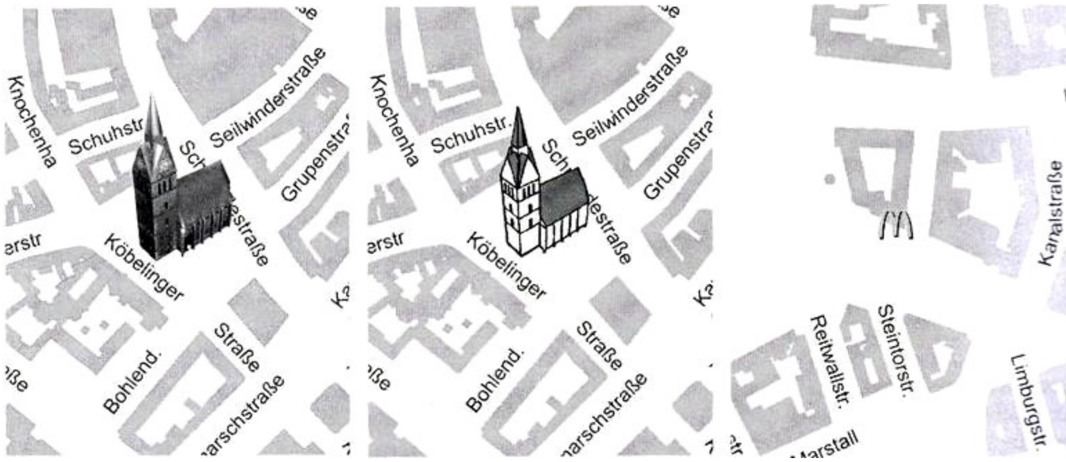


Figure 2.3 Different visualization cases. From left to right: using a building's image, a building's drawing and the logo of a shop on the map (Source: Elias, Paelke & Kuhnt, 2005)

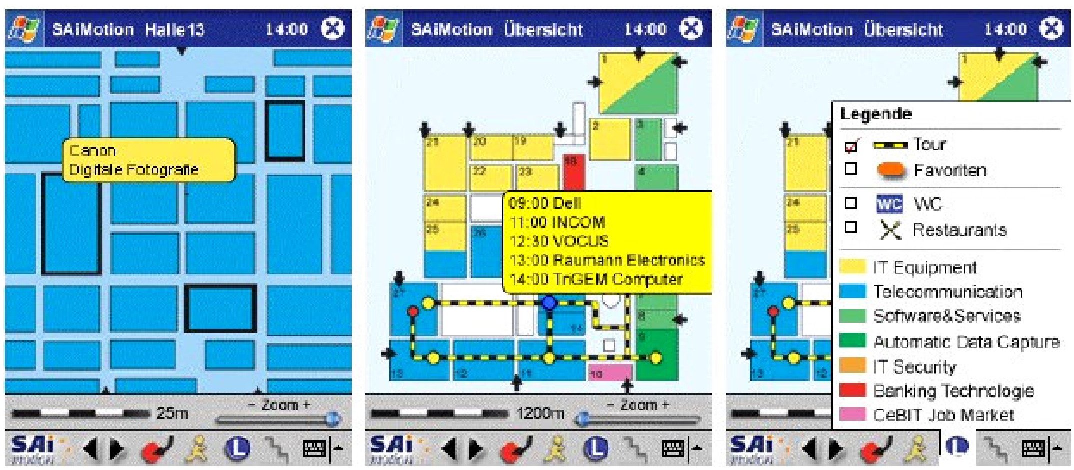


Figure 2.4 SaiMotion prototype interactive maps. From left to right: labels hidden inside tool tips, route and tour elements and a hidden label showing extra information, and user defined labels (Source: Heidmann, Hermann & Peissner, 2003)

The challenge of mobile map visualizations is to keep the graphical / mapping interface clear, but at the same time pleasant to the user, while providing him/her the amount and quality of information that can satisfy his/her requirements and expectations. The small screen does not allow for the visualization of more than a limited number of objects in a certain viewpoint, while the ability to involve usage of labels in order to describe these objects is also very limited (Heidmann, Hermann & Peissner, 2003).

Another issue in the use of mobile geo-applications is the limited capabilities of keyboard input in mobile devices; this enforces the mobile applications to evolve less user interaction during the operation. Location and profile of the user can be used in this case to adapt the application to the special user needs, thus eliminating the need for frequent keyboard use.

Finally, data exchange with the server should be the lowest possible, because of the cost of wireless / mobile data transfer as well as the continuously changing transmission factors that affect the quality and speed of communication (Markoulidakis et al., 2004). It has to be mentioned here that it is not necessary that mobile geo-applications make use of a mobile or wireless network, as the necessary geo-database can be inside the storage memory of the mobile device such as a PDA and the geo-location can be acquired through the use of an integrated GPS receiver, making the system autonomous. The weak point in this case is that this mobile geo-database cannot be updated in real time and it is in the responsibility of the user to update his/her data whenever updates are available and s/he is aware. This can lead to absence of important map changes as for example in car navigation or tourist information applications. The procedure of updating can be done through the use of a PC connected to the application's support centre via the internet.

2.3. Use of maps and tasks in mobile geo-applications: the vehicle navigation example

Vehicle navigation systems' purpose is to assist the drivers when they are following an unfamiliar route, while giving the minimum possible impact on their concentration on the road. They can be in-dash integrated, or mobile device-based, with the latter case being of interest for the current research. Their main function units are:

- The geographic location sensor, which is a GPS receiver integrated to the device or a separate GPS module wirelessly communicating to it via Bluetooth connection. GPS receivers are also giving information about the speed of the vehicle, the time of day, the altitude and the direction of movement.

- The digital maps, which are usually stored in the physical or expanded memory of the device (for example on an SD memory card), or can be retrieved from a remote server through a mobile data connection such as UMTS or General Packet Radio Service (GPRS).

- The software that makes the route calculation with the use of the CPU of the device.

- The map display / interface that is used for showing the navigation information in the form of maps and / or guidance graphic symbols and text, allowing at the same time the input of information from the user or the manipulation of the map through functions like zooming, panning, or the selection of features. As Kraak & Ormeling (2003) argue, navigation and orientation are virtually the crucial functions of maps, so they play a very important role in vehicle navigation systems too. They help the driver conceptualize the geographical surroundings, and have a better overview of the planned route and the movement of the car in relation to the surroundings. The direct manipulation of the map through zooming, panning or the selection of features directly on the map in order to retrieve information or use them for the routing process, are additional functions that can enhance the usability of the system.

A typical mobile vehicle navigation system that uses an offline map database stored inside an SD memory card is shown in Figure 2.5.

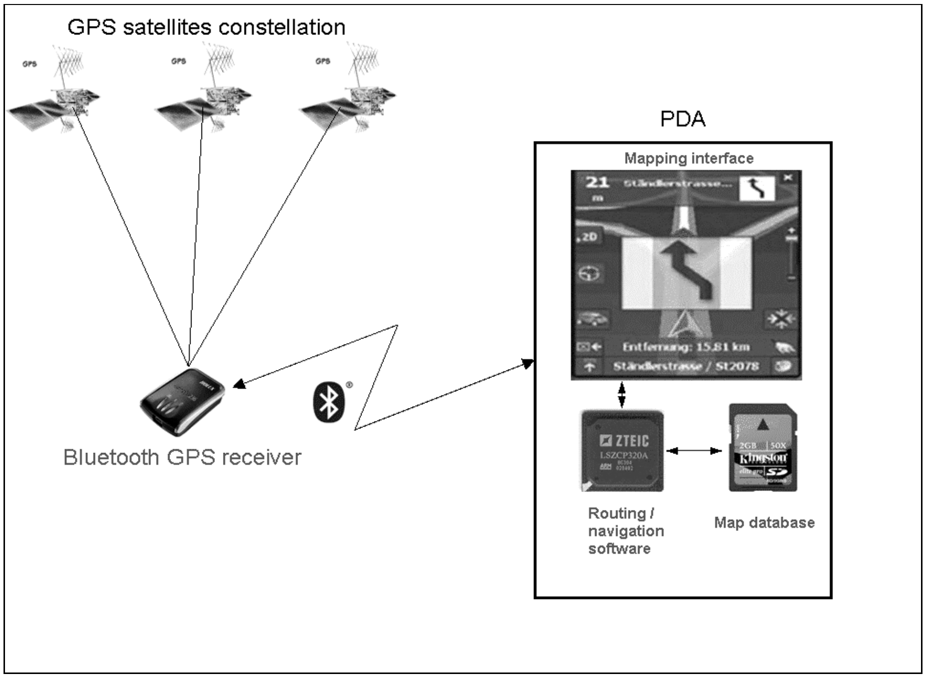


Figure 2.5 A typical (offline) mobile vehicle navigation system with the use of PDA

The driver can insert the desired destination through different ways, such as directly typing the name, searching it in a favorites' list, pin-point it on the map or retrieving it from a proximity-based search or characteristics-based process (e.g. searching for the nearest gas stations). The easiness of selection of a target is still low in these systems, particularly when the user is driving at the same time (Tijerina, Parmer & Goodman, 1998 as cited in Lee, Forlizzi & Hudson, 2005; Burnett, Summerskill & Porter, 2004). Since there is the ability of location tracking, recalculation of the route can be done automatically in case that the vehicle goes out of this route, in contrast to turn-by-turn instructions or paper maps. Although mobile vehicle navigation systems provide a lot of advantages in comparison to traditional means, such as paper maps, they can also issue some serious safety manners, due to the disturbance of the concentration of the driver. Information is usually being supplied without taking into account the context of use including the driver's cognition (Lee, Forlizzi & Hudson, 2005).

The usual map visualization in mobile navigation systems are maps that show the vehicle's position in the middle, while the map is scrolling dynamically according to the movement and orientation of the vehicle. Heading up in 2-D or 3-D perspective is the default mode of presentation, while the user can change it to north-up in most of the systems. Current map visualizations in vehicle navigation systems are often based on paper map styles, producing a complicacy which is worse than that of paper maps, as the displays of mobile devices are very small in comparison to paper maps. Zooming functions help to manage this shortcoming, but unfortunately they also lead to a kind of shortsighted perspective of the geographic space, reducing the understanding of the overall route. Moreover, drivers still have to pay attention in order to comprehend the displayed information, even when the screen shows only a small fraction of the overall map.

The use of easily understandable graphic symbols instead of maps to represent turns or crucial parts of the route is another method that has been investigated together with concurrent voice instructions (Figure 2.6).

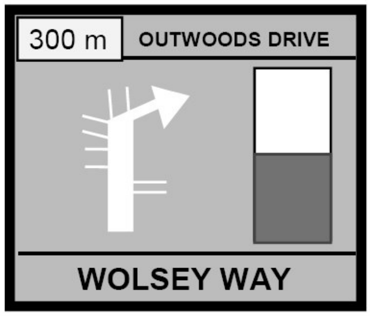
DISPLAY	VOICE
	"Turn right in 300 metres"

Figure 2.6 A Typical non-mapping display and voice instructions in a vehicle navigation system (Source: Burnett, 2000)

With the use of these graphic symbols alone, the visual exploration task is simpler, but not without negative consequences. These are the elimination of any context-related information that helps the driver preserve a conceptual model of the geographic space and his position inside it, together with the absence of a visual representation of map features that can help him/her make a decision concerning alternatives when he is planning or following a route (Lee, Forlizzi & Hudson, 2005). Realizing the importance of maps in mobile navigation systems as well as in mobile geo-applications generally, lead to use of abstraction methods that have been already used in cartography, and especially generalization, in an effort to make maps use easier and more usable in mobile environments. Easier and faster map reading and understanding are the goals. According to MacEachren, 1995, as cited in Lee, Forlizzi & Hudson 2005, the understanding of a map consists of three stages. These are first a "precognitive visual array", where a detection of shapes, boundaries and edges is taking place. Second is a "2.5D sketch" wherein the user's short-term memory is storing a visual outline of the representation and compares it with his/her actual knowledge. And third, the interpretation of this representation by the user.

Navigation is a procedure that has the aim of accomplishment of moving inside the geographical space. Wickens and Hollands (2000) cited in Lee, Forlizzi & Hudson (2005) argue that intellectual cognition of geographic space is founded onto three knowledge types. These are the landmarks, routes and surveys. Landmarks are representations of the characteristic spatial features of a particular area, like train stations, governmental buildings and parking areas. Routes are the representations of the way to go from one place to some other. And surveys are synoptic geographic perception of the real environment that enables the user to sketch a precise map. The most objective knowledge is the last one, as it is the outcome of the generalization of experiences. Usually users obtain cognition about an area regarding landmarks, routes and surveys, in the order that they are presented here.

There have been several proposed models about how navigation through driving is realized. Michon (1985) as cited in Lee, Forlizzi & Hudson (2005) splits the task into strategic, maneuvering and control levels. The strategic level involves the planning of the route, the maneuvering level involves the preservation of the position of vehicle inside the route, and the control level involves the manipulation of the vehicle in order to obtain its movement according to the traffic rules and the route. An expanded model based on Michon's concept, including trip planning, preview, identification, confirmation, trust and orientation was proposed later by Burnett. This model incorporates driver's requirements together with navigation system capabilities (Burnett, 1998; Ross & Burnett, 2001).

But what are the geographic questions that (mobile) vehicle navigation systems try to answer? According to Nyerges (1991) as cited in Nyerges (1997), the categorization of geographic questions can be done according to their *location and extent, distribution and pattern or shape, spatial association, spatial interaction, and spatial change*. In mobile vehicle navigation systems' case, the questions are mostly associated with location and extent, involving temporal characteristics too. Some examples of these questions are the following:

- Where I am now?
- What is around me?
- Where is the place / feature I want to go?
- What else is there?
- What is the distance to get there?
- How much time I need to get there?
- What is the fastest / shortest route to get there?
- How can I avoid / include a place inside my route?

According to the above geographical questions, we can determine the tasks of the users of this type of mobile geo-applications. These can be addressed through the answers to three major questions:

- What the users want to do?
- What information is required in order to do this?
- What actions are required?

Let us now try answering these questions. First of all, we have to know who exactly the users of the system are, and in this case they are the drivers of cars. And they want to navigate in order to go to a desired destination, which means to make a plan of a trip from a point A to point B. Navigation through driving of a vehicle requires a lot of attention, as the driver has to make an efficient planning in terms of time, distance and trip comfort (e.g. avoidance of very small roads). At the same time, s/he has to keep the vehicle inside a safe speed limit and inside a right path, according to traffic and road conditions, manipulating vehicle controls accordingly (Mitrovic, 2001). Vehicle navigation systems aim at reducing the attention needed from the driver and help him/her in navigating. The driver's overall task is to navigate safely and efficiently to the desired destination while assisted by the system.

As it comes to the second question, it is obvious that in principle, the system must be informed about the user's requirements and preferences regarding the guidance driving task. The system must know the required destination and the trip parameters, such as the desired maximum speed and time arrival, together with the specifications of location, speed, altitude, time and movement's direction that can be acquired through a GPS receiver.

The required actions, as posed in the third question, are to enter the route planning parameters to the system, to check the planned route in order to find possible shortcomings, overviewing the map of the route and start driving while paying attention and following the system's visible and / or audible instructions and warnings.

2.4. Context-based mobile map interfaces

Maps, as already mentioned, are very often playing the role of interfaces in mobile geo-applications, efficiently communicating geospatial data. At the same time, mobile applications that are based on the use of maps are regarded as promising field for the application of the latest mobile device and communication technologies (Hampe & Paelke, 2005). In order to override the display limitations of the mobile devices, a series of techniques are used, which try to utilize the small, low-resolution screens optimally. Their main goal is to represent sufficient information, while it is still clear, understandable and helpful to the user. Different techniques have been implemented and a lot of research is taking place nowadays in this field, while still there is a long way to go. Some examples of this research are thereupon presented.

Bjørke (2003) investigated the use of new algorithms in order to achieve automated generalization of road networks, by eliminating the conflicting arcs in regards to the scale of maps and road hierarchies. Adaptivity of maps to dynamic contexts of use and maximization of the useful information in the map representations were the motivations for this research. A basic parameter, named T, defining the separation of points is changed according to map scale and the capability of user in terms of visual perception, changing the generalization result. It is stated that this method, giving good results in the conducted case study of the research, can be efficiently adjusted and parameterized so that can be used in different mobile geo-applications.

Goldsberry & Dillemuth (2006) researched the possibilities of the implementation of context-aware mobile mapping applications based on the GPS receiver measurements in order to dynamically change the mapping representations. The study associates the GPS data specifications of location, time, altitude and speed and direction of movement with the cartographic elements of map centre, level of contrast and colour scheme, map scale and extent and orientation. The main subject is the investigation of techniques that can sense user activities in order to not only affect the generalization and selection functions during map representation, but also the map scale and extent of view.

Adaptation is a key concept in mobile geo-applications, as it can help overriding the limitations of the current mobile software and hardware platforms (Frank, Caduff & Wuersch, 2004). It adjusts a system by changing its characteristics in order to respond according to a particular context of use. Adaptation role is to make the system able to cover the specific user group requirements efficiently, by giving them the needed information in dynamically changing conditions, taking into account their personal profile parameters (Goulimis, Spanaki & Tsoulos, 2004).

According to Dey (2001) as cited in Nivala (2003): "context is any information that can be used to characterize the situation of an entity, where entity means a person, place, or object, which is relevant to the interaction between a user and an application, including the user and the applications themselves" Context awareness in mobile geo-applications can be active or passive. In the active case, the user can provide special personal information to the application so that it adapts to his/her special needs, while in the passive the system can retrieve it through the way of use of the application, as for example the response of the user to system question pop-up windows (Nivala, 2003).

Context aware mobile geo-applications are obviously user-centered, taking into account several parameters available to the system in order to adjust better to the user tasks. Location, system properties, purpose of use, time, physical surrounding, navigation history, orientation, cultural and social issues, and user personal profile are the most important. Maps have the role of graphical interfaces in mobile geo-applications as it was mentioned earlier, so the mobile user becomes a mobile map user. (Figure 2.7)

Adaptation can be divided into four elements: personalization, which refers to the adaptation of the content to a particular person, individualization, which refers to the adaptation based on the requirements of a group of persons, localization, which refers to the adaptation requirements of a particular location such as a specific interface language, and customization, which refers to the adaptation of services to a particular user (Reichenbacher, 2004).

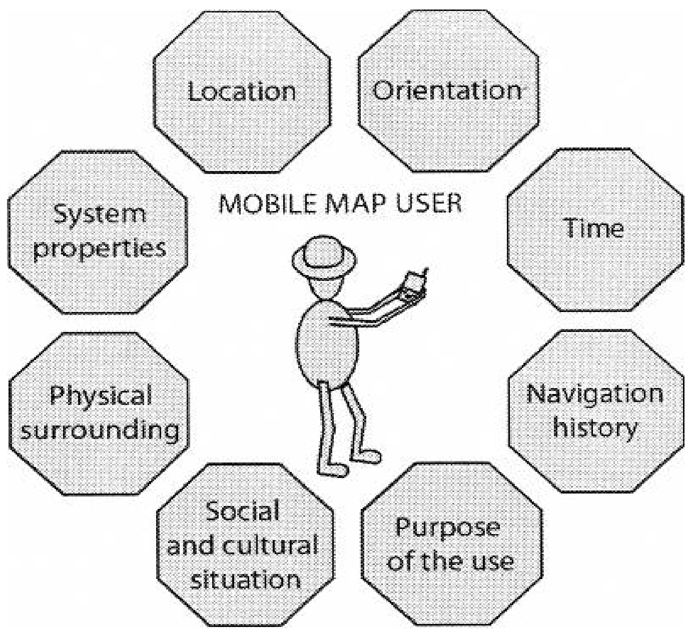


Figure 2.7 Mobile map variety of contexts that circle the user (Source: Sarjakokski et al., 2004)

In mobile geovisualization, an adaptive map is a map that automatically changes its visualization parameters according to a specific user profile, the user's activities, the particular mobile platform technical specifications, or the required information. This, from the other side, doesn't mean that the application will generate a different result for every individual user, but rather this user will fall into a group of predefined categories of solutions that are available (Reichenbacker, 2004). Regarding abstraction, adaptation represents an interpretation amongst the User Reference Frame which is his location and orientation, the Map Reference Frame which is a real world's abstractive representation and the Absolute Reference Frame, which is the real world. This means that any single point in one of these three frames can be projected to the others (Figure 2.8). This interpretation amongst the different frames enables the system to adapt the mapping interface, while adaptation consists of four elements: panning, rotating, zooming and selection (Frank, Caduff & Wuersch, 2004).

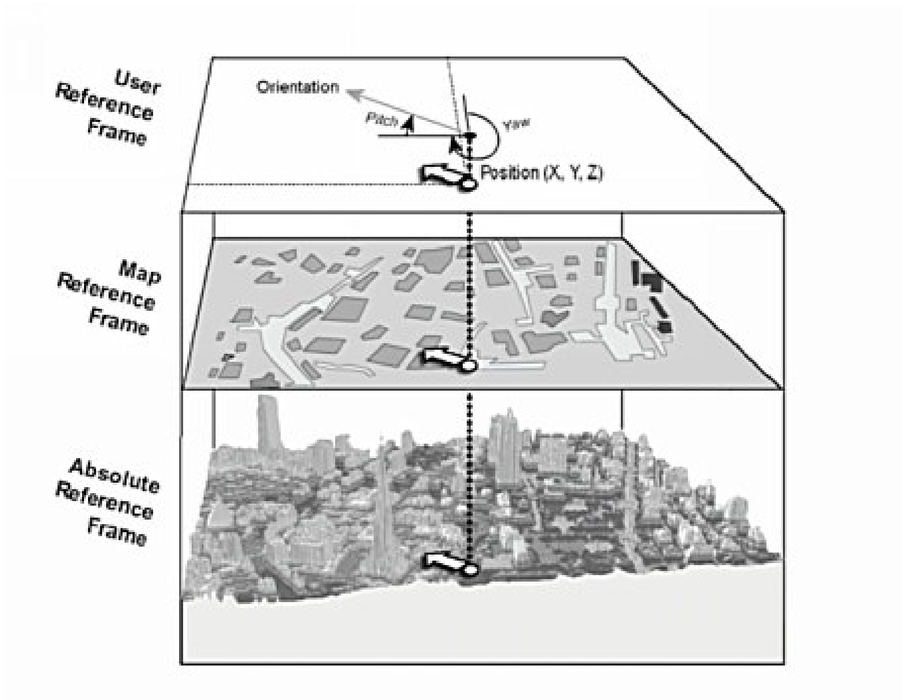


Figure 2.8 Relations between the real world, the map frame and the user's location and orientation (Source: Frank et al, 2004)

2.5. Conclusion

This chapter dealt with the mobile geo-applications, regarding their uses, functions, and different technologies applied. Maps usually play an important role in these applications, trying to support the user's geographic decision making and his activities, such as the navigation through a mobile map. A vehicle navigation system, which is then described, shows how an interaction between a user and a mobile navigation system can take place. The user's questions that such a system has to answer, according to the tasks that s/he has to carry out, reveal the roles of context awareness of a mobile geo-application and adaptation to specific user needs as contributors in order to enhance user-friendliness.

It is argued that there are different kinds of mobile geo-applications under development or in the market today, aimed at covering different user needs in different contexts of use. But how can they success in meeting these requirements? Usability testing, studied in the next chapter, is one of the solutions regarding this issue. An overview of the methods and techniques for usability testing will be given there, together with the unfolding of the shortcomings and advantages of each method. A deeper look into the usability testing that has been already applied to mobile geo-applications, through several examples will then be presented.

3. Usability testing methods and techniques

3.1. Introduction

This chapter starts with a close look at usability, HCI and UCD. After that, an overview of the different usability testing methods and techniques is given, including the laboratory-based and field-based ones, the advantages and disadvantages of each of them and the stages of a design that they can be applied to. Finally, conclusions are drawn, according to which the next chapter's research will be carried out. Is usability testing beneficial to mobile geo-applications? Is there space for more research on usability testing and improvement of current methodologies? These are some of the questions that this chapter will try to answer

Usability is actually a measurement of the usefulness a system from the user's perspective. It can be divided according to ISO definition (ISO 9241-11) into three fundamental elements, which are: effectiveness, efficiency and user satisfaction in a specific context of use (Jordan, 1998; Williams, 2004). These are significant parameters for mobile geo-applications, and although they have been researched thoroughly on desktop computers, in the mobile field there is still a long way to go. Usability testing is important in order to assess the effectiveness, efficiency and satisfaction of the user-mobile device interaction in real conditions. In this way, lack of usability testing during the design and development of an application often leads to user dissatisfaction and rejection of an ulterior use (Kimber, Georgievski & Sharda, 2005).

The usability testing involves participants that are executing predefined tasks. The feedback that they give through this procedure is collected through different methods and then used to improve the application and correct possible design mistakes or problems. The majority of mobile geo-applications usability testing is done nowadays in the laboratory while there is a strong need for testing on the field too, which is relatively rare, conducted by only 19% of the current studies. (Kjeildskov & Graham, 2003; Kaikokken et al., 2005). This need comes from the fact that the interaction of a mobile user with the mobile device / application involves interaction with the surrounding real environment too, and usability cannot be properly checked in means of controlled laboratory environments alone, as a big part of the contextual information cannot be investigated (Mennecke & Strader, 2003; Leitner, Plattner & Hitz, 2006; Kaikokken et al., 2005). As Chan & Fang (2003) inside Lim & Siau (2003) argue: "Traditional means of user interviews or usability testing in a laboratory environment cannot reveal insights into users' activities and mobility in real life. Context becomes a critical consideration in gathering information about user requirements".

According to Nielsen (2003), usability has a dual meaning as a definition. It outlines both a variable of quality of a system in terms of easiness of user interfaces, and a collection of methods, techniques and procedures that are applied through the design and development phases of a project in order to make it easier to use. Bevan (1999) refers to usability as "quality in use", together with functionality, reliability, efficiency, maintainability and portability of software products. Koohang & Harman (2007), argue that a distinguishion between usability, usability engineering and usability testing should definitely be made, in order to avoid confusion regarding their exact meanings. Usability refers to the theory and its application of making usable products, usability engineering is the materialization of usability in the design process of a product, and finally, usability testing, is the evaluation process of the usability of the implemented product. "The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order" (Koohang & Harman, 2007). This fact is very important in software applications and user interfaces, such as mobile mapping interfaces, where there should be a good connection between the user's conceptual models of the world and their artificial representations on the interface.

3.2. Variables of usability

Usability, as described earlier, is the degree of effectiveness, efficiency, and satisfaction of the user of a product in a specific use context. The specific context of use is a very important issue, as it is obvious that a product that is designed to cover lucid requirements of usability for a specific use under specific conditions, in any other case it may behave totally different. Navalkar (2004) gives an example of this fact, mentioning the alteration of usability value of a car's steering wheel that is designed to be used by a usual driver when it is used by a disabled one. So a user-centered design and implementation approach would be very beneficial, where the precise context of use and the exact user's nature should be deeply studied in order to meet the three main aspects of usability that were presented earlier (Gulliksen et al., 2003).

Usability should not be confused with utility, which addresses a different way of assessing product or service implementations. While usability refers to the easiness of use under specific contexts of use, utility refers to the product quality, independent of the context of use, described also as functionality by Erdogmus & Tanir (2002), For example, in most cases the web sites lack usability, but their utility is good, meaning that the user can retrieve the information s/he wants, but this is done through spending a lot of time trying to figure out how this can be done (URL_3.1).

Usability of a product can be described as a qualitative evaluation attribute, which consists of several elements, expanding the three fundamental ones into six (Nivala, 2005):

- Effectiveness: the capability of the product to enable users completing missions and satisfying objectives with accuracy and completeness.
- Efficiency: the performance of the users in completing tasks in relation to the resources spent for this process.
- Satisfaction: the level of pleasure that the functionality and friendliness of the product give to the users.
- Learnability: the level of easiness in learning the manipulation of the product the first time that they come in touch with.
- Memorability: the sufficiency of recalling the handling of the application after an idle period.
- Errors: the amount of errors during the use of a product by the users, the severity, and the easiness of getting over them.

It can be argued that satisfaction and memorability are somehow variables for assessing effectiveness and efficiency of a product, as a product which can be more effective and efficient when it is learnable and memorable. Similarly, errors' amount is a variable for assessing effectiveness.

Therefore, usability is actually assessed all the time by the three fundamental variables that were initially presented).

The upcoming ISO 13407 standard integrates usability testing methods within a general user-centered framework of design and implementation following an iterative approach according to Bevan & Curson (1998). This framework is shown in figure 3.1.

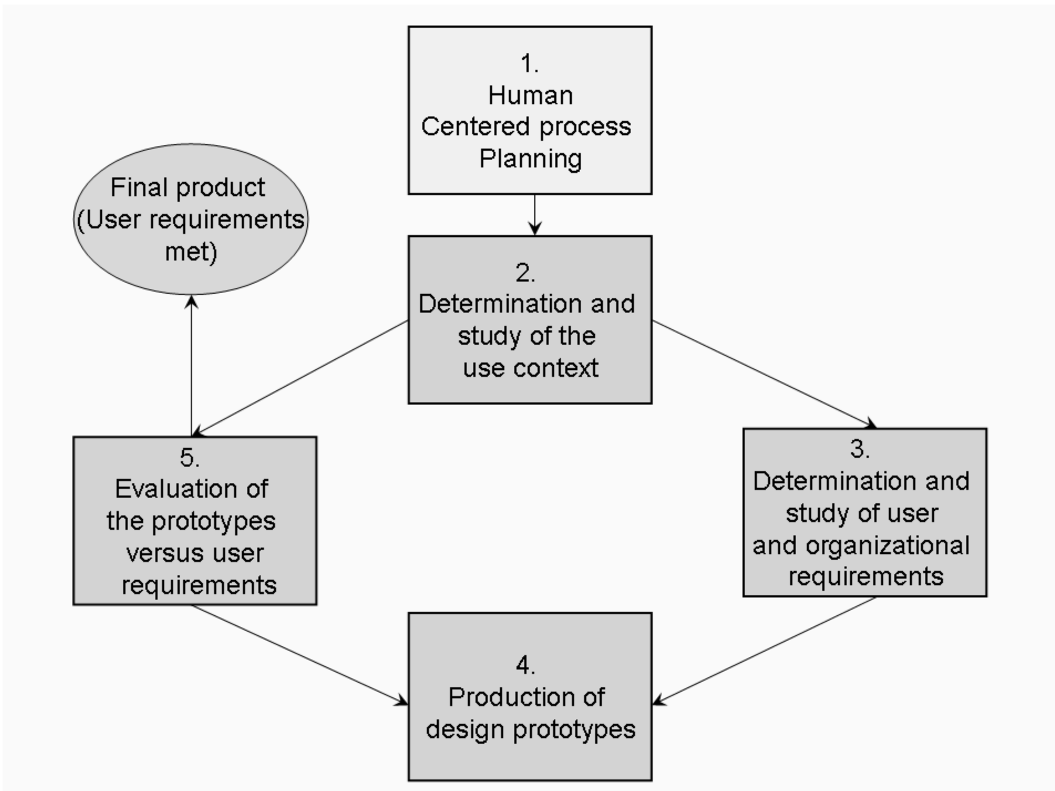


Figure 3.1 Overview of ISO 13407 user-centered design framework involving usability testing (adopted from Bevan & Curson, 1998)

The usability testing is an empirical study or analysis of usability that can provide a measurement of the users' easiness of going through a particular system or prototype as well as exploring its specific contents of it (Rosson & Carroll, 2002). To be usable, it has to derive users' prospect regarding the level of difficulty in utilizing an application, the satisfaction they obtain through the use, and the effectiveness and efficiency of use. (Wichansky, 2000, cited in Zhang & Adipat, 2005).

Although usability testing can be conducted through several methods and techniques, the most helpful is the user testing, consisting of a series of actions (Brink, Gergle & Wood, 2002):

- Planning of the test, by specifying the objectives and tasks
- Bringing all the necessary materials for the testing
- Preparing of the test site, and equipment in order to be ready for the testing
- Pilot testing in order to investigate possible problems in the testing process

- Contacting, selecting and scheduling appointments with the participants
- Carrying out the testing, by welcoming, instructing, and give the tasks for testing to the participants. Applying the pre-selected usability testing methods while the participants conduct the test. Debriefing of participants at the end of the testing
- Analyzing the results, by investigating the problems that were found, classifying them according to their frequency and seriousness, and finally try to determinate possible ways of solving these problems
- Preparation of the testing site and equipment again in order to conduct the next test, and applying the solutions of the problems in the testing process that were found in the previous part

Currently the amount of resources spent for the usability improvements is still generally very low (Nielsen, 2003). Usability testing for mobile geo-applications is a challenging and still young field of research, dealing with the restrictions and special features of the mobile devices. The dynamically changing mobile environments, heavily dependent on contextual factors such as time, location, social and cultural environment, natural surroundings, orientation, navigation history, purpose of use and system properties (Sarjakokski et al., 2004) require approaches that are apparently different from that of desktop computers and their applications. Appropriate methods and techniques for the usability testing of mobile geo-applications are strongly needed, in order to investigate the special human-application, human-device and human-interface context-specific interaction issues, and contribute to the solution finding and improvement making processes.

What are the different methods used generally for the usability testing in the field of human and computer interaction? What are the advantages and disadvantages of each of them? In which stages of the design and development process they can be applied? Are some of the questions that this chapter is aiming to answer.

3.3. HCI and usability

The science of Human-Computer Interaction is a rapidly developing field apparently triggered by the conference “Human Factors in Computer Systems”, which was organized by the National Bureau of Standards of USA, on March 1982. HCI is studying and practicing usability engineering and testing regarding software and technological applications that are made to be used by humans. These applications should be usable. HCI combines computer science R&D with social and behavioural science’s applied theory (Kumar, 2005).

The main four areas of research that contributed through collaboration into HCI are:

- Iterative design and prototyping used in software engineering field
- Human factors and software psychology used in computing systems field
- User interface software used in computer graphics field
- Frameworks used in cognitive science

HCI concentrates mostly on software interfaces (Bennet, 1984; Shackel, 1984 as cited in Han et al., 2000). There are several ISO standards related to usability under HCI. They can be categorized to attributes as follows and be visualized in figure 3.2.:

- Product's use including efficiency, effectivity and user satisfaction in a specific context of use
- User interface and interaction of the user with it
- Product's developing process
- Organization's capability of applying a user-centered design

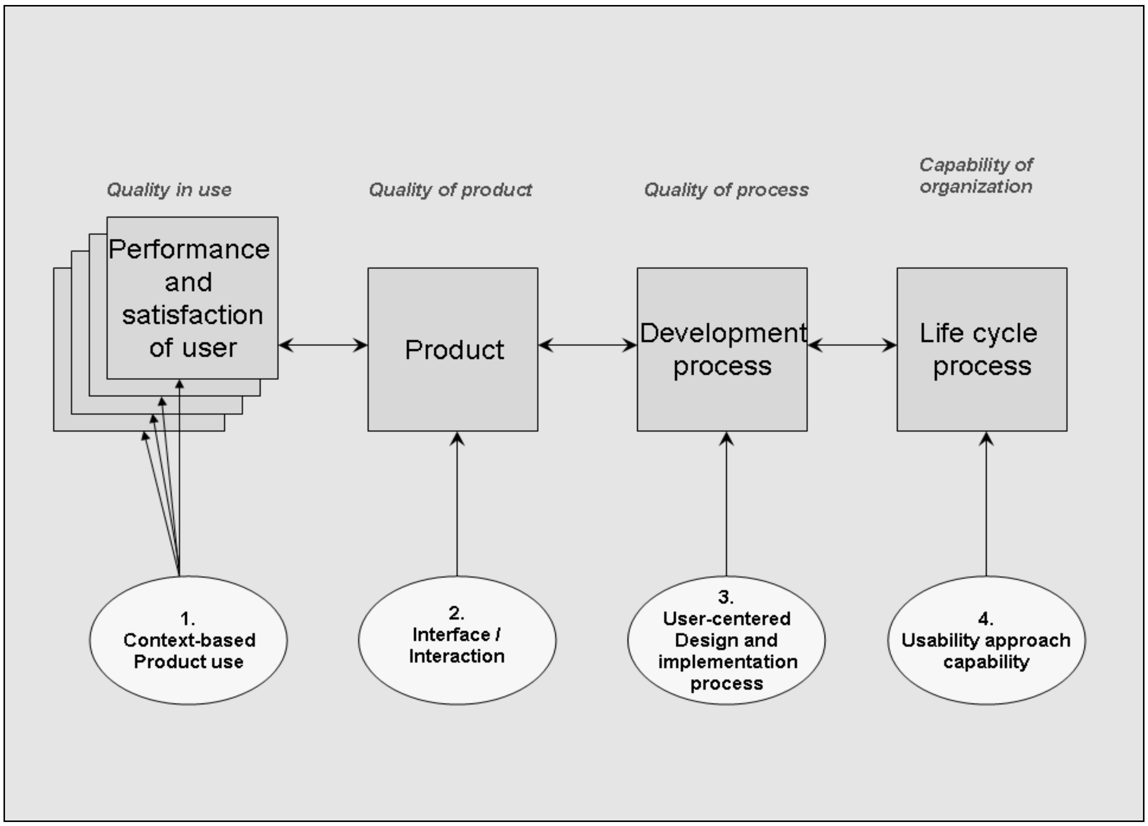


Figure 3.2 Categorization of HCI standards related to usability as quality in use (Adopted form URL_3.2)

3.4. User – Centered Design

The concept of User-centered design (UCD), rose from HCI theory, is a philosophy and at the same time a set of methods, that encompass the end-users' participation to a design development and realization, in a dynamically influential way. Specifically, UCD is aimed at verifying the usability of a particular system according to users' requirements and expectations. The central idea behind it is that the actual user has to be a central part of the design and development stages of any system, and in collaboration with the designers and experts, to make it capable of addressing the requirements and special needs of him/her. There are two basic variations of users' participation in UCD. The first is involving the users at special stages of the design development, usually in the determination of requirements and usability testing phases, while in the second they regarded as actual design partners who strongly influence the whole design from the start until the end (Abras, Maloney-Krichmar & Preece).

According to Norman (1988), cited in Abras, Maloney-Krichmar & Preece (2004), the requirements and expectations of the users should be deeply studied, and the design's usability should be the main consideration. A design should cover the following targets:

- Easiness of determination of possible actions regardless the phase of design
- Transparency of every part of the design, such as the fundamental system's model, the possible alternative processes , and their effects
- Easiness of inspection of the current system's state
- Adoption of natural analogies to investigate the relations between goals and required activities, activities and their effects, and information and system's state

The user is here obviously in the core of the design, and the designer's goal is the implementation of a usable product.

The initial action in the UCD process is the determination of the final users of a product together with their actual contexts of use and missions. According to Macaulay (1994) the users can be separated into three categories: the primary, secondary and tertiary users. These parameters are then used to specify the requirements from the final product. A design prototype is afterwards implemented and evaluated by the users, who give a feedback in relation to the previously determined requirements. This feedback can be used to improve or even substantially modify the initial design. This procedure is continuously applied, until the prototype meets the users' requirements, so that it can be put into mass production.

A new UCD – based product implementation according to figure 3.1, starts by Requirement Analysis, where a deep investigation of the main users and tasks takes place, together with a study of the possible contexts of use. The problems and the level of satisfaction of the users with existing products are identified through user research, and the first organization and users' requirements from the new design are set. Several scenarios of use are formulated into use cases in this stage, leading later to the initial design of the interface which is task-oriented, determining both the system functionalities and the user-system interaction queue (Hermann & Heidmann, 2002). The available resources determine the way that will be followed for the UCD implementation, taking into account the parameters of usability testing that will be used, such as the methods, the participants, and the applicable project phase (Nivala, 2005). After that, the development of the first prototypes is realized, aimed at assessing the capability to meet the needs of the users in relation to their actual tasks. Representative user evaluation in this stage is very useful, as the cost of prototyping is much lower than that of the final product and can help in discovering possible design problems from an early stage. The advantages of the UCD approach are the cost reductions of production, and more usable products. The UWSM2 project development is also based on a UCD approach, based on the ISO 13407 user-centered implementation framework. The diagram of this approach can be seen in Figure 3.3.

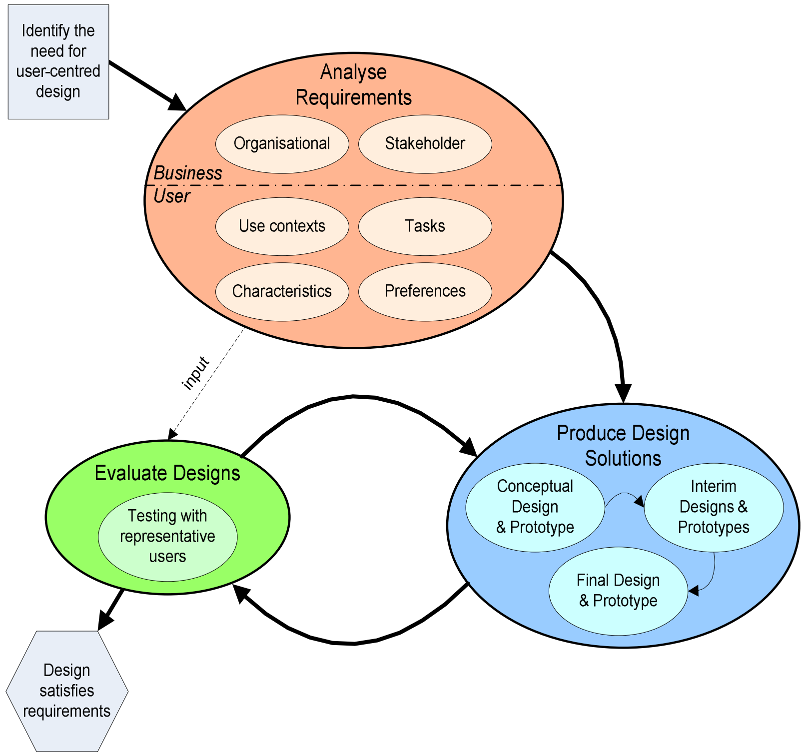


Figure 3.3 A general diagram of the UCD project design and development process (Source: Van Elzakker & Wealands, 2007)

3.5. Usability testing methods, validity, reliability and participants

Usability testing is a procedure that should be ideally applied through the whole design cycle (Genise, 2002). It can be performed by real or representative users, while it is testing real or representative systems. (Whitefield, 1991, as cited in Nivala, 2005), and it can be divided into three categories: testing, inquiry and inspection.

The testing deals with the evaluation of prototypes or final products, by observing the execution of several tasks from the users. The inquiry is based on communication of the researcher with the users, where a feedback about possible improvements of the product and the level of effectiveness that is obtained. And, finally, the inspection is based on analysis of the product by users, but it can involve designers and experts as well.

Examples of testing are the remote testing and the think aloud protocol, examples of inquiry are the observation, focus groups, interviews, data logging, and questionnaires, and examples of inspection are the Heuristic Evaluation, Cognitive and Pluralistic Walkthroughs and Formal Usability Inspection (Ivory, 2003, URL_3.3).

Heuristic Evaluation is using experts that assess the interface of an application, utilizing rules of thumb method for the guidance. Heuristic evaluation is generally an easy to implement, and able to find a respectable amount of usability problems in short time and without the need of using prototypes or real users. Usually 3 to 5 experts are needed for the evaluation, who must be very experienced though, in order to produce the expected good results (Po et al., 2004).

Cognitive Walkthrough is using experts too, utilizing checklists and use scenarios in order to investigate the usability of an interface, focusing on the learnability attribute of usability. A set of cognitive procedure–based criteria are used in order to enable the expert to evaluate the predefined tasks together with the interface of the application. Generally, in this method the expert is virtually performing the actions that a real user would perform during the execution of a task. It is a structured method of analysis of user interface; it is used in the early stages of a design cycle, it is not difficult in learning and application, and its findings are valuable. As disadvantages of this method we can include the validation necessity, the possibility of long sessions during complicated tasks, the subjectivity of the findings and the requirement of approach to the real end users of the interface tested. It is also require high skilled analysts (Stanton et al., 2005).

Pluralistic Walkthrough is the cognitive Walkthrough involving group meetings of people, consist of developers of an application, users and usability experts. Amongst the advantages of this method is the easiness of learning, iterative testing capability, involvement and therefore feedback of all the participatory parties. One of the disadvantages is the difficulty of finding proper task-based contexts for the usability testing.

Empirical methods are a standard for usability testing, using data empirically collected form users working with the real application or product. Usually it is not easy to conduct empirical methods, as it is generally expensive and time consuming and it can pose some critical problems, regarding the validity of the results. Field studies can be used in order to warrant validity; nevertheless the results are of qualitative nature and need high expertise in order to be interpreted correctly (Rosson & Carroll, 2002).

An overview of the classic methods of usability testing, including their advantages and disadvantages is presented in table 3.1. All of these methods can include different data collection methods (tools).

Method	**Advantages**	**Disadvantages**
Heuristic Evaluation	Easiness of learning, inexpensiveness of implementation, early development stage problem identification	Necessity of debriefing session towards finding ways of fixing the problems
Cognitive Walkthrough	Ability in diagnosing problems, investigating cognitive processes	Requires a trained skilled practitioner, focuses on one usability attribute
Pluralistic Walkthrough	Easiness of learning, iterative testing capability, satisfying criteria of all test participatory parties	Difficulty in finding proper task-based context for usability testing
Empirical Methods	Capability of finding the reasons and results, answering particular questions	Expensive and time consuming, requires a trained skilful practitioner

Formal Usability Inspection	Ability of representing different domains of knowledge, getting usability problems and solutions list, evaluating both knowledge processing and behavioural tasks	Difficulty in finding proper task-based context for usability testing, actually no participation of end users
Formal Design Analysis	Ability of finding early stage problems, comparing different usability design	Difficulty of learning and using, applicable to only expert behaviour analysis

Table 3.1 Different usability testing methods

Usability testing by real users is usually more expensive approach in terms of time and money, but proves to give more valid results about real-world tasks (Nivala, 2005). Redish & Dumas (1999), outline usability testing through five main features:

- Improvement of usability of a product is the main goal. This can be divided into several sub-objectives and questions of each test, that are defined from the start of the project
- The real users are actual representatives of the participants of the tests
- Real tasks are done by the participants
- Users' reactions and verbal expressions are observed
- Data analysis, investigation of actual problems and recommendations about correction modifications are taking place

Every user evaluation method has some advantages as well as some disadvantages. According to Kumar (2005), the particular methods that will be used for a collection of data should be selected based on some important criteria. These are the mission of a study, the available resources and the researcher's capabilities. This means that although a method could fit very well in an evaluative research, it couldn't be used due to practical limitations, affecting the data as it comes to quality issues. A very important issue regarding the selection of methods is the nature of the participant itself. For example, social, economic and demographical variables, like the previous experience with similar product designs, his/her level of education, age and gender, can determine the use of questionnaires, interviews or observation as appropriate methods (Brink, Gergle & Wood, 2002). It is also important that the purpose of the testing is clearly and comprehensibly illustrated to the participants, especially when the participant has to carry it out alone, and without the ability of getting further explanations from the researcher. This is the case in questionnaires method.

But let's have a look at the main data collection methods and recording techniques that are used generally in usability testing, while they have been adopted from the general research methodology theories.

A. Data collection methods:

- Observation's main purpose is to investigate the actual interaction of the participant with the tested system when this takes place. It can uncover specific issues that couldn't be investigated through others methods, and especially in cases where the participants are not willing or it is difficult for them to express their opinions while they are interacting. It is a meaningful and analytical way of collecting data, and can be separated into direct, covert, and participant observation (Salmon & Stanton, 2005). In the first case the researcher passively supervises them and collects the interaction facts, in the second s/he does the same but without being visible to the participants, and in the third s/he is actively involved in the tasks done by the participants being a part of them. The basic problem that this method may pose is that it can be intrusive, as the participants' performance may change due to the fact that they are aware of being observed. This factor makes this method not very convenient for field-based testing (Jordan et al., 1996).However, the application of proper technical solutions, where the observer stays somehow "invisible" to the participants can reduce his/her influence to them. Observation can be applied to throughout the usability process.

- Interview is a method that involves the verbal interaction between the researcher and the participants for a specific intention, where the former keeps a record of the process output. It can be structured, following a strict predefined frame of construction, content and questions, unstructured, which a more liberal type, allowing the researcher to modify any part of the process according to his/her will or semi-structured, which is a combination of the other two. The structured interviews are easier for analysis, as through their equable data results they allow for immediate comparative study. In-depth and Focus Group interviews are alternatives belong to the first category. The focus group involves a group of participants, while the second a single one each time.

  An interview is generally capable of investigating more deep information, it can be combined with other methods like Observation easily, in order to further support the results of a research, and there is the ability of clarification from the side of the researcher about possible misconceptions.

  In the disadvantages of interview are amongst others the higher resources and time requirements, the affection of data quality by the Interview's quality and the researcher's personal profile, and the possible appearance of bias affecting both the interviewer and the participant. Interview is used after the completion of the practical part of usability testing of the product.

- A questionnaire comprises a textual form of questions, given to participants in order to answer them after they read them carefully. In this case there is no physical interaction of the participant with the researcher, so it is very important for the questions to be given in a way that they are clear and understandable from the participants. It is very much dependent on the skills of the researcher to make the structure of it similar to the flow of a verbal interview, and desirable to be answered.

  A questionnaire can be provided by mail, collective or administrative forms. A questionnaire by mail is a mailed envelope that contains the questionnaire together with a cover letter that gives all the essential directions to the respondent. A usual shortcoming of this method is the low rate of response. A collective questionnaire is using a population that is gathered together, such as students in a university or people attending a social event. The advantage here is that it is a fast and voluminous responding method, giving to the researcher the chance to explain by him/herself every possible issue of it to the respondents. And finally, an administrative form is conducted in the public, involving occasional people from whom the ones that fall inside the population criteria of the research are selected for questioning.

A questionnaire is generally an inexpensive method, better preserving the privacy of the participants and easier to administer, since the questions are answered structurally. On its negative side, we can include the low response rate, it is possible to show a bias originated in the participants' reasons of responding, the inability of further explanation if questions arise, the possible alteration of initial answers by other questions or other respondents, and the incapability of collecting more information than what the answers of particular questions pose.

- The Think Aloud protocol concentrates on the investigation of what the users are actually thinking and doing when they are interacting with a design. It can be realized through a similar to a controlled experiment form, where the participant is given several tasks to carry out in specific order, or through a free exploration, where the researcher asks the participant to copy his actions in order to explore the design. The controlled one is good in discovering particular design errors, while the free one can highlight the parts of the design that are more preferred from the users.

  The Think Aloud protocol is proved to be an efficient method of collecting big amounts of data when the number of participants is small (Jordan, 1998). The main disadvantage that can arise from the use of this method is the possible conflict between the participant's task execution and verbal expression / description, as they are trying to do the first while at the same time they communicate the second. Two more problems that could be found is the effort of the participant to give a logic explanation to what s/he is doing even if it was accidental, or the opposite, to try to follow a logic framework that s/he verbally expressed at the beginning, showing s/he is logically consistent. The researcher has to be skilful in balancing the level of intervention to the participant, avoiding both the lack of data collection and biasing of him/her.

B. Data recording techniques:

- Video recording / analysis is aimed at capturing the participants' interaction with a particular design during testing, which can be done in the laboratory or in the field and maintained for further analysis. Efficient analysis is a very important part, as it determines valuable this method can be. One of these is the Diagnostic Recorder for Usability Measurement (DRUM) (Macleod & Rengger, 1993, as cited in Jordan et al., 1996). This is a special software program for the support of researchers in organizing and analyzing video-recorded usability testing observations in order to reduce the time needed and improve the investigation results. Camera alignment issues for best visibility of the participants' interactions and time consuming procedure are some of the disadvantages of this method.

- Data logging is another technique, used for the recording of timing and type of incidents that happen during the evaluation process of a design, mainly applied to computer-based environments, as for example computer interfaces and software evaluations. Data analysis through Data logging can be up to eight times more efficient than other methods, depending on the context of use (Hammontree, 1992) The required time for the completion of a task, or the errors found during the testing of a design are some of the objective variables that can be efficiently investigated through this method, with the addition of easiness and rapidity of producing graphical representations of the collected data. Disadvantages of this method are the possibility of problems issued by the excessive amounts of data to be analyzed, possible violations of participants' security and privacy matters, and possible need for special logging software customization in order to be used by a particular hardware system.

The different data collection / recording methods are shown in table 3.2 and their comparison in table 3.3.

Technique	Description
Observation	User's behaviour is observed through the whole usability testing process
Interview	User's verbal report is collected using interviews after the completion of the usability testing process
Questionnaire	User's opinions and attitude towards the usability of application are collected through question items
Think Aloud protocol	User's thoughts are collected through his/her audible expression during the usability testing process
Video Recording / analysis	User interaction data is captured via video recording through the usability testing process
Data-Logging	User's actions tracking is collected via data logging software applications during the usability testing process

Table 3.2 Different data collection / recording methods

Technique	Advantages	Disadvantages
Observation	It captures actual user behaviour, and not what s/he thinks that is doing. It can later be compared it with other data collection methods for validity check, inexpensive method	Cannot capture insights about what the person is really thinking or the reasons behind a particular behaviour or comment
Interview	Questioning level can be changed to meet the context, interesting matters can be investigated as they appear	Analysis of results can be problematic, researcher style and personality may affect the response of the participant

Questionnaire	Fast, less expensive, can be easily analyzed, offers better anonymity, can be used throughout the design process	Inability of questions clarification, less flexibility than interview, as questions are decided and typed in advance, lack of probing possible
Think Aloud	Reveals the reasons behind user actions, simultaneous collection of preference and performance information, less expensive method	It is often unnatural and distracting to the user
Video Analysis	Immediate view of the actions of the participant, capturing of face expressions possible	Difficulties in analyzing the data, difficulties in right alignment of the camera for best view
Data Logging	No obstructive to the participant, it can investigate real-world actions in real environments	It maybe need special software, it cannot capture the participant's face expressions

Table 3.3 Comparison of Different data collection / recording methods

The validity of the collected data is a very important parameter, which determines the quality of research methods and techniques based on the relation of the research results with the actual research questions. It assesses the consistency of the research, considering the selection made regarding the target population characteristics and its sample number, the types of methods used for the collection of the data, the way of processing the data, the statistical analysis techniques used, and the final conclusions presented. The research method should measure what it is designed to measure, something that can pose problems in case of qualitative methods that assess for example the efficiency of an application, where subjectivity of results could be observed (Kumar, 2005).

Reliability is another issue that can determine the quality of a research. Stability, consistency and predictability of the research methodology are reliability's main characteristics. Reliability can generally be assessed by the ability of the research to be reproducible, providing the same results when the same research methods are applied.

In case of usability testing, where the research is always, though not necessarily completely, of qualitative nature, the validity cannot be generally statistically verified, as it is mostly insight-driven. The number of participants for example, does not have to be as large as what is used in quantitative approaches. Faulkner (2004), states that this number should be generally between fifteen and fifty, depending on the method used, and she is giving theoretical background and research evidence for this. It is proven that even small numbers of participants can still produce good validity results (URL_3.4).

3.6. Usability testing and mobile applications

While in desktop PC applications there are a lot of already established usability testing HCI-based methodologies, these are not all the time suitable for the mobile ones (Jones et al, 1999, cited in Zhang & Adipat, 2005). In a UCD approach there is a need for usability testing with the end users ad infinitum during the whole design cycle, as they are the ones that definitely judge the design. Laboratory is the classic environment that this testing is conducted, actually being detached from the concept of mobility (Isomursu, Kuutti & Väinämö, 2004). This happens because of the different environment of mobile applications involving the mobility of the users together with the real context of use. In order to apply real contexts into laboratory usability testing, the data collection methods should be specifically designed.

The difficulty in conducting usability testing for mobile applications is the unpredictable conditions in which a user can be present when using them in real life, including walking on a street, driving, standing still or meeting different lighting environments. Thus usability testing for mobile geo-applications often includes only a few of the aspects that can be applied, even when it is taking place in the field with the use of the actual mobile devices. Mobile context ignorance and inadequate control of testing procedures are some of the most usual confronted shortcomings. Consequently it is very important to establish substantial methods for usability testing in mobile geo-applications (Zhang & Adipat, 2005).

Mobile applications' evaluation is generally a difficult activity. New methods and techniques have to be implemented in order to counteract the problems that are faced in the laboratory testing, such as the lack of the real context, as well as in the field, where mobility generates controllability and validity issues. Using video recording as a recording method in the field, is for example an easy approach, but poses user disturbance issues. However, it is still useful when the target is a more deep investigation rather than a massive user testing. Various other techniques for field usability testing have been investigated, including click sequence data logs, voice recording on the mobile device or installing a micro camera on it (Kjeldskov et al., 2005).

Laboratory testing has proven to be more effective than field-based one, especially when strictly user interface and navigation issues have to be investigated. However, when contextual factors have to be studied, a field testing, studying the behaviour of user in the real environment is an appropriate approach (Kaikkoken et al., 2005). Until there is a better understanding of the mobile environment and use, so that these can be better recreated in the laboratory, data collection in the field should keep on being involved in the usability testing (Kjeldskov et al., 2005).

3.7. Examples

In this part we will investigate some characteristic examples of research that have been conducted in mobile geo-applications design and implementation projects, presenting the goals of these projects, the techniques that they utilized and the results that they achieved. Additionally, a look upon the usability testing that was applied to them will be given, in order to obtain a general idea about the magnitude and necessity of this research.

3.7.1. The GiMoDig project

The GiMoDig project (Geospatial info-Mobility service by real-time Data-integration and generalization), was a project conducted by several European universities, where the Department of Geoinformatics and Cartography of the Finnish Geodetic Institute was the coordinator. The objective was to develop and evaluate guidelines for appropriate mobile map representations, by applying appropriate real-time generalization and data-integration techniques. A prototype mobile geo-application was developed in order to conduct this research, which would be able to serve the mobile users with geo-information provided by geographic databases of national mapping agencies (NMAs).

The mapping representations should be dynamically changeable according to the different use cases and users' preferences and requirements, while the spatial data in vector format could be retrieved through a wireless connection to a remote map server. Adaptation of the map interface according to the context of use was one of the essential methods that were applied. The mobile map should be legible regardless the zoom level and easily understandable from the users, while visualization aspects such as the symbols of objects, the colours used and the text displayed should follow rules that were created through the realization of the research and based on cartographic theory.

Different use cases were initially developed in order to study the requirements from the final product, which were then converted into four use cases. These were a hiker's outdoor activities in a national park, a fire exigency in the University of Hannover, a cyclict's crossing of the Danish-German borders, and an expert's use of the system (Nivala & Sarjakoski, 2005) (Figure 2.9).

The project final results showed that the framework that was implemented for mobile geo-applications generally met the initial requirements, although it still cannot meet the requirements of particular types of applications, as for example the ones that involve route planning or thematic data association. Several advanced functionalities, such as context aware adaptation of maps, dynamic icon placement and text placement and POI integration, were implemented, although they were not included in the initial plans, giving a hint for further future research. The investigation of real-time generalization methods' necessity as part of the context-based dynamic adaptation abilities of mobile maps, showed that such types of innovative functions need further research, as the current users are not yet familiar with them. Even so, a global system philosophy that is carefully applied can produce user-friendly and non-complex designs.

The system's web-based framework for utilizing different national topographic databases, although it is working efficiently as it is parameterized now, it showed that it would be greatly benefit from a standardization of the topographic databases of NMAs of the different (European) countries according to the global specifications that were proposed, and especially from the application of a common coordinate system. The need for a central administrative system for services like the GiMoDig, that would take care of up-to-date the databases of the service in interrelation with the several source national databases was also mentioned.

In GiMoDig, usability testing was applied already from the beginning of the project, starting with a pilot user test in the field with existing hardware and software, in order to discover several usability issues connected to them. There wasn't any prototype of GiMoDig available at that stage, and that was the reason for using relevant existing applications, which they were not assessed for their own usability, but used as a mean to test the current mapping interfaces and visualizations. The usability test was conducted with the attendance of usability experts, in order to find problems regarding the software functionality, the quality of cartographic representation, and the usability of maps in mobile devices in general (Sarjakoski & Sarjakoski, 2005).

Amongst the problems that were found, was the inadequate use of cartographic symbols, the inefficient visualization of the maps in the small display, and the luck of consistency in zooming between different scales. Next was the identification of requirements for the new design based on the results of the surveys.

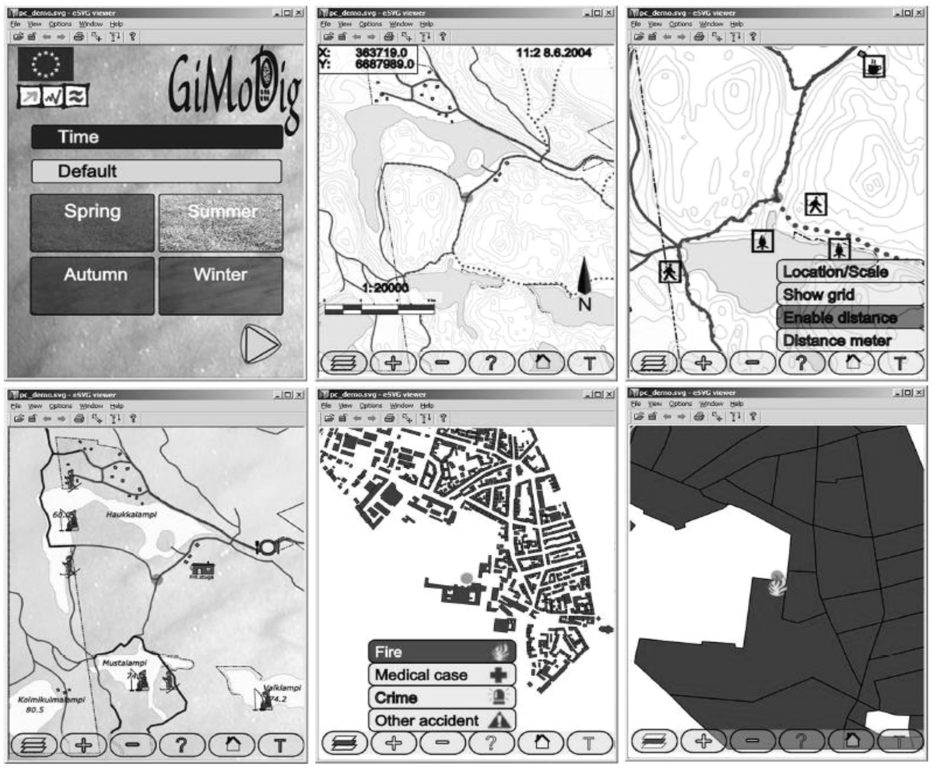


Figure 3.4 Examples of different visualizations in GiMoDig project, showing the adaptation of the maps according to the particular context of use and user preferences (Source: Nivala & Sarjakoski, 2005)

The project followed an iterative UCD-based evaluation approach, including assessment of the prototypes that were later implemented, regarding the satisfaction of the requirements and usability criteria that were initially set. These were the easiness of the user interface, suitability of cartographic presentation, integration of various representations, and context awareness of the maps according to a predefined group of contexts of use. These were categorized as computing, user, physical context, time, and history contexts.

Heuristic and expert evaluations together with usability and intuitivity testing were the methods conducted for the usability testing of the project, while no real context user field surveys took place. The implemented application should be suitable for use in laptop as well as PDA environments, utilizing easy to understand self-explained feature pictograms, combining different topographic datasets and adapting according to different contexts of use.

GiMoDig generally followed an efficient methodology for the design, implementation and testing of the software application. We could say that the addition of an end users' usability evaluation of the final prototypes in the field in a real context of use would be beneficiary, combined with the laboratory testing that was part of the iterative design approach of the project.

Actually, the final usability testing was done with potential users; consisting of private companies' members. This population was met during six group meetings, discussing the whole project's achievements regarding the system and the different interfaces, the context aware adaptation of maps, the use of different datasets, the technologies of mobile devices, and applications with the best potential to achieve commercial success.

Some additional usability issues could be investigated, that the laboratory setting is hard to find due to lack of a real testing environment, if field-based real user testing was applied. The high value of field testing for context aware mobile geo-applications was already indicated by several researchers such as Mennecke & Strader (2003), Leitner, Plattner & Hitz (2006) and Kaikokken et al. (2005).

In this research we will conduct different field-based usability testing methods and techniques in order to check whether and how this testing can provide additional and valuable information that would be probably impossible to find during laboratory-based evaluations.

3.7.2. The Trammate project

Trammate, a participatory project involving Aalborg University of Denmark, The University of Melbourne and Novell Company, was initiated as part of a PhD research, intended to build a context aware mobile guidance and route-planning system, dedicated to users of public transportation in Australia, in the city of Melbourne (URL_3.5). The main target group of users of the system was comprised of enterprise employees that are on the need of gathering meetings in different places of the city. Inside the Trammate project, the development of a unique user interface including map and textual elements was realized, having the target of efficiently eliminating the information represented, based on contextual factors, and increasing in this way the user-friendliness.

The prototype realization was done by the Department of Geomatics of the University of Melbourne, and offered the capability of route planning, considering the location of the user which was acquired by a GPS receiver. The interface was a combination of mapping and text information, and the application was running on an IPAQ PDA with integrated Wireless Application Protocol (WAP) client browser that could get data from a server through a GPRS connection. Three were the main functions of the system, as of displaying the tram timetables, planning a trip, and overviewing the route. Route calculation was taking into account the location of the user, the selected destination, the number of stops of the trams until the target, and their time schedule (Kjeldskov et al., 2005).

In the Trammate project, the initial requirements were identified and divided into four functions, including the supply of travel information, route planning and alerting. The project doesn't seem to have followed a specific methodology of design and implementation, and is more focused on the comparison of different usability and data collection methods in order to assess the usability of the implemented prototype. These are the field and laboratory evaluation, Heuristic Walkthrough and Rapid reflection. Actually, the comparison between laboratory and field usability testing was the inspiration behind the development of the Trammate project.

In the field testing, four tasks that represented real-world situations were given to four users for completion in order to assess the usability while they were mobile, while the GPRS connection was actually a simulated and not real-time one, due to connectivity issues that were difficult to overpass during the evaluations. In the laboratory, similar tasks were given but in static mode this time. The heuristic walkthrough was done by four expert evaluators and the rapid reflection included all the participants.

Among the results of the project was the finding that there are advantages in all the four usability testing methods that were used, depending on the type of usability problems that have to be investigated. Most of the problems that were identified were associated with the map representation, structure and accuracy, as well as text-based information amount and quality. Each method was able to discover exclusive usability problems; therefore a combination of these methods gave the best results. Especially as it comes to the comparison between laboratory and field testing, it is argued that, while the former allows for a very detailed investigation of interface and representation problems, the latter can additionally uncover problems of use in real mobile environments.

The study concludes that there is still a strong need for more research in usability testing for mobile applications, and especially regarding field survey methods, as laboratory testing cannot, at least until now, be established in a way that allows for adequate exploration of real usage situations.

Figure 3.5 Map interface of the Trammate application, laboratory, and field evaluation (Source: Kjeldskov et al., 2005)

An issue regarding the Trammate field-based usability testing that is taken into account in this research, is that there were three researchers during the evaluations, one managing the sessions, the second for video recording the participants, and the third for taking notes. So it is obvious that the human resources used were high, making the duplication of the experiment difficult and costly. The laboratory testing again involved three researchers, while it would be easier and more efficient to use automated procedures.

Additionally, the research concludes that there is not so much difference between the results gathered by using the four different methods. But it does not mention the possible influence of the researchers on the participants' behaviour, as their number is big and could easily pose a bias that would alter the current findings. Moreover, the Trammate's researchers argue that their research is not intended to be regarded as statistically strong, and is more focusing in general conclusions and critics about the different usability methods applied.

In this research, we will try to find technical solutions in order to reduce the number of researchers needed for usability testing in the field, while at the same time the influence of the researchers on the participants should remain the lowest possible.

3.7.3. The TIP system

TIP comprises a mobile tourist information system, which utilizes the location, interests and travel history data of its users in order to provide them with recommendations about tourist routes and information about sights according to Hinze & Buchanan (2005). The system consists of two basic interconnected processing elements: that of location identifier and that of filtering of data sources. The latter is processing the information available from various databases in order to select what will be provided to the user according to the context of use. Depending on the personal background, preferences and previous activities of the user, as for example visiting specific types of historical buildings, relevant information and suggestions is been given to him/her, making the system a highly personalized tourist decision support assistant. The user interface is map and web browser-based (fig. 2.12).

The intention behind the development of TIP was to create a system that would support different application implementations, through its context awareness and adaptability. In order to achieve this, a context management model was applied, including model, observe, store and access aspects. Several issues, including the interactivity patterns between the user and the service or the objects of interest depending on the type of mobile geo-application, methods for efficient manual or automatic observation of user context, efficiency, security, and privacy of data storing were investigated throughout the design and development of TIP.

Among the findings of the research through which TIP was implemented was the difficulty to satisfy the user's need for clearly knowing the reasons behind system's decision into providing him/her with particular information without at the same time adding more complication in the user-application interaction process. Consistency of content and presentation in interactive mobile geo-applications was another issue that needs more investigation and evaluation, as current approaches of these applications, often offering context-based adaptivity in order to overcome the limitations of mobile device use, may produce confusion and reduce the performance of the user. Finally, unreliability of location sensing and network connection in mobile geo-applications can lead into a disconnection between the real context of use and the one that is perceived by the system, resulting in false information provided to the user.

There is no user testing in real conditions mentioned in the TIP research, which could be very useful in order to investigate usability problems as they arise in real contexts of use of the application.

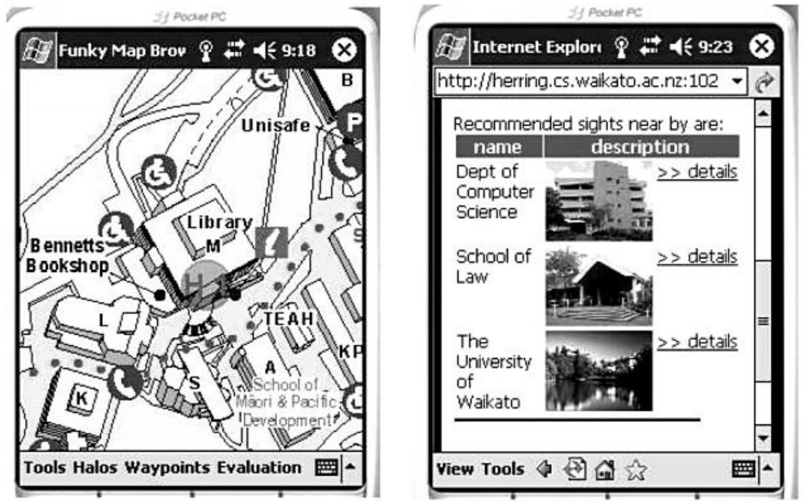


Figure 3.6 Screenshots of TIP showing the mapping interface and the internet explorer-based one (Source: Hinze & Buchanan, 2005)

3.8. Conclusion

In this chapter, a broad overview of usability methods and techniques used for the usability testing was given, together with an evaluation of the testing methods applied to examples of mobile geo-applications. It is clear that there are two basic methodologies for usability testing, the laboratory-based and the field-based. Each one has its pluses and minuses, depending on the type of application and the type of functions that will be assessed, together with the context of use. The combination of laboratory and field testing seems to be the best solution as it comes to deep investigation of different usability problems that cannot be found with each approach alone. Based on what was found while studying recent examples of mobile geo-applications for implementation or already implemented, there is still a need for research on usability testing methods and techniques.

Especially field-based usability testing is lacking thorough research, as it is more difficult to be carried out due to reasons such as real environmental constraints, difficulties of controlling the experiments and difficulties in applying several data collection methods, as for example voice recording, which can be heavily disturbed by the high background noise. The need for building a framework for field testing using methods and tools that can overcome the shortcomings of current approaches is obvious.

In chapter four, we will try to combine all the knowledge that was gained through the research work of the second and third chapter, in order to build and propose an appropriate usability testing methodology that can be used for the evaluation of the UWSM2 project prototypes. This methodology can cover the part of field testing of mobile geo-application prototypes and particularly the ones that use map interfaces that utilize map scaling functionalities, such as the UWSM2 project. Modifications and adjustments can later be applied to this methodology in order to meet the requirements of various mobile geo-applications' field-based usability testing.

4. Research methodology and user tests

4.1. Introduction

In this chapter we will combine all the knowledge gained through the literature review, together with the application of new ideas, in order to build a methodology for usability testing of mobile geo-applications, focusing in the field-based evaluation of the prototypes of the UWSM2 project.

The chapter starts with a description of the UWSM2 use cases, tasks that the user of this system is performing and general questions that the system intends to answer. The description of what are the goals of the proposed usability testing methodology is then given, including the variables of usability that will be assessed and the part of the project's functionalities that will be investigated in this research. The advantages of the selected methods are discussed, and the possibilities of reduction of some of the problems that are usually faced, through the use of advanced techniques, are presented. As a part of this approach, a unique system that was developed in order to support the application of the proposed methodology is then outlined, justifying the reasons behind its conception. Finally, the criteria-based selections of testing participants and applications similar to UWSM2 are described and directions for the next part of this research are given.

Generally, in order to build a usability testing methodology, first the goals of this testing should be determined. What are the parts or functions of the system under investigation that have to be measured for their usability? What variables of usability will be measured? What methods and techniques should be used in order to measure these variables and where will the surveys take place? What should be the number of participants and based on what criteria they should be selected? The answers to these critical questions enable the researcher to build the testing methodology and adjust it to the specific needs that are figured out during its development, based on facts and problems that are confronted throughout this procedure.

A general framework for the design of a usability testing methodology for mobile applications can be seen in figure 4.1. This framework divides the building process of a usability testing methodology in four stages. In the first one, the research questions and objectives are set, and the decision whether the experiment should take place in the field, in real contexts of use, or in the laboratory is made. In the next stage, the tools that will be used for the testing, being actual devices or simulations are selected according to the previous stage. After that, the variables of usability that have to be tested are decided and the way that these variables will be actually measured is identified. Finally, the methods and techniques that are aimed at measuring the previously defined usability variables are set and the setting of the survey is made. The paths that were followed during this research in order to build the proposed usability testing methodology are represented in the darker boxes.

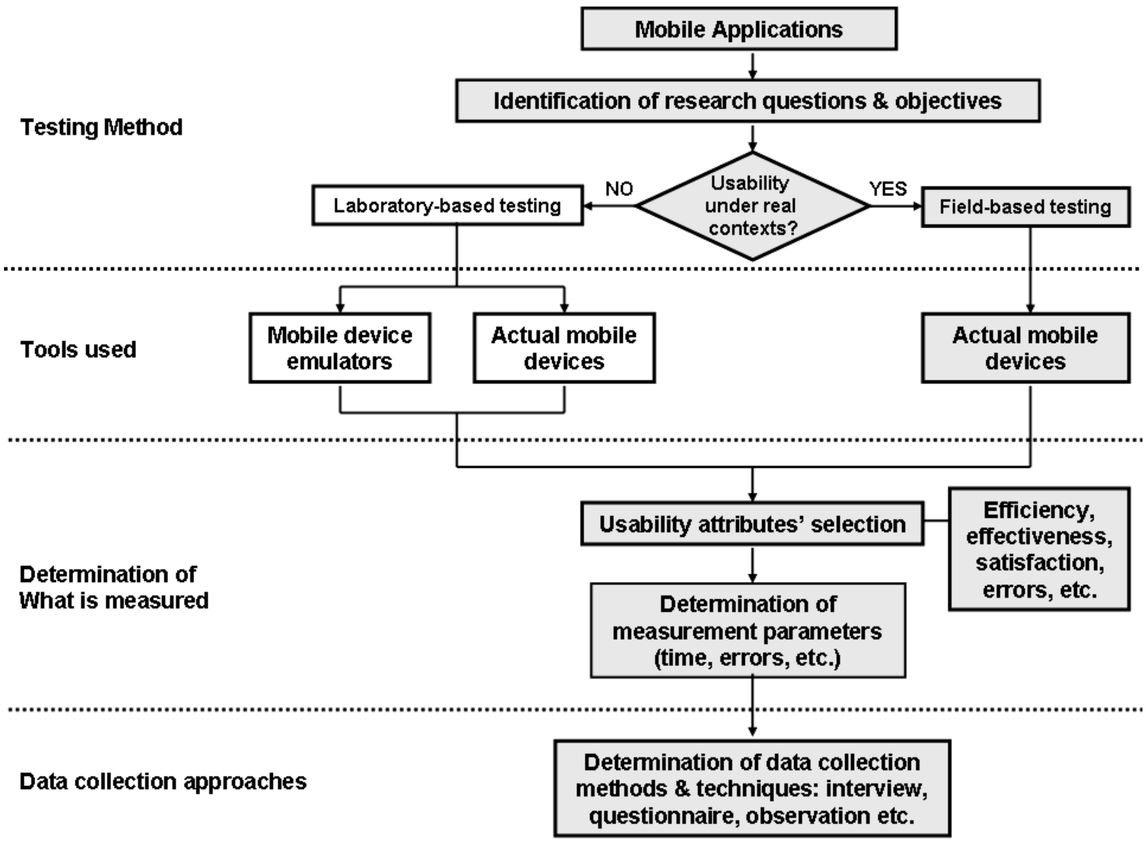


Figure 4.1 A General framework for selection of usability testing methodologies (adopted and adjusted from Zhang & Adipat, 2005)

4.2. UWSM2 use cases and related questions to be answered

As this research intends to build a methodology that can be used for the usability testing of the prototypes of the UWSM2 project, the methodology that will be proposed here should be able to investigate usability issues related to the application functions that this project utilizes, and particularly that of appropriate map scaling. UWSM2 focuses, as its name reveals, to the supply of well-scaled mobile maps to costumers. However, it is out of the scope of this study to deal with the special processes that will run between the mobile server and the mobile device / application that will be implemented by UWSM2, such as generalization and progressive data transfer in order to achieve map representations at different scales on the mobile device's display, described in Van Oosterom, De Vries & Meijers (2006). What the user sees in the mobile display and how can s/he can manipulate this information in a usable way is the subject of the present research.

Before the design phase of the USWM2 project, two use cases were defined, involving users of the imagined system in real world scenarios in specific contexts of use. The first use case, named "service to citizens" takes advantage of the very detailed topographic maps maintained by the municipality of Amsterdam in order to provide the citizens of the city with location-based services, such as parking information to car drivers. Information regarding availability, price, and distance from the current driver's location is presented on the mobile device's display, using maps at different scales and levels of detail, allowing him/her also to save the location of the car and assisting him/her to navigate through the city. A hypothetical scenario for this use case, proposed by UWSM2 researchers, is involving a woman who wants to visit the Rijksmuseum in Amsterdam, wants to find parking for her car in order to go to the museum and then to find her car again and to leave Amsterdam (please refer to Appendix A).

The second use case of UWSM2, originated from ANWB, which is the Royal Dutch Touring Club (URL_4.1), is named "mobile tourist information", aiming at supporting the decision making of a tourist traveling in the Netherlands and using a mobile device. This resembles a mobile tourist guide, utilizing geographic positioning technologies, such as GPS, together with different sources of information, such as traffic, map, transportation, tourist attractions-related, and can supply the user with paper-like maps that can efficiently utilize the small screen of the mobile device. Fast and continuous zooming and panning functions, without interruptions, routing capability and different levels of detail in the maps are the fundamental features that this guide should include.

The two use cases that are presented above will be integrated in one mobile geo-application that will be able to support both user groups and requirements. As we can see, the main tasks for the user of such a system, generally, and according to this research's findings, are:

- To understand his/her geo-location in relation to the surrounding environment, through the use of a mobile map, adjusted according to the data from a geographic position acquisition system (e.g. GPS receiver, GSM cell positioning)
- To search for Points of Interest (POIs) that best fit his/her needs, where the response of the system is giving the geo-location in the map together with the closest street name / number
- To retrieve online location-based information such as weather forecasts and ticket prices.
- To get a presentation of a calculated route from his/her current location or a manually defined one to a selected POI on the mobile map together with information about it (e.g. distance, time of travel needed)
- To get navigation assistance through several information on the map display (e.g. names of streets, one-way roads)
- To get navigation assistance through graphical symbols

These tasks can be transformed into various related questions that the system is aimed at answering in order to support the user's traveling decision making. These questions include amongst others: mobile interface, mapping, data input, personalization and online information functions. Some of the more abstract geographic questions that the map in this case helps answering are:

- "Where am I?"
- "What is in front of me?"
- "Which and where in the map is the closest appropriate POI for what I need now?"
- "What is the distance and time needed to go that POI depending on the media I use (foot, vehicle, bicycle)

- "Which route should I follow?"

We focus on the map scaling features, which is the core of UWSM2 research, and would then issue questions as the following:

- Is the map representation clear, consistent and comprehensive enough at different scales?
- Is it easy for the user to understand his/her position on the map in relevance to the real surrounding environment in all zoom levels?
- Are the zooming / panning / rotating / overviewing functions helpful for the better understanding of the map / route?

These geographic and map scaling-related questions together with the previously defined tasks will be used in order to build the methodology to assess the usability of a mobile map scaling function, and to select a geo-application relevant to the UWSM2 that can be used to test this methodology.

4.3. Variables of usability to be assessed

The three variables of usability are efficiency, effectiveness and user satisfaction. In order to practically assess these variables and given the fact that the testing will be field-based, appropriate methods and techniques have to be selected. UWSM2 project's "well" scaled maps refer to appropriate map scaling in specific situations. That means that what is "well" depends on the particular use and user context, which have a great impact on the usability of the application that will be developed for that project.

Interacting with a mobile geo-application in the real world, as it is already discussed in the previous chapters, is a potentially different case than what happens inside controllable laboratory environments and field testing can give a great insight into the actual user-application interaction (Kaikkoken et al., 2005; Streefkerk et al., 2006; Zhang & Adipat, 2005; Kimber, Georgievski & Sharda 2005). Field-based testing is trying to more closely recreate the real contexts of use, although it is not possible to include every possible parameter, regarding the resources and time available for the testing together with the design and techniques used. So a well defined evaluation setting should be followed, involving amongst others a careful selection of participants, area for the testing procedure, predefined tasks to be done for specified data collection techniques, and an estimation of the duration of the testing process. The validity of the evaluation procedure should be deeply concerned too, as the setting of the experiment should meet the purpose of measuring what is initially designed to.

In order to select the method that will be used in this research, we studied the advantages and disadvantages of the usability methods and techniques that have been applied for the evaluation of mobile geo-applications as well as other applications, as presented in chapter 3. Additionally, ways to improve these techniques in order to better fit in the field-based testing environment were considered.

4.4. Methods and techniques that will be applied

The techniques that were chosen to be used in this research are observation, Think Aloud, video recording and semi-structured interviews. Observation was chosen because it can provide qualitative assessment of the context of use, especially as it comes to the real environment inside which the usability testing is taking place. The video recording is used to capture this interaction and give validity and evidence to the conducted research (Kimber, Georgievski & Sharda 2005). In order to assess this combination of techniques, three variations are studied, applied, applied into three groups of participants. Observation and semi-structured interviews are applied to all the groups of participants, as they are proposed as appropriate, while video recording and think aloud are used in three combinations. The setting of the techniques applied to the three groups is as follows (figure 4.2):

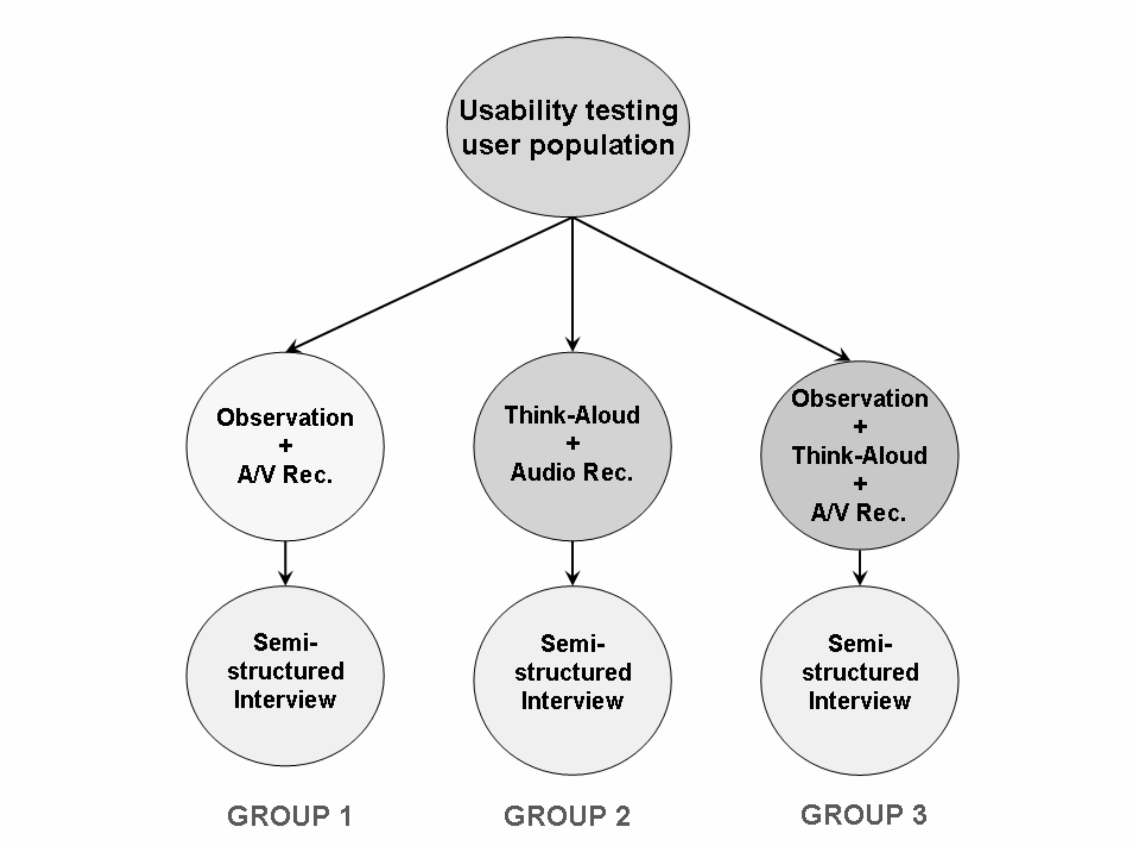


Figure 4.2 The combination of techniques used in each participants group

The participants are given several tasks to finish, which are designed to investigate the usability variables that are needed to be assessed, which are in this case, as it is already mentioned, the efficiency, effectiveness and user satisfaction through the use of a mobile geo-application involving map information.

One point that has to be carefully considered here is the least possible influence of the observer to the participant. It was also found, during the preliminary tests, that the use of a usual camcorder for the recording of the users is not the most efficient way, as then more human resources must be used for the testing, there are alignment problems, as the camcorder should continuously try to keep the subject inside the frame and only a part of the context of use can be recorded. It is also a more intrusive way, which makes the participants often feel uncomfortable with their monitoring, something that have impact in their results. The application of observation together with Think Aloud techniques in the field can pose an amount of problems, so new ways of conducting these techniques should be researched (Beck et al., 2003, cited in Zhang & Adipat, 2005). In classical usability testing, apart from the participant, two more persons are used; the observer and the facilitator.

Think Aloud is selected because it can give qualitative assessment of problems, strategies and user expectations identification (Jordan et al., 1996), while it uncovers a plethora of usability problems and the reasons why they happen. It is probably the most valuable of all the usability testing methods (Nielsen, 1993), requires a relatively small number of participants, from 3 to 5, and little experience from the facilitator (Andrews, 2002). Although it has also some disadvantages, like the difficulty to be applied in performance assessments, its very rich qualitative output from a small number of participants makes it a very good choice, if of course is used carefully. Think Aloud provides information that helps answering why participants do the things they do. Field-testing, reported in Chapter 5, will investigate this fact.

The semi-structured Interviews, as an indirect method, was chosen against questionnaires, as in its structured part it can investigate similar issues to what would be received through the questionnaires, and at the same time it allows, in its unstructured part, a more free interaction between the researcher and the participant. This unstructured interaction can uncover deeper thoughts of the user and help discovering new problems, ideas and suggestions. Although the risk of bias from the researcher to the participant is bigger here, the rich outcome of this method does reduce this risk's significance. Additionally, through interviews, misunderstandings regarding the questions of the researcher can be solved easily and straight, something that cannot happen with the questionnaires unless the participants are filling them while they are still next to the researcher. Audio recording is also used throughout the interviews, in order to enable a later confirmation of the results written in paper notes from the researcher.

The present research is experimenting with the use of the observer in the role of the facilitator too, through the use of special technologic supporting tools. The Observation and video recording of the participant together with Think Aloud protocol, as the proposed methodology suggests, is difficult to be conducted by one researcher only and classical techniques. For example, the use of a camcorder was considered to be difficult in the field, as the owner must continuously try to align it in order to have a proper view of the user / PDA display. The video recording through a camcorder could also influence the behaviour of the user, as the observer should stay very close to the user, making him/her feeling possibly uncomfortable. In order to overcome these issues, a special audio/video recording system was implemented, allowing the observer / facilitator to stay in a fair distance (10 – 25 meters) from the participant and remotely observe and communicate with him/her while s/he is carrying out the predefined tasks.

4.5. The implemented system for field-based usability testing

The system that was implemented initially, graphically represented in figure 4.3, comprises of three wide angle B/W mini cameras, one with integrated microphone, two pairs of video/audio transceivers, a video quad processor, a laptop, a handheld hard disk video/audio recorder, a pair of Private Mobile Radio (PMR) transceivers, a PDA, a Bluetooth GPS receiver and batteries.

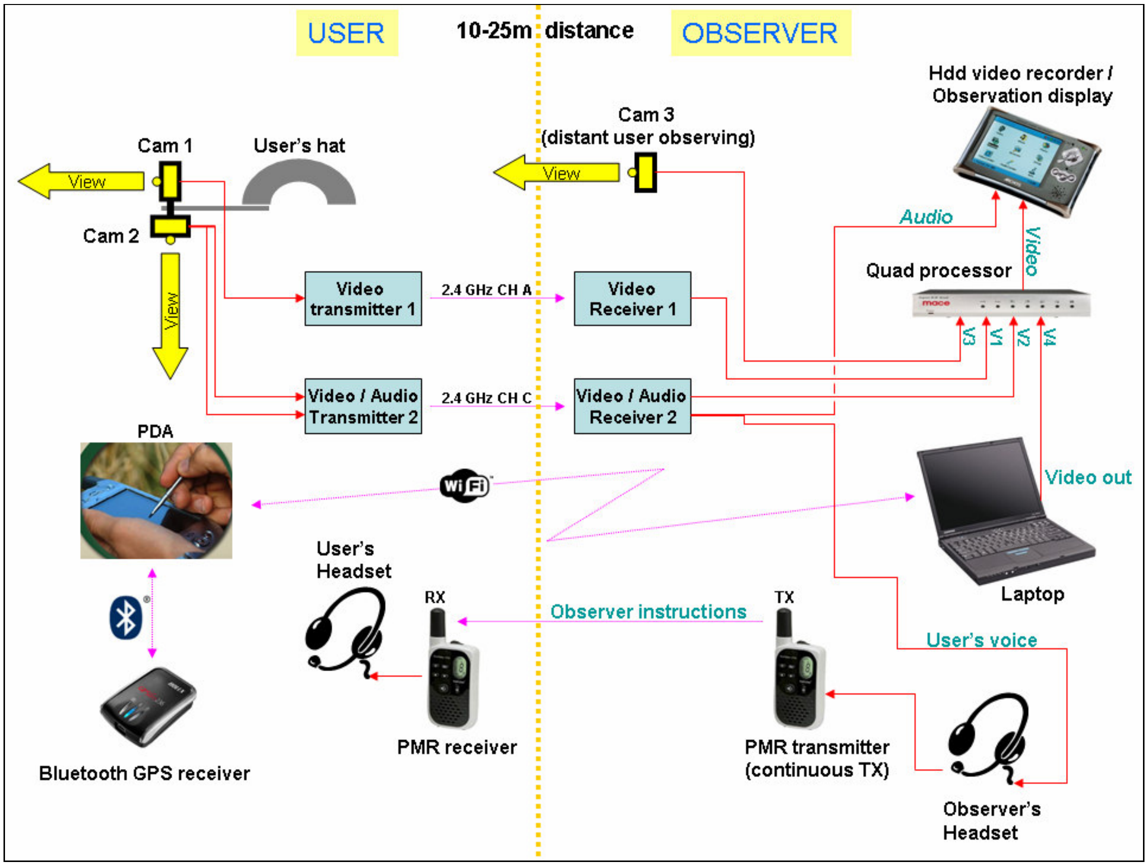


Figure 4.3 The diagram of the initially implemented field observation system

The first two B/W cameras are put on hat that the participant wears, capturing the environment in front of him/her and the movements on his / her hands interacting with the mobile device. The PDA, an Ipaq model hx4700, is the mobile device where the selected mobile geo-application is running and the test user is carrying. The video and audio signals from the user's cameras go to a pair of audio/video transmitters and then transmitted to a pair of receivers that the observer carries. The observer also carries the laptop, which is wirelessly connected to the user's PDA through a Wi-Fi connection and MS Windows ActiveSync software. Through this connection, and with the use of a remote controlling / capturing software, named Pocket Controller, a screen capture of the PDA's display is then transferred in almost real-time to the laptop, which has a video output capability (figure 4.4). This video output is connected to the video quad processor, together with the received two video signals from the remote user cameras, and the third camera that the observer carries and is capturing the user from a distance (figure 4.5).

Figure 4.4 Remote PDA screen capture on the laptop via Wi-Fi connection

Figure 4.5 The researcher as observer / facilitator and the user while being remotely captured

The output of the quad processor is an image including all the above four video signals in one screen, in a cross-like frame, together with actual date / time and a name on each image quadrant. This output is recorded from a pocket-sized Archos Av-400 40Gb hard disk video recorder, together with the audio signal that remotely captures the user's voice and comes from the one of the two audio/video receivers (figure 4.6).

Figure 4.6 Example of the quad-screen video recording of the system

The suitable laptop does not have to be a powerful one, as it just plays the role of an interface / converter between the digital screen capture of the PDA through Wi-Fi connection to a composite analogue video output. Small size, weight, and long lasting battery are the most important characteristics in this case. During the initial system tests, an old Pentium III Compaq Armada laptop running Windows XP was used, not only without any problem, but also giving very long working times without recharging, between 5 and 6 hours, which is important for a field testing.

While the observer receives the voice of the user through the A/V transmitter to his headset the observer's voice is transmitted to the user through a PMR radio transmitter, modified in a way that it is continuously in transmission mode, and not only when the transmission key is pressed.

Figure 4.7 The wireless communication headset used in all the groups and the two cameras installed on the user's hat

A PMR receiver at the user's side connected to a headset that s/he wears enables him/her to hear the voice of the observer when it is needed. The user's headset together with the hat that has the two mini cameras on it, described earlier, can be seen in figure 4.7.

This remote capturing / observing system was inspired by the research of Van Elzakker (2004), who used a combination of camera / video quad / video recorder in the laboratory in order to investigate the interaction of test users with a GIS for the exploration of geographic data. Kaikokken et al. (2005) had also used a remote observation / capturing / recording system in order to assess the usability of a data transferring application for mobile phones in real conditions. In that case, the user carried three remote cameras, one of which was capturing his/her face and the other the mobile phone display / keyboard.

For mobile mapping applications, the remote capturing of the display of the mobile device through Wi-fi that is applied in the present study was found to be more efficient as the screen is not covered by the hands of the user. Video recording of the face of the user is not also used, as the system is aimed at being less intrusive.

The initial idea was to use a mini handheld PC, like the Oqo 01+, which would play the role of the video quad processor and the recorder through a four-input USB video capturing device. The Oqo, being at the hands of the observer, could also easily remotely control the mobile geo-application running on the PDA, giving a lot of capabilities in task control and data logging. This solution would also be easier to carry and less complicated, but unfortunately at the time that this research was conducted, it was difficult, in financial and availability manners, to be applied.

4.6. Participants' group forming criteria

In a field-based usability testing, where the interaction of the actual users of a product with the product has to be investigated, a careful selection of participants has to be done. In the current study, these users should be representatives of the target population of the UWSM2 project. This population, generally, includes citizens of Amsterdam or tourists that travel in the Netherlands, that can be native or foreigners. Since this population was difficult to be conducted for the testing surveys, considering the limitations of the current study, a solution was found to be a good alternative. This solution was the use of ITC students, who, although have at least a basic knowledge of paper and digital maps, they represent at the same time tourists, being mostly foreigners. These students were highly available, could form homogeneous and thus comparable groups of users, and could be contacted easily and fast.

Taking into account that this study does not intend to test a mobile geo-application itself but the combination of different methods and techniques in order to adjust and support the proposed usability testing methodology which could be later used for the evaluation of the prototypes of UWSM2 project or any other mobile geo-application, some compromises could be made regarding the test setting. For example, using students of ITC would more acceptable as a solution, if the tests were done in an area unknown to them. As Streefkerk et al. (2006) argues, the use of students as participants in usability studies is acceptable when the design of an application is more common and it applies to a broader population.

In order to select the participants that would form the three user groups in this research, a pre-selection questionnaire was given to forty people, mostly students, asking questions about their age, gender, and knowledge and experience on GPS systems, paper and digital maps, mobile devices, specifically PDAs, and mobile navigation / mapping applications. They were also asked to rate their knowledge and practical experience, if they were answered to be present, inside a range of poor / modest / very good, and their familiarity with two selected areas outside the city of Enschede were they live, which were the villages of Lonneker and Glanerbrug. This questionnaire is presented in Appendix C

According to the answers to these questionnaires, the 3 groups of users, including 6 participants each, as this number is acceptable for the core technique of the proposed methodology, which is the Think Aloud, were formed. In fact, Nielsen (1994) argues that the use of three to 5 participants, and not more than 6 or 7, for a think-aloud research, is a good balance between cost and output validity of such a study. Six persons per group, leading to a final total number of eighteen participants, was the maximum feasible number, considering the time and resources available for this research. Each group consisted of 4 men and 2 women, with different levels of knowledge and experience, especially in the field of mobile navigation systems and PDAs, aging from 25 to 40 years old. The groups selected to be homogeneous to each other, and the participants are one-to-one identical between each group, as for example people with excellent, moderate or no knowledge of mobile navigation systems are equally spread in the 3 groups. The area that most of the candidate participants were not familiar with was that of Lonneker, which was then selected for the field-based testing of the three user groups. The average ages for the 3 groups were 31.2 with a standard deviation of 4.88 years for group A, 31.7 with a standard deviation of 4.50 years for group B, and 30.3 with a standard deviation of 3.08 years for group C.

From now on the test participants of this study, for practical and confidential reasons, will be identified with a code that indicates their position inside the group forming table. This code consists of a letter from A to C and a number from 0 to 6, with the former indicating the group they belong to. The methods and techniques used in each of the group are presented in Section 4.4, in figure 4.2. The group forming table can be seen in table 4.1, where their similarity selection criteria are also shown. The number of crosses, from one to three, determines the knowledge / experience level in different fields, and is low, moderate and excellent accordingly.

Testing participants			**Knowledge / experience**				
Group 1	**Group 2**	**Group 3**	**GPS**	**Paper maps**	**Digital maps**	**PDA**	**Navigation**
A1	B1	C1	++	++	++	NO	+
A2	B2	C2	+++	+++	+++	+++	+++
A3	B3	C3	++	+++	+++	NO	NO
A4	B4	C4	++	++	++	++	+
A5	B5	C5	++	++	++	++	NO
A6	B6	C6	+++	+++	+++	+++	+++

Table 4.1 Testing participant groups forming

4.7. Similar geo-application selection criteria

In order to test the proposed methodology and find out which combination of usability testing methods and techniques give better results in a particular context of use that intends to approach the general UWSM2 context of use, a selection of a mobile geo-application, similar to UWSM2 project's one have to be done. This selection should be based on a series of criteria that are defined in order to help comparing different instances of software and selecting the one that best fits the testing environment conducted here. After studying the UWSM2 on-going research findings, and taking into account the limitations of this thesis, these criteria are:

- Ability to answer the abstract form of the geographic questions that UWSM2 is aimed at answering
- Zoom / pan / (manual) rotation functionalities during navigation
- Capability for non-automated zooming during navigation
- Movie-like zooming function preferable
- Different map detail in different scales
- Vario-scale map ability (desirable but not obligatory)
- Cost
- Availability
- Detailed coverage of the study area
- Compatibility with the mobile hardware platform used
- Clear map scale
- Easily changeable orientation of the map (north-up / heading-up)
- Non-overlapping of POIs
- Routing capability

UWSM2 intends, as it is discussed in earlier chapters, to use advanced techniques, such as on-line retrieval of maps from an up-to-date remote map server which applies real-time modification processes such as generalization to the geographic data. As we concentrate on the map scaling functions, we are not going to investigate the network communication influence on the usability of the application. So it is acceptable to use a mobile geo-application for the testing, which can utilize off-line mapping data, stored in local databases inside the mobile device's memory storage. The proposed usability testing methodology can be easily adapted and modified after the implementation of the first prototypes of UWSM2 in order to evaluate the usability of additional functions. Further modifications could be applied in the proposed methodology to make it suitable for the usability testing of different mobile geo-applications too.

The ability to answer the abstract form of the geographic questions that are posed inside UWSM2 usage scenarios is easy to be found in a lot of commercial mobile geo-applications. Actually this is a fundamental part of most of vehicle navigation systems, although the way it is realized, and the data source that is being used differ from implementation to implementation.

Zoom / pan / rotation functions are also common amongst available mobile navigation / mapping systems, nevertheless in a lot of cases, such as in TomTom Navigator v. 6, during navigation the map is automatically scaled by the software, even after the user changes the zoom manually. Movie-like continuous zooming, which is the idea in UWSM2 project, is more difficult to find in commercial mobile geo-applications, as in most cases when the user changes the scale of the map, there is no transient state between the initial and final scale. There is an abrupt change, which can confuse the user orientation ability making him/her lose the mental connection between the real environment and its abstract representation on the screen (Sarjakoski & Nivala, 2005).

The different map detail in different scales is something common, as most of the mobile mapping applications use a series of pre-composed stored forms of the same area maps that are then applied to different scale ranges. These forms involve different levels of applied generalization and different layers visualization. For example, from scale of 1:70.000 until 1:140.000 the same map is used (Destinator v.6), which is just zoomed in and out without change in the detail supplied.

Although vario-scale mapping will probably be used in UWSM2, it is not yet widely used in commercial geo-applications, as it still is a technique under research. This is however not a critical point for the present study, as its influence on the usability of the scaled map can be later evaluated with the use of the proposed methodology.

The availability of the mobile geo-application and its reasonable cost, considering the low budget of a research are some additional selection criteria, together with the good map coverage of the study area and the compatibility with the hardware architecture used. The area of Lonneker, as the area of field testing for the proposed methodology should be then covered in detail, something that a plethora of mobile navigation applications with map coverage of Netherlands can support. The hardware platform of IPAQ PDA running Windows mobile 2003 or 5 is a very common one, allowing for the use of a big number of different geo-applications.

The clear map scale, meaning that it should be visible to the user and changeable according to the zooming level, is something not happening in some applications as for example the Destinator v.6 and, and the same applies to the manual rotation and non-overlapping of POIs in several zooming levels too (Route 66 Navigate 7, TomTom Navigator v.6).

Taking into account all the discussed criteria and after comparing several mobile geo-applications that were or were not directly available to the researcher, including: Destinator v.6, TomTom v.6, Igo 2006, Route 66 and Marco Polo mobile Navigator 3, the choice is the Igo 2006. This application covers all the requirements for the testing, except only the luck of a vario-scale map capability, which however was not also found in the other applications (figure 4.8). The comparison between the characteristics of the five studied mobile geo-applications based on the selection criteria can be seen in table 4.2.

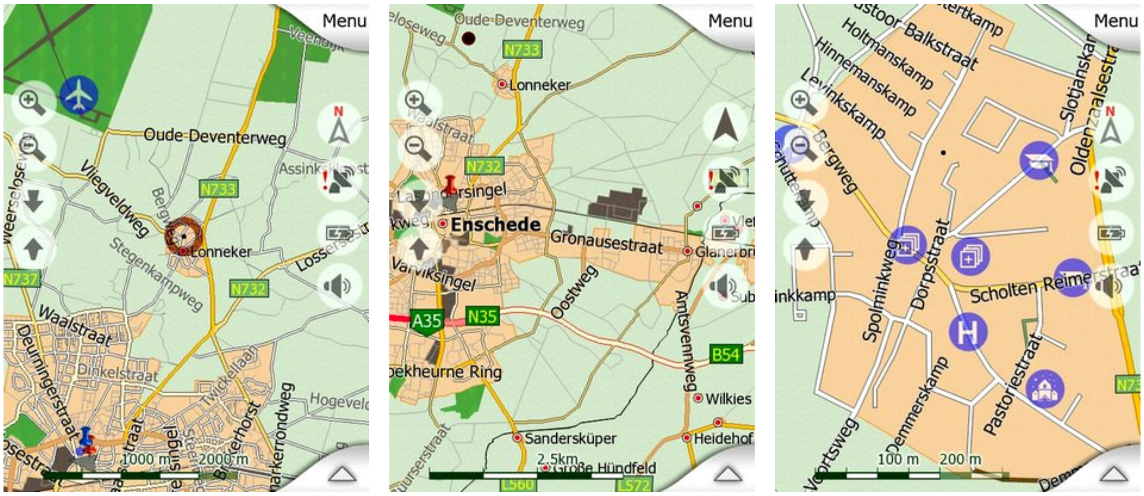


Figure 4.8 Different scales of Lonneker and Enschede area map in Igo 2006

Criteria	TomTom v.6 Western Europe	Route 66 Navigate v.7 Europa	Destinator v.6 Western Europe	Marco Polo v.3 Europe	Igo 2006 Western Europe
Geo questions answering	YES	YES	YES	YES	YES
Zoom / pan / rotation	YES/NO/NO	YES/YES/YES	YES/YES/YES	YES/YES/NO	YES/YES/YES
Non-auto zooming	NO	YES	YES	YES	YES
Movie-like zooming	NO	NO	NO	NO	YES
Different map detail / zoom	YES	YES	YES	YES	YES
Vario-scale map capab.	NO	NO	NO	NO	NO
Cost	156 EUR	209 EUR	193 EUR	170 EUR	170 EUR
Available to the researcher	YES	NO	YES	NO	YES
Detailed study area coverage	YES	YES	YES	YES	YES
Platform Compatibility	YES	YES	YES	YES	YES

Clear map scale	NO	YES	NO	NO	YES
Changeable orientation	YES (via menu)	YES (via menu)	YES (via menu)	NO	YES (direct)
POIs Non-overlapping	NO	N/A	NO	N/A	YES
Routing capabilities	YES	YES	YES	YES	YES

Table 4.2 Comparison of navigation software according to the defined criteria

4.8. Setting of the surveys

The setting of a usability testing is a very important issue as it can greatly influence the output of a usability study as well as the aspects of validity and reliability. According to Kjeldskov & Graham (2003), the evaluation of mobile systems can be realized in three different general settings. These are the environment independent, real environment based and artificial. In this research, the real context of use is selected, concerning the context awareness of UWSM2 project together with the fact that field-based usability testing is less researched, although it poses a great potent in further understanding and analyzing human-application interaction in the real world.

One of the bigger real context of use difficulties, as already discussed in the previous chapters, is the controllability of the dynamic context variables involved in the user interaction with the mobile application. As we use three groups of participants, involved in three different combinations of usability testing methods and techniques, and we want to compare the output of these three methodologies, the dynamic parameters that can influence the output should be kept inside reasonable limits. In order to achieve this, we have to make a descriptive definition of the testing environment, limiting the level of uncertainty that can influence the validity and alter the results. The surveys should not be performed, unless the survey area's conditions meet the predefined ones. However, as it is not possible to delimit all the contextual parameters regarding a real testing environment, we focus on the most important ones:

- Amount of light: the tests will be performed in daylight with the backlight of the PDA's display turned on and in the maximum setting
- Time of day: the surveys will be done from 8:00am until 17:00
- Place: the area of Lonneker, starting at a predefined point, the same for all the participants
- Weather: the tests will be done during sunny or cloudy weather, with the wind not exceeding the speed of 20 km/h. During raining, hailing or snowing the tests will not be performed. The temperature should not be below -2, concerning that it is a winter season; otherwise it can strongly influence the movement of the hands of the user. The weather will be monitored through online weather reports and forecasts for the selected area, and the participants will be informed on time in case of cancelling of the testing due to uncomfortable weather conditions

- Environmental noise: the tests will not be performed in case of demonstrations, big social celebrations or other noisy social or structural events in the area of surveys during the time of the tests

4.9. Tasks to be performed and time needed

In order to assess the map scaling usability of the selected application, tightly related to the functions of zooming, panning and rotation / orientation of the map, the participants of the research are asked to execute a tasks-based use scenario. In this scenario, the user represents a visitor / tourist in an unfamiliar town to him/her, wanting to navigate to 7 points of interest in that town, while using only a mobile map. S/he has to navigate to each of the points in a predefined order, starting from a point 0, and returning there after the completion of the tasks. Each of the 8 points (including point 0) is highlighted on the map with the use of different colour pin-points and when the user clicks on them, their name starts with “Point n” where n is the number of the particular point, numbered from 0 to 7.

The user is not allowed to use the routing and auto-navigation functions of the application, as the aim of the test is to assess the efficiency and effectiveness of zooming of the map at different scales / zoom levels from the user’s side. S/he is allowed to use the functions of zooming, panning, rotation, and change the orientation of the map to North-up, Heading-up and overview modes. The researcher is following the user in a distance of 10 to 25 meters, trying to avoid optical contact with him/her and talking to him/her in a way that is allowed and necessitates according to the methods used. For example, whenever Think Aloud is used, the researcher tries to keep the user talking and expressing his/her thoughts while using the application, by asking questions, while in the case that only observation and video recording is used, he remains generally silent. When the methods used include observation, the researcher takes notes describing the user mistakes, confusion moments and writes down general comments when it is necessary, in order to support the later analysis and confirmation of the results, based on the data colleted through the recording techniques.

The efficiency of the map scaling is reflected in the easiness with which the user understands his / her position inside a real environment using a mobile map and the ability to navigate to different points that are displayed in that map. In order to assess the attribute of efficiency, we assess in this research the time that the user spends in order to navigate from one predefined point of interest to another. There is also important to determine the time limits for completing each navigation task, which, when reached, the user should be moved to the next point and start the next task. In order to set these limits, the researcher, knowing the area very well, and one pilot user, unfamiliar with the area, carried out the designed tasks while the time that was needed from him was assessed. It was found that in every case the pilot user didn’t exceed the double of the time needed from the researcher to complete each task. Therefore, this was defined as the time limit applied to all the survey participants later. In table 4.3, the parameters of the navigation tasks, including the names of POIs, the order in which the user has to approach them and the distances and the time limits of each task are presented.

Task	From	To	Dist.	Time
1	Point 0 (car)	Point 1 (Rabobank ATM)	140 m.	4 min.
2	Point 1 (Rabobank ATM)	Point 2 (Café Sprakel)	190 m.	4 min.
3	Point 2 (Café Sprakel)	Point 3 (lake)	340 m.	8 min.
4	Point 3 (lake)	Point 4 (Church Kapel UL)	600 m.	14 min.
5	Point 4 (Church Kapel UL)	Point 5 (Vakdrogist – post box)	440 m.	12 min.
6	Point 5 (Vakdrogist – post box)	Point 6 (Lonneker map – info)	220 m.	4 min.
7	Point 6 (Lonneker map – info)	Point 7 (kindergarden)	120 m.	3 min.
All	Total		2050 m.	49 min.

Table 4.3 User surveys navigation tasks' parameters

As it can be noticed in the above table, the time needed for each task is not directly proportional to the distance that is required to be walked. This happens because the timing of each task also depends on the difficulty to follow a particular route, as it was found during the pilot testing. For example, in cases were one point is connected to the next one through a straight road, it is faster to navigate. In the figure 4.9 the map of the area, as seen in the application display is presented.

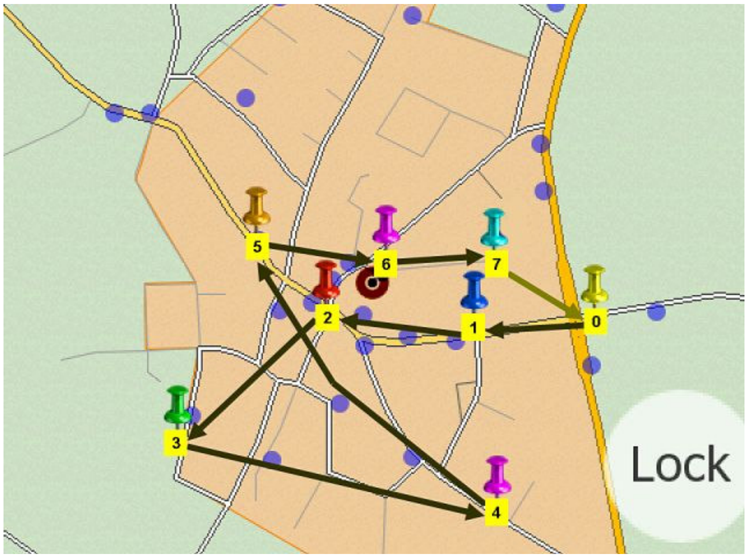


Figure 4.9 The pin-pointed POIs that the user has to navigate to

The numbers of the points are not directly visible to the user in the actual application and were put here only for reference, and the same applies to the lines connecting these points too. Actually, the user has to click on a particular pin-pointed POI in order to see the information about it including its point number if applicable, name of the POI, and address (Figure 4.10).

Figure 4.10 Visible POI symbols and information pop-up box of a POI after clicking on it.

During the actual navigation, of course, the user has to make also a lot of turns in different roads and not only walk in straight lines. S/he can always use the zoom-in function in order to see the symbol of a POI, as in smaller scales it is represented by a blue dot only on the map.

Each user's testing session is starting with a meeting of 15 minutes, where the scenario of a tourist in an unfamiliar city that the user represents is explained by the researcher, together with the map functions that can be used. The tasks that the user has to carry out are discussed with the help of a printed screenshot of the selected application's mapping display and after that a first contact with the mobile geo-application itself is taking place. During the transportation to the survey area, lasting 5 minutes, the user is exploring the full functioning application on the PDA, in order to get familiar with it. S/he is also helped by the researcher during this time, who is explaining again the functions and the capabilities of the application during movement, and giving advices that could help the user overcome possible problems in using the application.

After the arrival in the starting point I the Lonneker area, the user and the researcher spend 6 minutes to prepare all the devices and the systems that are needed for the testing, and they start the survey session. Finally, they return to the car and put out all the equipment that they are carrying, and the researcher asks for an interview from the user, which needs 10 minutes to be completed. The interview takes place in the car and the voice of the user is recorded in a small audio recorder. The user is then transported back to his place, again spending 5 minutes of time. The time needed for each part of the survey sessions is shown in table 4.4., where we can see that the total time of a survey session should not exceed 1.5 hours. These times were defined during the pilot testing and they were tested during the actual surveys in order to be adjusted if it is needed. In the next chapter, where the actual user surveys are investigated, we argue upon this matter.

Part of the survey session	Time needed (minutes)
Initial briefing	15
Transportation to the survey area	5
Preparation of equipment	6
Survey tasks execution	49 (max)
Unplugging the equipment and interviewing the user	10
Transportation back	5
Total time	90 (max)

Table 4.4 Time needed for each part of a survey session

4.10. Conclusion

In this chapter, the proposed methodology for the usability testing of mobile geo-applications in the field was discussed, based on the UWSM2 prototype evaluation, and focusing in the appropriate map scaling in specific contexts of use. The use cases, tasks and geographic questions that the UWSM2 project poses were studied, in order to extract a more abstract type of questions that an already existent mobile geo-application could answer. After that, the proposed methodology for the usability testing of this mobile geo-application was presented, together with its alternatives that will be evaluated during the field surveys. A detailed description of the electronic observation system that was developed in order to resolve some of the shortcomings of the combination of observation, Think Aloud and video recording techniques in the field was also given, as one of the most important parts of this research.

Following was the description of the process used for the selection of the testing participants, and the reasons behind the selection of ITC students for the tests. The choice of an existing (commercial) mobile geo-application that best resembles the UWSM2 functionalities was also justified, based on the meeting of a series of similarity criteria that were pre-defined.

Finally, the setting of the surveys was presented, considering the necessity of a well defined context of use and the description of the parts that constitute the survey sessions is given, together with the time needed for each of them, and especially the user tasks execution.

In the next chapter, the actual field testing process is presented, together with a description of the tasks that are given to the participants. The parameters that have to be adjusted during the testing in order to improve the process and several problems and unexpected events that the researcher faces are also discussed.

5. Results

5.1. Introduction

In this chapter, we will describe the actual user surveys procedure, based on the methodology that was proposed in the previous chapter. The problems that were found during the tests, the solutions that were given in order to overcome them as well as ideas for further enhancements that it was difficult to be applied during this research will also be discussed.

5.2. Execution of the tests

The testing surveys took place on January and February 2007, in the village of Lonneker, five kilometres from Enschede, as it is already mentioned. The total number of the participants was twenty persons, including eighteen persons for the actual tests and two additional, the pilot test user and one experienced usability advisor. According to the participants' answers during the interview part of this research, they were generally very satisfied with the whole setting of the experiment and they found the research objective very interesting and the methodology promising. Albeit some of the participants had to be re-scheduled for testing sessions due to weather restrictions, which are described in the next section, they always reacted with great understanding and zest.

Regarding the methods and techniques used for the surveys, the users' response was satisfactory, although there were differences, from person's to person's response, especially for the Think Aloud method. Some participants were very keen to talk, almost continuously, without needing any prompting from the researcher, while some others had to be triggered regularly through questions in order to express their thoughts as they were performing the tasks. However, there were moments were most of the participants, being in deep cognition process, stalled their talking and the researcher had to trigger them in a neutral way. There were also sometimes difficulties in the comprehensibility of users' words, as no one had English as his/her native language, something that applies to the researcher too, and the sound quality of the wireless audio transceiver was also moderate. During the interviews, the participants were generally very talkative, highlighting very interesting facts about the experiment's setting as well as mobile geo-applications and their usability in general.

5.3. Real and social environment issues

As the survey sessions had to follow a well defined context of use, including the environmental conditions, and considering the fact that the electronic and computer systems used for the testing are very sensitive to water, the weather conditions played a big role in the scheduling and execution of the surveys. The highly changeable weather in Netherlands, including a lot of rain and low temperatures during the winter season, led to a lot of delays and cancelling of scheduled user testing sessions. The maximum practically possible number of user testing sessions per day found to be three, limited by the battery-operated devices, and especially by the laptop and the portable video recorder. However, the number of one to two users per day proved to be more convenient, as it was allowing for full recharging of the devices and thorough checking of their functionality before each session. The low temperatures during some testing days also posed some problems, although the users could still work with bare hands without significant disturbance, except the participant B6 whose hands could not tolerate the below zero temperature and the survey session had to stop for a while until he was ready to continue again.

Another problem that was faced during the surveys was the disturbance of the researcher and the users from the social environment. There were nine cases that local people were interested to find out what was the actual subject of the research and why we had to carry all these electronic devices. Such an instance can be seen in figure 5.1, as recorded from the field observation system, where a woman is asking the participant about the research. As she could not communicate in English, the participant, who happened to be the advisor, explained her in Dutch what is this research about and why it is necessary to use this plethora of electronic and computer equipment.

Figure 5.1 Explaining the research objectives to local residents

Some of the local residents were very friendly and were asking just a few questions, while others were more curious and even suspicious sometimes, and the researcher had to totally stop the survey for some minutes in order to explain them the general facts about the experiment. The assessment time for the completion of a particular task was influenced in such cases, and the researcher had to take notes of the time delays due to such disturbances, in order to correct the measured time for the completion of that task. These time delays were later confirmed and corrected through the analysis of the video and audio recordings.

5.4. System issues

Although the unique system that was initially implemented for the usability surveys in order to enable the experiment to be done by only one researcher worked generally well, there were also shortcomings found, leading to improvements of the system. One of these shortcomings was the often interruption of the voice of the user coming to the researcher and being recorded to the video or audio recorder through the audio/video transceivers. This was considered to happen because of the interference of the PDA / laptop Wi-Fi connection with the audio/video transceivers, as they are both working in frequencies around 2.4 GHz. The solution here was to use a second pair of PMR radio transceivers for the voice of the user, not passing anymore through the audio/video senders. The headset connections had to be modified accordingly, in order to be connected to both the PMR transmitter and PMR receiver. A different frequency was selected for each PMR transmitter / receiver pair, so that the user and the researcher could talk and hear each other at the same time without problems. This modification was done after the initial participant's testing session, which was ceased due to the sound problems that appeared. The participant was not asked to contribute again, since a second contribution would decrease the validity of this research. He would be already familiar with the survey setting and this would possibly influence his performance, due to learnability and memorability issues. Therefore he was replaced by another person and he is not included in the user testing groups that are presented and investigated in this research.

Another issue was the quality of audio recording of both the researcher and the user, especially during Think Aloud sessions. In cases where both of them were talking at the same time, or there was high environmental noise, the discrimination of the voices during later auditing of the recordings was difficult. This problem was greatly reduced by recording the voice of the user in the left audio channel of the recorder, and the researcher's voice to the right. This modification was done after the fist two participants, and it took a short time. The modifications and improvements to the initial system described above are presented in figure 5.2.

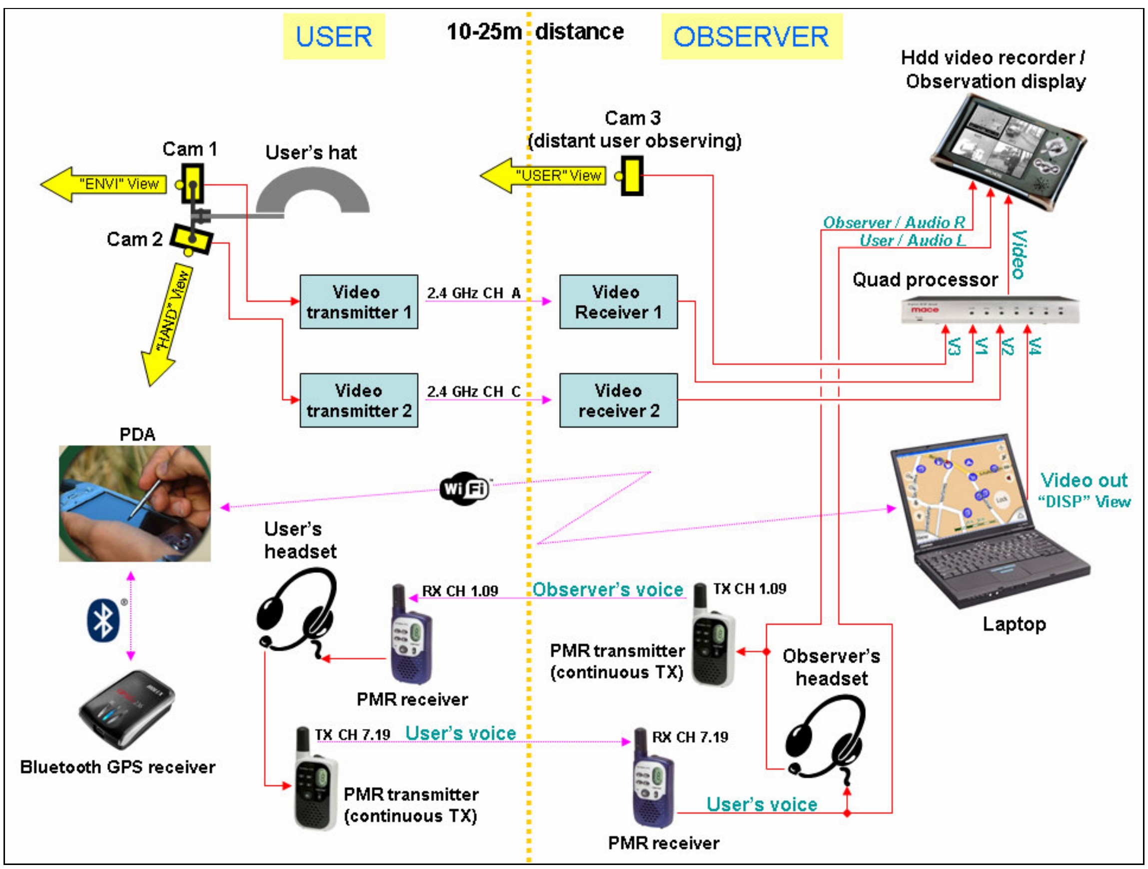


Figure 5.2 The diagram of the improved field observation system used for the tests.

The Wi-Fi connection between the PDA and the laptop, used for the screen capturing of the PDA display also showed some problems, resulting in often disconnections. In these cases the researcher had to stop the survey and try to re-establish the connection, spending the time that was arranged for the whole session. The solutions that found to highly improve this situation was the installation of an extended-range PCMCIA Wi-Fi card in the laptop and the use of a direct TCP/IP connection between the PDA and the laptop through the Pocket Controller software, removing the Microsoft ActiveSync involvement. The distance between the user and the researcher, although it could be up to 25 meters, in practice it was found that reaching the upper limit causes sometimes Wi-Fi instability and disconnections. Therefore a more safe distance proved to be between 10 and 20 meters.

5.5. Comparison between the three methodologies

As it is already mentioned, this research is not aimed at evaluating a particular mobile geo-application, but to use a mobile geo-application in order to evaluate 3 different methodologies for field-based usability testing. Therefore, what we are going to investigate is whether a particular combination of methods that were selected for this study is the most promising and if the unique technical solutions that were used support this methodology effectively.

5.5.1. Timing

In chapter 4, we described the process that was followed in order to determine the time needed for the completion of each use task, based on the researcher's and the pilot user's time measurements. But what were the actual times that the participants needed in order to carry out each task? Was the total time of 90 minutes available for the whole session of each participant enough? What were the differences between the different groups as it comes to time? In table 5.1, the timings of all the participants, divided in the 3 groups and in the different tasks, are shown. The number under each task (T1 to T7), is the time limit for that particular task, being set before the start of the surveys.

Group	User	Task 1 time 4 min limit	Task 2 time 4 min limit	Task 3 time 8 min limit	Task 4 time 14 min limit	Task 5 time 12 min limit	Task 6 time 4 min limit	Task 7 time 3 min limit	Tasks total time 49 min limit	Session total time
A	A1	x	2:38	5:45	7:42	5:43	x	x	32:48	101
	A2	1:27	2:12	4:34	5:21	6:35	2:57	1:30	24:36	79
	A3	2:05	3:36	6:21	7:03	5:58	3:49	3:12	32:04	76
	A4	2:00	3:41	x	11:51	x	x	2:28	44:00	103
	A5	2:33	2:27	5:23	9:30	5:10	3:15	1:34	29:52	97
	A6	x	3:41	6:01	9:12	8:55	x	2:49	38:38	88
B	B1	1:50	2:53	5:03	9:32	5:24	2:26	2:23	29:31	77

	B2	3:37	x	7:33	X	9:07	x	3:40	45:57	91
	B3	2:10	2:35	x	6:30	11:50	x	x	38:05	82
	B4	x	3:38	x	10:43	9:29	x	3:04	42:54	84
	B5	2:45	2:51	4:30	7:47	5:10	2:22	2:10	27:35	77
	B6	x	3:19	7:54	13:18	7:32	3:46	2:57	42:46	87
C	**C1**	2:41	2:28	5:00	9:14	7:08	3:21	2:36	32:28	83
	C2	2:49	2:57	7:51	11:23	7:00	x	3:55	39:55	125
	C3	1:30	2:32	3:27	5:41	9:24	3:52	1:43	28:09	74
	C4	2:35	4:00	x	6:04	11:12	x	2:38	38:29	107
	C5	3:03	2:34	8:00	7:35	9:11	2:00	1:33	33:56	78
	C6	1:53	x	x	6:09	9:27	x	2:34	37:03	117

Table 5.1 Timings of the participants for each task

In this table, the letter x represents the POIs that were not found inside the defined time limits. An interesting thing is the appearance of high values in the total session times of groups A and C. There are 3 out of 6 users in each of these groups that needed more than 90 minutes for completing a full session, including briefing, transportation, equipment preparation, tasks execution and interview. These users are the A1, A4, A5, C2, C4 and C6. In group B there is only one such instance, regarding the user B2, but even in this case, the difference is only plus 1 minute. These high values are the result of the disconnections of the Wi-Fi connection that affected the user sessions before the improvement of the system with a longer range Wi-Fi card, as described in section 5.4. The time needed for the process of re-establishment and testing of the Wi-Fi connection is excluded from the users' task execution time, but not from the total session time. Only groups A and C, as it is already known, used the Wi-Fi connection for remote PDA screen capture as part of the observation and video recording methods and techniques, so the problem appeared in these groups. Generally, it can be argued that there is no significant difference in the total time needed for each group, representing a different methodology for usability testing, considering the timings achieved after the improvement of the field observation / recording system.

Another interesting finding in this table is the high value of failure into completing 6th task, where half of the participants, equally distributed in all the groups, could not find the POI number 6, which was in reality a tourist information point, including a road sign with an "i" letter and a big tourist map of Lonneker (figure 5.3). This fact shows how important is a more descriptive presentation of some POIs on the map, as half of the users were standing a few meters from this point and could not find it at all, until they reached the time limit of that task, or they thought they found it, unluckily standing in another close point.

The interviews, as the final part of every testing session before transportation back, took an average of 9:42 minutes, ranging from a minimum of 5:26 to a maximum of 12:40 minutes. Considering the fact that initially the time reserved for the unplugging of the equipment together with the interview was 10

Figure 5.3 The difficult to find tourist map of Lonneker.

minutes, and the time actually needed for the unplugging is between 2 to 3 minutes, the interview / unplugging time has to be practically adjusted to 16 minutes.

As it comes to the first part, that of introduction and description of the research objectives, tasks and application functions, proved to be sufficient for the majority of the participants in terms of time, combined with the practical use of the geo-application inside the moving car during the transportation to the testing area. However, 3 out of the 18 participants, A1, A6 and B6, argued that they would better spend a little more to get more familiar with the application before they perform the actual tasks. These users were the 3 out of 4 that failed to find POI 1, as they were moving to the opposite direction of the correct one and they could not realize it through the use of the mobile map.

5.5.2. Effectiveness and efficiency measured

In order to measure the effectiveness and efficiency of the selected map application scaling, the achievement of the task goals, the time that was needed for the completion of each task, discussed in the previous section, as well as the number of mistakes that the user made while performing the navigation tasks are investigated. The researcher used a notes page, presented in Appendix B, with separated boxes for each task. There he was writing whether a task was completed, the time needed for this completion, using a chronometer, as well as the number of errors and the severity of each of them, together with general notes about a particular user. This information proved to be very useful later, during the analysis of the video and audio recordings, as it gives a general idea for the user performance and the parts of the sessions where something important happened. The same technique was applied to all the groups, although in groups A and C, where the remote observation / recording system was used, an easier and faster confirmation and investigation of timings and errors was possible, as the researcher has on the screen the time of day, the view of the surroundings that the user moves through and the map display of the PDA. During group B sessions, where only think-aloud method with voice recording was used, the process of confirmation and investigation became much more complicated and lengthy.

As a measuring unit of effectiveness, we use in this research, where there is only a fail / success state for each task, the percentage of the users that achieved the task goals. If we had used a wider range of task completion characterization, such as partially successful completion, then we should have used the average task completion given in percentage units. In figure 5.4 we can see the results of the measurements of effectiveness achieved in each group. The difference between the groups is not statistically significant, and it seems that it would be smoother if the user sample was bigger.

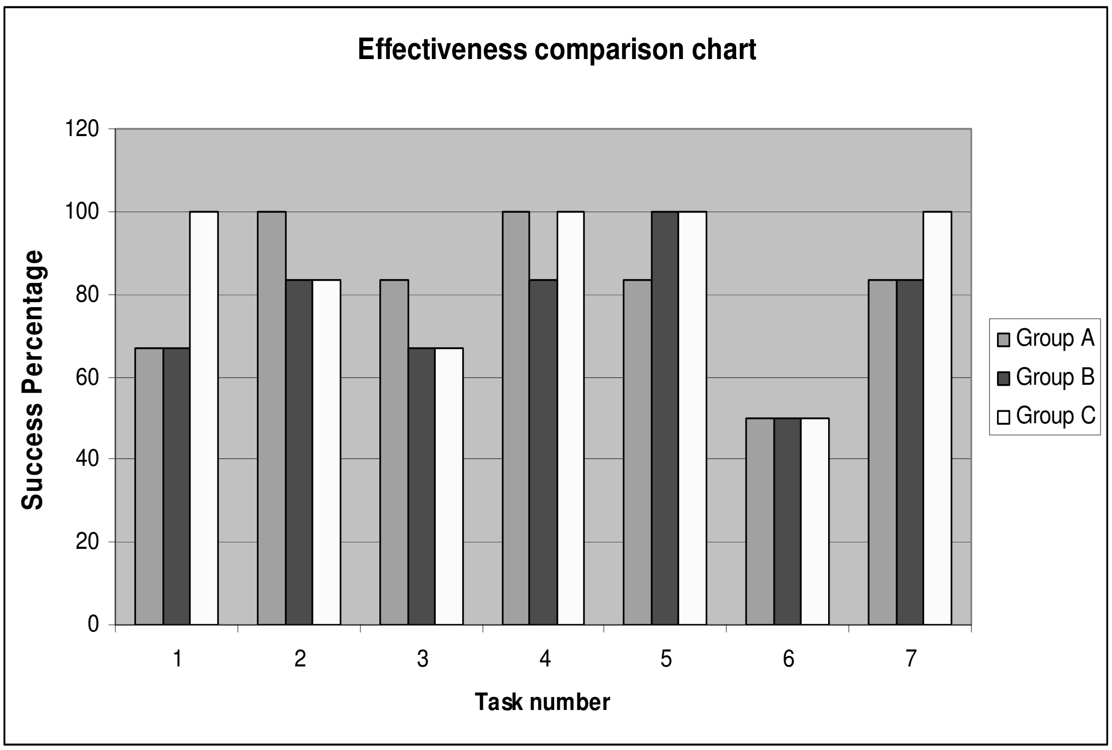


Figure 5.4 Effectiveness results comparison of the 3 groups

Efficiency is measured by the mean time needed for the completion of a task, related to the achieved effectiveness percentage. In figure 5.5 we can see the results of the efficiency achieved from the different groups of this research. Again there are some deviations, not very significant statistically, showing that the selection of the participants for the 3 different groups was appropriate and the results of the comparison valid. It seems again that the differences between he groups would be smoother if the user sample was bigger.

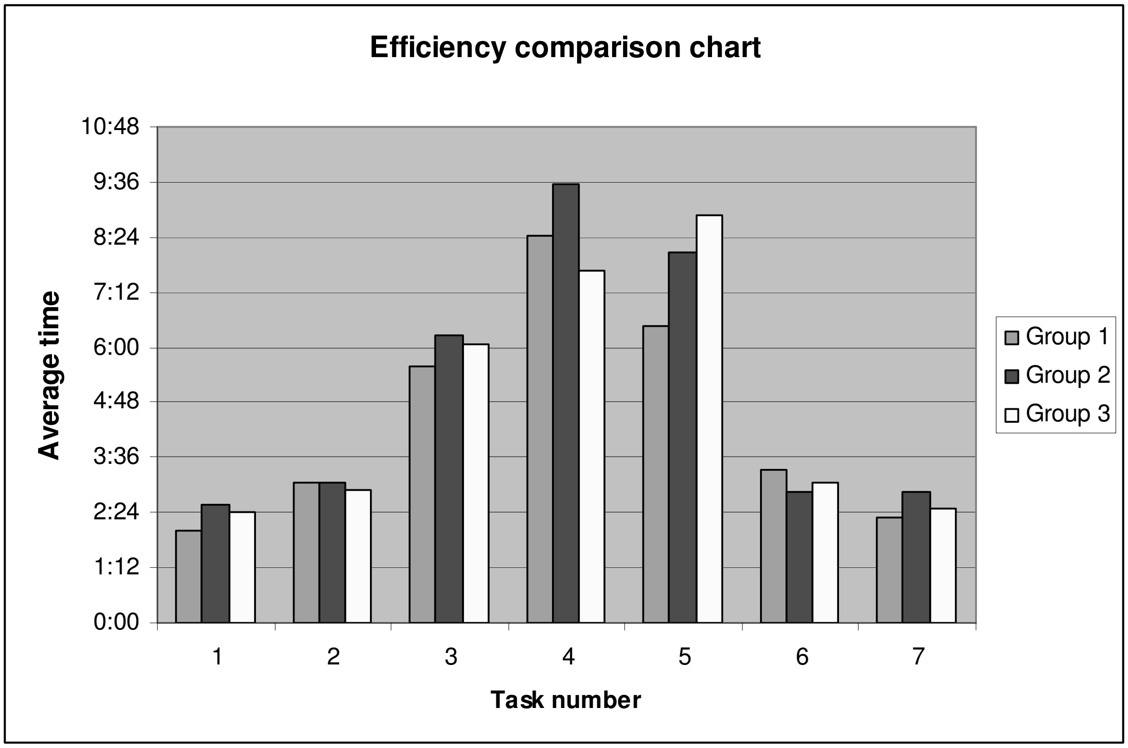


Figure 5.5 Efficiency results comparison of the 3 groups

5.5.3. Interviews and user satisfaction regarding the map scale and the test surveys

The user satisfaction was measured through the semi-structured interview that was held at the end of each testing session. Although the semi-structured interview gave very interesting and valuable results for the application investigated functions' usability, map scaling suitability and comprehensiveness and the experiment as a whole, it was found to be insufficient in the way it was structured, to properly scale the user satisfaction. Therefore it has to be reformed and adjusted in order to cover this need. The questions asked through the questionnaire are presented in Appendix C.

Generally the participants were satisfied with the map scaling of the mobile geo-application that they used for the tests, the information available in different zoom levels and the functions of zooming, panning, rotation and orientation of the map. However, they mentioned helpful additions, such as graphical landmarks that could help them orientate easier in their navigation tasks and not so sensitive GPS-based auto rotation of the map during heading-up mode. A lot of user mentioned that they found themselves confused in times when the direction arrow appeared on the map was showing the wrong direction. This generated to the researcher the following question that was added to the later users' interview questions: Do you think that it would help you orientate yourself and easier navigate if the map had the ability to rotate when you turn around yourself without having to walk? Most of the users were excited to this idea, although a few, mostly people that used only the north-up orientation of the map, did not think that this would make a difference for them.

As it comes to the survey setting, all of the users argued that it was very interesting experience for them and that mobile mapping is an exciting and promising field. Most of them agreed that mobile maps are more useful than paper maps, especially regarding their ability to zoom in and pan, while they are showing the current position of the user on the map, helping him / her navigate efficiently. The small size of the screen did not seem to bother them, and they seem to prefer it more than a big, unfolded map, where amongst others they could not click on a POI and find more information about it. The ability to have graphical information when they click on a POI, such as photos and historical / tourist facts was mentioned as very useful by some of the users too.

Regarding the think aloud sessions and the amount of comfort / discomfort that they felt, almost all of the participants agued that they would not behave in a noticeably different way if they were totally alone and used the same mobile geo-application in order to carry out the same tasks as in this research. Actually some of them even mentioned that the think-aloud helped those to better understand what they are doing and why they are doing that, resulting in the correction of their probable mistakes.

When the users were asked to compare the unique system that was used for the observation in this research with a case where a person would use a usual camcorder and record them from a close distance, all of them agreed that the latter would definitely influence them much more and make them feel uncomfortable. They prefer the idea of communicating wirelessly with the observer and not see him / her beside or in front of them.

5.5.4. Potential shown by each methodology

Based on the outcome of the usability surveys we will try to make a comparison between the 3 different methodologies that were used. There is certainly not perfect methodology, since every combination of methods in usability testing always has advantages and disadvantages. Context of use also plays a big role in the selection of the most appropriate amongst them.

The most easy to implement and less technically complicated methodology was the one that was applied to group no. 2, involving Think Aloud, audio recording and interview. The small number of electronic devices used, led to faster preparation times and less technical problems. Also, the battery power needed for the operation of the devices was low. However, the actual session times do not show any noticeable difference in the total survey times, considering the times of the other groups after solving the technical problems that appeared. Think Aloud proved to be a very useful method in revealing the user's cognitive process during his/her actions of orientation and navigation. Unfortunately, the analysis of the results based only on audio recordings, proved to be much more difficult than the audio/video ones. There is no indication of time, and the researcher had to keep a chronometer from the start of the auditing of the recordings in order to keep a track of time. Again, there is no indication that the user is for example approaching a POI, or s/he is going in the totally opposite direction, due to the lack of an observation method in this group.

In the group no. 1, where observation together with video recording and interview was used, the researcher could easily analyse the video recording, as he could see on the screen exactly what was happening, and even small disturbances, confusions or mistakes of the user could be investigated during the watching of the recordings of the sessions. The displayed time and date ensures the validity of the findings, as it prevents the researcher from making possible timing mistakes, while it allows for a confirmation of the participant's name through the visual recognition and according to the survey schedule. Observation with video recording was an efficient method, which additionally allows for even deeper analysis of the results from usability experts that do not have to be in the field during the testing. However, the lack of Think aloud does not allow the investigation what the user is thinking during the tasks' execution. The analysis of the results can be based only on the actions of the user, but what is missing is the motivation behind these actions. A way to solve this problem was thought to be the addition of user-specific questions regarding the tasks execution in the unstructured part of the interview. The researcher was asking the users why they reacted in a particular way during a specific task, in order to investigate the reasons behind their actions. For example, the participant A1 started walking in the totally wrong direction during the task no. 1, until the time determined for this task was over and the task was considered failed. During the interview, when the question was whether he could easily understand his position through the map, he answered: "yes, easily". After that, the researcher asked him why then he was walking to the opposite direction. The participant answered that during the first task he found himself confused with the mobile geo-application, as he forgot how to use the functions of orientation and zooming correctly, and after that he did it fine. If we have now a now a look at table 5.1, we will see that this participant finally failed in 3 out of the 7 tasks, which can have different meanings. Here is the part where Think Aloud could help better understand the causes of the user's particular actions. An issue regarding this group is the high battery needed for the devices, as it uses the laptop, the video quad processor, the cameras and the video recorder.

As it comes to methodology no. 3, using observation, Think Aloud, audio/video recording and interviews, it proved to be the most usable. In this group, a combination of all the initially selected methods was applied, and all the parts of the uniquely implemented field survey system were used. This was the most technically complicated methodology, requiring slightly longer times for the preparation and disconnection of the equipment. Here is also were the problems of wireless connectivity, together with group 1 appeared. However, after the adjustment and improvement of the system, these problems eliminated. This methodology gave lot of information about the user, and allowed for a deeper analysis of the results. The audio and video recordings, combining the Think Aloud protocol with the views of the user, his/her surroundings, the movements of his/her hands and body, the PDA screen capture and the time/date stamp, is giving a complete overview of the user-application interaction in the real context (table 5.2). Battery power is also an issue regarding this group, as it uses all the devices of the implemented system. In order to eliminate this problem, special high capacity lithium batteries were used.

Methods / Group 1		**Methods / Group 2**		**Methods /Group 3**	
Advantages	Disadvantages	Advantages	Disadvantage	Advantages	Disadvantages
Video recording allows visual investigation / confirmation	No insight in the user's thoughts during task execution	The easier to implement	No visual investigation / confirmation	audio/visual investigation / confirmation	The most complex
Observation allows for thorough investigation of user actions	Complexity not significantly lower than group 3	Investigation of user's thoughts	audio recordings alone make analysis more difficult	Observation allows for thorough investigation of user actions	Systems used need a thorough testing before applying

	Battery power an issue		Long battery power	Insight of user's thoughts	Battery power an issue

Table 5.2 Comparison of the 3 methodologies

5.5.5. Additional remarks - adjustments

In this research, a fundamental assessment of efficiency and effectiveness was done. But it is assumed that the capabilities of the technical solutions used, allow for the investigation of additional aspects of these variables. For example, remote data logging, a technique not used in this research, can be applied through appropriate software, allowing to automatically measuring how many times the user used the zoom-in, zoom-out or panning functions, where the higher the number, the less the usability of the map. These measurements can also be done through the current video recordings of the methodologies 1 and 3, although much more time and effort is needed for this aim as it has to be done manually by the researcher.

As it comes to the technical problems that were appeared during the testing, it is argued that it would be beneficial if more pilot testing was done before the execution of the main survey. In this case, most of the problems would have been tracked earlier and the improvements of the system would have been done on time.

6. Conclusions and recommendations

6.1. Summary

This research thesis investigated the selection and use of different methodologies for the usability testing of mobile geo-applications, and particularly of those that can be applied for the evaluation of the UWSM2 project prototypes. Usability testing proved to be a very important part of the design and implementation of such applications under an iterative process and considering a well defined context of use. Although still conducted by limited research, field-based usability testing is a very promising field, especially when it comes to the evaluation of context-aware mobile geo-applications. During this research, the special requirements, limitations and advantages of mobile geo-applications were studied, together with the usability testing methods and techniques that can be applied for their evaluation. A methodology for field-based usability testing was then proposed, using a selection of different methods and techniques in 3 different combinations. These were the observation, Think Aloud, audio and video recording, and semi-structured interview. In order to investigate the most resultful combination of them, a mobile geo-application, similar to that which UWSM2 will soon implement as a prototype is used in order to be evaluated using this methodology. Not all the functions of this application are evaluated, but especially the quality of map scaling and the zooming, panning, rotation and orientation functions. A unique system was implemented in order to support and enhance the proposed methodology, studying the possibility of using high technology in order to replace human resources and time needed in order to conduct field-based usability testing. The user tests that were performed led to improvements of the system and adjustments and finalization of the methodology.

In the first chapter, a general introduction to this research was given, presenting the achievements, problems and challenges of mobile maps and mobile devices and the need for more research on the usability testing of mobile geo-applications, especially the field-based one, which is less developed but poses a very good potential. Additionally, UWSM2 project will soon implement a prototype mobile geo-application, and there is a need for an appropriate modern usability testing methodology in order to evaluate this prototype. These were the main motivations behind this research.

In the second chapter, the mobile geo-applications were investigated, together with the questions that these applications are aimed at answering and the use cases that can be used to describe the user-application interactions in real context. Through the study of a vehicle navigation system, it was shown that the context of use plays a very important role in mobile geo-applications, and should be always taken into account, as it can greatly change the mobile geo-applications usability.

The usability methods and techniques are the subject of the third chapter, under the concepts of UCD and HCI. Each method and technique proves to have several advantages and disadvantages, making the choice of the appropriate of them a matter of different variables, such as the main objective of the survey, the context of use, the setting of the research, the participants' characteristics and so on. After that, three examples of mobile geo-applications are discussed. Their requirements, goals and results, together with the methods of usability that were applied in order to assess them are presented. It is argued that the usability testing needs further research, especially as it comes to field-based testing, which is more difficult to be done due to the lack of controllability of the surveys and the higher resources needed.

In order to contribute to the further development of field-based usability testing of mobile geo-applications as a way to investigate usability problems that cannot be found in the non-real laboratory environments, we proposed a field-based usability testing methodology that can be applied to several mobile geo-applications. The usability testing of the forthcoming prototypes of the UWSM2 project is the base on which this methodology was built. The use cases, questions and functions of UWSM2 prototypes were studied, in order to determine a series of criteria for selecting a similar mobile geo-application that would play the role of UWSM2 prototypes, testing the methodology that was implemented. This application was the pedestrian and vehicle navigation software named Igo 2006. The variables of usability that were selected in order to be assessed in the user surveys were the fundamental usability variables of effectiveness, efficiency and user satisfaction. The methods and techniques that were selected as appropriate were the Think Aloud, observation, audio / video recording and semi-structured interview. Three different combinations of them were formed, in order to be tested during the user surveys. The interview was common to all the methodologies, assessing the user satisfaction through the use of the mobile geo-application, while the effectiveness and efficiency was measured through the successful completion and the time it took for the successful completion of 7 user tasks that were designed. These user tasks were actually navigation tasks, from one Point of Interest to another. The zoomability of the map was checked, as it influences the ability of a user to navigate successfully. ITC students were used as participants for the user testing, which were selected through pre-selection questionnaires. The researcher believed that with the proper technical solutions a field-based survey could be undertaken by only one person, and in order to try this possibility, he implemented a unique system that could support the selected methodologies.

In the fifth chapter the user testing was presented, discussing the problems that were found and the solutions that were applied in order to eliminate them. Most of the problems would have been revealed earlier, if there was time for more pilot testing before the user surveys. The results of these surveys were then presented, showing that the selection of the participants that formed the 3 methodologies groups was appropriate, during the common to these methodologies tasks, the results were similar. However, the 3 methodologies had their own advantages and disadvantages, and the most appropriate found to be the 3rd one, involving the combination of all the selected methods for this research, as it can investigate more usability issues than the other methodologies and through the use of the implemented system it is efficiently applicable.

The mains results of this research will be synopsized by the answering of the research questions that were framed in the first chapter, under section 1.5.

What are the investigated problems in already done research regarding usability evaluation for mobile geo-applications?

While there is a lot of research on these applications' usability, there are only a few studies regarding field-based usability testing, and the absence of modern technical solutions poses a series of shortcomings, such as high human resources and time needed, while the test user still feels uncomfortable and "unnatural" with the current usual testing settings.

Which methods and techniques or which combinations of different methods and techniques would be the most appropriate for field-based usability testing of mobile geo-applications and particularly for the assessment of mobile map scaling, as a fundamental part of UWSM2?

There are several methods and techniques that show a good potential and they could be used for this matter. Observation, Think Aloud, video or / and audio recording and interview are classical methods, proved to be efficient and not generally very difficult to be applied. There is however no "best solution" recipe, as everything depends on the context of use, the special characteristics of the test users and what exactly the research wants to investigate. In case of UWSM2, where appropriate map scalability is the research core, a combination of laboratory and field-based usability testing, using the above mentioned methods and techniques would be a proper approach. During this research, where 3 combinations of different methods were assessed, the most appropriate method found to be the one that combines Think Aloud together with observation and semi-structured interview. Video together with audio recording is also used during the sessions. This combination proved to be able to assess all the variables of usability that were initially set and at the same time to make a deeper investigation on the causes of the user reactions. Validity and reliability of the research can also be further verified through the recordings.

What unique technical solutions could be developed and applied in order to overcome current problems of usability testing in the field?

The current problems have to do with the difficulty of conducting a field-based survey itself, as the needed resources to conduct such a test are usually high and discourage the researchers. High technology can be used in order to implement systems that are capable of replacing the human potential needed with electronic and computer devices, and enable one-person (researcher) solutions. Remote screen capture, wireless voice and video transmission and recording, and remote controlling of mobile devices are some of the solutions that were investigated during this research through a uniquely implemented field survey system. These solutions can be further investigated and adjusted in order to support faster, easier and less expensive future field-based usability evaluations.

Which is the usability testing methodology that is found to be the most efficient for the evaluation of the mapping scalability functions of the prototypes of UWSM2 in the field and why? How could this methodology be adapted in order to be used for the evaluation of other (map) functions and different contexts of use in different mobile geo-applications?

The combination of observation, think aloud and interview, together with a remote observing / recording / communicating system shows a good potential as it can investigate several usability problems, poses a low bias to the test user because of the relative "invisibility" of the researcher to the user and can be further expanded to carry out more advanced functions such as automatic analysis of the results. Surely, there are additional methods and techniques that could be used for the methodology testing, such as the questionnaire after the survey and the data logging. The limited time available, however, would not allow for further tests with more methods and techniques. The combinations of methods that were selected for the formations of the 3 methodologies were representative examples, and through the comparison of them valuable conclusions were drawn. Actually, it is proven that, as every method and technique has its advantages and disadvantages, a proper combination of them can eliminate these particular disadvantages and investigate a lot of usability problems under real context of use. One important thing that has to be considered, still, is the cost of the methods in human and material resources. It can be argued that the system that was implemented succeeded into reducing the researchers needed for the execution of the testing surveys, although it is still to be investigated whether the results would be different if more human potential was used, different technical approaches or participants with different knowledge and background.

6.2. Recommendations

This research work focus on the development of a field-based usability testing methodology for mobile geo-applications, which could be used for the usability testing of the UWSM2 prototypes. It was found that the use of advanced technical solutions can efficiently support the application of such methodologies in real contexts and reduce the resources needed. In order to support this research‘s findings, and motivate new ideas, some aspects for further research will be given:

- There are more usability testing methods and techniques that could be used for the usability testing in the field. It would be interesting to combine additional methods in order to see whether the results are changing

- ITC students, having specific geographic knowledge were used for the testing surveys in this research. But how would the participation of other categories of users influence the outcome?

- The testing surveys were done in specific context of use, such us predetermined time of day, amount of light from the environment, weather conditions etc. How big would be the influence of a change in these contexts, as for example if the surveys were done in night hours?

- For the usability testing, PDAs with relatively big displays were used. What would be the outcome if devices with smaller displays, such as smartphones, would be used?

- The participants, during the interview, mentioned the lack of accuracy in the arrow showing the direction of their movement on the map. When they were asked whether they would like the possibility of a rotating map without the use of the GPS signal, but with other methods, that could work even when they just rotating around themselves, they answered positive. How this solution be implemented?

- Would the combination of a usability survey in the laboratory with a survey done in the field, similar to the one applied in this research lead into finding more usability problems, especially related to map zoomability?

- For the usability testing in this research, a unique system was implemented in order to support the application of the selected methods by only one researcher. How would the findings of this research change if simpler, usual techniques, such as ordinary camcorders were used?

7. References

7.1. Bibliography

Abras, C., Maloney-Krichmar, D. & Preece, J. (2004), User-Centered Design. In: Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications. Draft obtained on 25.11.2007 from *http://www.ifsm.umbc.edu/~preece/Papers/User-centered_design_encyclopedia_chapter.pdf*

Anderson, P. & Blackwood, A. (2004), Mobile and PDA technologies and their future use in education, JISC Technology and Standards Watch: vol. 04-03, November 2004. Paper obtained on 7.11.2006 from *http://www.jisc.ac.uk/uploaded_documents/ACF11AE.doc*

Bevan, N. (1999), Quality in Use: Meeting User Needs for Quality. In: Journal of Systems and Software, vol. 49, pp. 89-96.

Bevan, N. & I. Curson (1998), Planning and implementing user-centred design. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 1998), 18 - 23 April 1998, Los Angeles, California, pp. 111 - 112.

Bjørke, J. T. (2003), Generalization of Road Networks for Mobile Map Services: An Information Theoretical Approach, Norwegian Defence Research Establishment, Horten, Norway. Paper obtained on 15.11.2006 from *http://www.survey.ntua.gr/main/courses/geoinfo/admcarto/lecture_notes/map_complexity/bibliography/bjorke_2003.pdf*

Brink, T., Gergle, D. & Wood S. D. (2002), Usability for the Web: designing web sites that work. San Francisco: Morgan Kaufmann Publishers.

Burnett, G. E. (2000), Usable Vehicle Naviagtion Systems: Are We There Yet? In: Vehicle Electronic systems 2000 - European conference and exhibition, 29-30 June 2000, pp. 3.1.1-3.1.11. Paper obtained on 7.11.2006 from *http://www.cs.nott.ac.uk/~geb/VES2000-burnett.pdf*

Burnett, G. E., Summerskill, S. J. & Porter, J. M. (2004), On-the-move destination entry for vehicle navigation systems: Unsafe by any means? In: Behaviour & Information Technology, vol. 23, no. 4, pp. 265–272. Paper obtained on 8.11.2006 from *http://www.cs.nott.ac.uk/%7Egeb/BITdestentry.pdf*

Burnett, G. E. (1998), "Turn Right at the King's Head". Drivers' Requirements for Route Guidance Information. PhD Thesis, November 1998, Loughborough University, United Kingdom. Obtained on 8.11.2006 from *http://www.cs.nott.ac.uk/~geb/GEB-PHD.pdf*

Davis, D. E. (2003), GIS for Everyone: Exploring your neighborhood and your world with a Geographic Information System. Redlands: ESRI Press.

Elias, B., Paelke, V & Kuhnt, S. (2005), Concepts for the cartographic visualization of landmarks. In: Proceedings of the 3rd Symposium in Location based Services & Telecartography, Vienna, Austria, 28 - 30 November 2005. Geowissenschaftliche Mitteilungen, Heft Nr. 74, Schriftenreihe der Studienrichtung Vermessungswesen und Geoinformation, Technische Universität Wien, pp. 149-155.

Erdogmus H. & Tanir O. (2002), Advances in Software Engineering: Comprehension, Evaluation, and Evolution. New York: Springer-Verlag.

Frank, C., Caduff, D. & Wuersch, M. (2004), From GIS to LBS: An Intelligent Mobile GIS. In: GI-Days 2004, Muenster, Germany, IfGI Prints, pp. 277-287.

Fraunholz, B. & Jung J. (2002), Evaluation of Mobile Applications-Software-Technical and Human aspects. In: Proceedings of the 9th European Conference on Information Technology Evaluation (ECITE 2002), Universite Paris-Dauphine, Paris, France, 15-16 July 2002, pp. 165-172.

Goldsberry, K. & Dillemuth, J. (2006), Spatial Context in Mobile Map Design: Display Scale and Viewable Map Extent. In: Proceedings of GIScience 2006, Muenster, Germany, September 19-23, 2006, pp. 77-81.

Goulimis, E. D., Spanaki, M., & Tsoulos, L. (2004), Context-based cartographic display on mobile devices. In: Proceedings of the 2nd Symposium in Location based Services & Telecartography, Vienna, Austria, 28-30 January 2004. Geowissenschaftliche Mitteilungen, Heft Nr. 66, Schriftenreihe der Studienrichtung Vermessungswesen und Geoinformation, Technische Universität Wien, pp. 25-31.

Gulliksen, J. et al. (2003), Key Principles for User-Centred Systems Design In: Seffah, A., Gulliksen, J. & Desmarais M. C. (eds), Human - Centered Software Engineering. Dordrecht: Springer. Chapter 2, pp. 17-36.

Karimi, H. A. & Hammad, A. (2004), Telegeoinformatics - Location-Based Computing and Services. Boca Raton: CRC Press.

Heidmann, F., Hermann, F. & Peissner, M. (2003), Interactive maps on Mobile, Location-Based Systems: design solutions and usability testing. In: Mobile Cartography, Proceedings of the 21st International Cartographic Conference (ICC 2003), Durban, South Africa, 10 - 16 August 2003. CD-ROM, pp. 1299-1306, International Cartographic Association.

Hermann F. & Heidmann F. (2002), User Requirement Analysis and Interface Conception for a Mobile, Location-Based Fair Guide. In: Fabio Paternò (ed.), Human Computer Interaction with Mobile Devices, Proceedings of the 4th International Symposium on Human Computer Interaction with Mobile Devices and Services (Mobile HCI 2002), Udine, Italy. Berlin etc.: Springer, pp. 388-392.

Hinze, A. & Buchanan, G. (2005), Context awareness in Mobile Tourist Information Systems: Challenges for User Interaction. In: International Workshop on Context in mobile HCI at the 7th International Symposium on Human Computer Interaction with Mobile Devices and Services (Mobile HCI 2005), Salzburg, Austria, 19 - 22 September 2005. Paper obtained on 2.11.2006 from *http://www.cs.waikato.ac.nz/~hinze/isdb/publications/hinze_buchanan_CONTEXTmHCI.pdf*

Isomursu M., Kuutti K. & Väinämö S. (2004), Experience clip: Method for user participation and evaluation of mobile concepts. In: Proceedings of the 8th Conference on Participatory Design, (PDC 2004), Toronto, Canada, July 2004, pp. 83 - 92.

Ivory, Y. M. (2003), Automated Web Site Evaluation - Researchers' and Practitioners' Perspective. Dordrecht: Kluwer Academic Publishers.

Jordan P. W. (1998), An introduction to Usability. London: Taylor & Francis LTD.

Jordan P. W. et al. (1996), Usability evaluation in industry. London: Taylor & Francis LTD.

Hampe M. & Paelke V. (2005), Adaptive maps for mobile applications, Institute of Cartography and Geoinformatics, University of Hannover, Hannover, Germany. Paper obtained on 11.11.2006 from *http://www.ikg.uni-hannover.de/publikationen/publikationen/2005/hampe_mobilemaps2005.pdf*

Kaikokken A. et al. (2005), Usability Testing of Mobile Applications: A comparison between Laboratory and Field Testing. In: Journal of Usability Studies, issue 1, vol. 1, pp. 4-16.

Karimi, H. A. (2004), telegeoinformatics: Current Trends and Future Direction. In: Karimi, H. A. & Hammad, A. (eds), Telegeoinformatics: Location-Based Computing and Services. Boca Raton: CRC Press. Chapter 1, pp. 3-25.

Karimi, H. A. & Hammad, A. (2004), Telegeoinformatics: Location-Based Computing and Services. Boca Raton: CRC Press.

Kimber, J., Georgievski, M. & Sharda, N. (2005), Developing Usability Testing Systems and Procedures for Mobile Tourism Services. In: Annual Conference on Information Technology in the Hospitality Industry (HITA 2005), Hospitality Information Technology Association, 19-20 June 2005, Los Angeles, USA.

Kjeldskov, J. & Graham, C. (2003), A Review of Mobile HCI Research Methods. In: Proceedings of the 5th International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI 2003), Udine, Italy, 8-11 September 2003. Berlin etc.: Springer-Verlag. Usability and HCI Research methods, pp. 317-335.

Kjeldskov et al. (2005), Evaluating the Usability of a Mobile Guide: The influence of Location, Participants and Resources. In: Behaviour and Information Technology (BIT), vol. 24, no.1, pp. 51-65.

Koohang, A. & Harman, K. (2007), Learning Objects: Theory, Praxis, Issues and Trends. Santa Rosa: Informing Science Press.

Kraak, M.J. & Ormeling, F.J. (2003), Cartography: Visualization of Geospatial Data, 2nd edition. Harlow: Prentice Hall.

Kumar, R. (2005), Research Methodology: A Step-by-Step Guide for Beginners, 2nd edition. London etc: Sage Publications LTD.

Leitner, G., Plattner, S. & Hitz, M. (2006), Usability Evaluation of mobile Applications in leisure-oriented Contexts. In: Proceedings of the 13th International Conference of Information and Communication Technologies in Tourism (ENTER 2006), Lausanne, Switzerland, 18 - 20 January 2006.

Lee, J., Forlizzi, J. & Hudson S. E. (2005), Iterative Design of MOVE: A Situationally Appropriate Vehicle Navigation System, Human-Computer Interaction Institute and School of Design, Carnegie Mellon University, Pittsburgh, PA, USA. Paper obtained on 9.11.2006 from *http://goodgestreet.com/docs/lee_IJHCS.pdf*

Li, B. (2005), Probabilistic Algorithm to Support the Fingerprinting Method for CDMA location, School of Surveying & Spatial Information Systems. PhD thesis, University of New South Wales, Sydney, Australia. Obtained on 10.11.2006 from *http://www.gmat.unsw.edu.au/snap/publications/lib_etal2005e.pdf*

Li, C. & Maguire, D. (2003), The handheld revolution: towards ubiquitous GIS. In: Longley, P. & Batty, M. (eds), Advanced Spatial Analysis. Redlands: ESRI Press, pp. 193-210.

Lim E. P. & Siau K. (2003), Advances in Mobile Commerce Technologies. London: Idea Group Inc.

Livingston, A. (2004), Smartphones and other mobile devices: the Swiss army knives of the 21st century. Educause Quarterly, vol. 2, pp. 48-52. Paper obtained on 5.11.2006 from *http://www.educause.edu/LibraryDetailPage/666&ID=EQM0425*

Lopez, X. R. (2004), Location-Based Services. In: Karimi, H. A. & Hammad, A. (eds), Telegeoinformatics: Location-Based Computing and Services. Boca Raton: CRC Press. Chapter 6, pp. 172-188.

Macaulay, L. (1994), A Human Centred Approach to Requirements Specification. In: Schmidt, F. et al. (eds), Computer Integrated Production Systems and Organizations: the Human-Centred Approach.Berln etc.: Springer-Verlag.

Markoulidakis, J. et al. (1997), Mobility Modeling in Third-Generation Mobile Telecommunications Systems. In: IEEE Personal Communications, vol. 4, no. 4, pp. 41.56.

Mennecke B. E. & Strader T. J. (2003), Mobile Commerce: Technology, Theory, and Applications. Lndon: IRM Press.

Mitrovic, D. (2001), Machine Learning for Car Navigation. In: Proceedings of the 14th Conference on Industrial and Engineering Applications of Artificial Intelligence (IEA/AIE 2001), Budapest, Hungary, 4 -7 June 2001. Berlin etc.: Springer-Verlag.

Nagi, R. S. (2004), Cartographic visualization for mobile applications. MSc thesis, International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, the Netherlands, December 2004. Obtained on 2.11.2006 from *http://www.itc.nl/library/Papers_2004/msc/gfm/nagi.pdf*

Navalkar, A. (2004), Usability Engineering – Quality Approach (ISO 13407). Paper obtained on 17.11.2006 from *http://www.humanfactors.com/downloads/documents/UsabilityISO.pdf*

Ng'ang'a S. M. (2005), New Directions for Coastal and Marine Monitoring: Web Mapping and Mobile Application Technologies. In: Bartlett, D. & Smith, J. (eds), GIS for Coastal Zone Management. Boca Raton: CRC Press. Chapter 6, pp. 63-76.

Nielsen, J. (1993), Usability Engineering. San Francisco: Morgan Kaufmann Publishers Inc.

Nielsen, J. (1994), Estimating the number of subjects needed for a think aloud test. International Journal of Human-Computer Studies, vol. 41, no. 3, pp. 385-397.

Nielsen, J. (2003), Usability 101: Introduction to Usability. Obtained on 10.11.2006 from *http://www.useit.com/alertbox/20030825.html*

Nivala, A-M. (2005), User-Centered design in the development of a Mobile Map application. Licentiate Thesis, Helsinki University of Technology, Finland.

Nivala, A-M. & Sarjakoski, L.T. (2003), An Approach to Intelligent Maps: Context Awareness. In: Workshop "HCI in mobile Guides", 8 September 2003, Udine, Italy.

Nivala, A-M. & Sarjakoski L.T. (2005), Preventing Interruptions in Mobile Map by Personalisation. Department of Geoinformatics and Cartography, Finnish Geodetic Institute, Masala, Finland.

Nyerges, T.L. (1997), Unit 007 - Asking Geographic Questions, NCGIA Core Curriculum in Geographic Information Science (study material). Obtained on 14.11.2006 from *http://www.ncgia.ucsb.edu/giscc/units/u007/u007_f.html*

Düpmeier, C. et al. (2005), Using SVG-based Maps for Mobile Guide Systems: A Case Study for the Design, Adaptation, and Usage of SVG-based Maps for Mobile Nature Guide Applications. In: Proceedings of the 4th Annual Conference in scalable Graphics (SVG Open 2005), 15-18 August 2005, Enschede, the Netherlands.

Peng, Z. R. & Tsou M. H., (2003), Internet GIS: Distributed Geographic Information Services for the Internet And Wireless Networks. Hoboken: Wiley & Sons.

Po, S. et al. (2004), Heuristic Evaluation and Mobile Usability: Bridging the Realism Gap. In: Proceedings of the 6th International Symposium in Mobile Human-Computer Interaction (Mobile HCI 2004), Glasgow, UK. Berlin etc.: Springer-Verlag.

Ramm K. et al. (2006), Map-Based Positioning using Mobile Phones. In: Proceedings of the XXIII FIG Congress, Munich, Germany, 8-13 October, 2006.

Rao K. R. et al. (2006), Introduction to Multimedia communications: Applications, Middleware, Networking. Hoboken: Willey & Sons Inc.

Redish, J. C. & Dumas J. S., (1999), A Practical Guide to Usability Testing. Exeter: Intellect Books.

Reichenbacher, T. (2001), Adaptive Concepts for a Mobile Cartography. In: Supplement Journal of Geographic Sciences, vol. 11, pp. 43-53. Paper obtained on 15.10.2006 from *http://129.187.175.5/publications/reichenbacher/adaptive_ag_2001.pdf*

Reichenbacher, T. (2004), Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices. PhD Dissertation, Department of Cartography, Technical University of Munich, Germany. Obtained on 15.11.2006 from *http://tumb1.biblio.tu-muenchen.de/publ/diss/bv/2004/reichenbacher.pdf*

Reichenbacher T. & Töllner, D. (2004), Design of an adaptive mobile geovisualisation service. In: 2nd Symposium on Location Based Services & TeleCartography, Vienna University of Technology, 28-29 January 2004, pp. 17-23.

Reilly D. et al. (2006), Marked-up Maps: Combining Paper Maps and Electronic Information Resources, Personal and Ubiquitous Computing, vol. 10, no. 4, pp. 215-226.

Ross, T. & Burnett, G. E. (2001), Evaluation of the Human-Machine Interface to vehicle navigation systems as an example of ubiquitous computing. In: International Journal of Human-Computer Studies, vol. 55, no. 4, pp. 661-674.

Rosson M. B. & Carroll J. M. (2002), Usability Engineering: scenario-based development of human-computer interaction. London: Academic Press.

Salmon, P. M. & Stanton N. A (2005), Human Factors Methods: a practical guide for engineering and design. Hampshire: Ashgate Publishing Limited.

Sarjakoski, L. T. & Nivala, A-M. (2005), Adaptive Visualization of Landmarks. In: Meng, L. et al. (2005), Map - based Mobile Services: Theories, Methods and Implementations. Berlin etc.: Springer.

Sarjakoski, L. & Sarjakoski, T. (2005), A use case based mobile GI service with embedded map generalization. Finnish Geodetic Institute, Masala, Finland.

Sarjakoski, T. & Sarjakoski, L.T. (2005), The GiMoDig public final Report. Finnish Geodetic Institute, Masala, Finland. Paper obtained on 02.11.2006 from *http://inspire.jrc.it/uploads/reference/GiMoDig-D1231-Final-Report-public.pdf*

Shamsi, U.M. (2005), GIS Applications for Water, Wastewater, and Stormwater Systems. Boca Raton: Taylor & Francis CRC Press.

Stanton, A. N. et al. (2005), Human Factors Methods – A Practical Guide for Engineering and Design. Hampshire: Ashgate Publishing Limited.

Streefkerk, J. W. et al. (2006), Selecting, Combining and Tuning Evaluation Methods for Context-Aware Mobile User Interfaces for Professionals, TNO Defense, Security and Safety, Soesterberg, the Netherlands, Delft University of Technology, Delft, the Netherlands.

Van Elzakker, C.P.J.M. (2004), The use of maps in the exploration of geographical data. In: Cartographic Renaissance, Proceedings of the 21st International Cartographic Conference (ICC 2003), Durban, South Africa, 10 - 16 August 2003. CD-ROM, pp. 1945-1955. International Cartographic Association.

Van Elzakker, C.P.J.M & Wealands, K. (2007), Use and Users of Multimedia Cartography. In: W. Cartwright, M.P. Peterson & G. Gartner (eds), Multimedia Cartography, Edition 2, chapter 33. Berlin etc.: Springer Verlag.

Van Oosterom, De Vries, M. & Meijers, M. (2006), Vario-scale data server in a web service context. In: Workshop of the ICA Commission on Map Generalisation and Multiple Representation, 25 June 2006. Obtained on 20.12.2006 from *http://aci.ign.fr/Portland/paper/ICA2006-VanOosterom2.pdf*

Wiegand, P. (2006), Learning And Teaching With Maps. London: Routledge.

Williams, J. R. (2004), Developing Performance Support for Computer Systems: A Strategy for Maximizing Usability and Learnability. Boca Raton: Taylor & Francis CRC Press.

Zadorozhny, V. I. & Chrysanthis, P. K. (2004), Location-Based Computing. In: Karimi, H. A. & Hammad, A. (eds), Telegeoinformatics: Location-Based Computing and Services. Boca Raton: CRC Press.

Zhang, D. & Adipat B. (2005), Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. In: International Journal of Human-Computer Interaction, vol. 18, no. 3, pp. 293-308.

7.2. URLs

URL_1.1: http://www.alicell.com/images/tomtom/550_SonyEricsP910.jpg

URL_1.2: http://www.blueunplugged.com/xsdbimgs/tomtom-screenshots.jpg

URL_1.3: http://www.archimuse.com/mw2005/papers/arts/arts.fig1.jpg

URL_1.4: http://www.handheld-pcs.com/products/116/arcpad-large.gif

URL_1.5: http://www.gdmc.nl/uwsm2/

URL_3.1: http://usability.about.com/od/aboutusability/a/usability.htm

URL_3.2: http://www.usabilitynet.org/tools/r_international.htm

URL_3.3: http://www.usabilityhome.com/

URL_3.4: http://experiencedynamics.blogs.com/site_search_usability/2005/01/latest_research.html

URL_3.5: http://www.cs.auc.dk/~jesper/html/projects.html#trammate

Appendix A: UWSM2 use case 1 scenario

A woman is planning to visit the Rijksmuseum in the city of Amsterdam in the Netherlands. She is going there by driving her car, thinking to park in the Museumplein parking. Unfortunately, this parking has no free space, so she wants to find another one, not far from the museum, by using a mobile geo-application running on her mobile device. Information about the cost of the alternative parking place is also of interest to her. The system has made earlier a user profile of her by asking several questions, such as her age and physical endurance, concluding that she can walk up to a 3km distance during good weather and 1km during bad one.

Although the system through an online weather forecast service then notify her for possible rain, she informs the system that she still likes to check the parking places within a distance of 3km from the museum. The response of the system is to show her the parking places around the museum together with their tariffs per hour, where the cheaper ones are at the other side of the street named Baerlestraat. She takes an overview of the map and the route / distance to that place and chooses to follow it. Navigation-supporting information, such as one-way road is given through the map to assist her in her driving.

When she arrives at the parking, she asks the system to inform her about the closest ticket machine to her, and it responses by showing in the map where this machine is, together with the nearest house address to it. After she gets the parking ticket, she enters the car's location to find it again easily and in order to know her position in relevance to the Rijksmuseum and what should be the route to go there she again uses the map.

After visiting the museum, she goes for shopping, walking much far from it and she decides finally to return to her car. Now she just wants to know the direction of where the car is in the horizon, and not a map of the area between her and the car. By walking she finds the car and driving out of Amsterdam.

Appendix B: Users' pre-selection questionnaire

Part 1: Introduction

Dear candidate survey participant,

My name is Ioannis Delikostidis and I am a Geoinformatics M.Sc. student in ITC.

During my Thesis research, with the subject of Usability Testing for Mobile Geo-applications, I am building a methodology for the usability testing of mobile geo-applications using field surveys. Usability is generally the effectiveness (the capability of the product to enable users completing missions and meeting objectives with accuracy and completeness), efficiency (the performance of the users in completing tasks in relation to the resources spent for this process) and user satisfaction (are the users satisfied through the use of the product?).

These surveys will involve users categorized in 3 different groups of 6 people each, with a different combination of usability testing methods applied to each group.

I am asking for your kind participation in my research, which is the first of this kind for ITC, and its results may be applied in a larger research project of the Netherlands on "Usable (and well scaled) mobile maps for consumers". WWW site: http://www.gdmc.nl/uwsm2/

This questionnaire is aiming at selecting users that can be fitted in homogeneous groups that can be comparable to each other.

The users that will be selected to participate in the surveys will carry-out tasks, as for example to successfully navigate to a point on the map with the use of a PDA/GPS/mobile mapping application combination. These tasks will be clearly given to the users, and the usability of specific functions of the mapping application will be measured, (through for example the time that the user needs in order to complete a task) mostly related to appropriate scaling / zooming / orientation of the map. Remote Observation and video recording of the user will be the most important technique that will be used, together with Think-Aloud, Interview and Questionnaires, depending on the group that the participant will belong after the forming of homogenous groups according to this questionnaire.

All the surveys will take place near Enschede, in a pre-selected area that should not be familiar to the user. The users will be transported in the field survey area by car and brought back after the end of each survey. Each user will not spend more than 1.5 hours total time, including the transportation. He/she will participate only once.

All the personal information provided here will be strictly kept confidence, and only code names will be used for the users (A1, B2, etc.)

Part 2: Questions

1. Please indicate:

(a) Your name / surname:

____

(b) Your age:

____

(c) Your gender: male female

(c) Your profession:

____

(d) Your current studies' field*:

____

(e) Your previous studies' field*:

____

(f) How long have you been in the Netherlands? ______________________________________

2. Do you have knowledge of GPS systems? Yes No

3. Do you have practical experience with GPS systems? Yes No

4. If you answered Yes to question 3, how, would you rate your knowledge on this field?

Poor Modest Very good

5. Do you have practical experience with paper maps? Yes No

6. If you answered Yes to question 5, how would you rate this experience?

Poor Modest Very good

7. Do you have knowledge of digital maps? Yes No

8. If you answered Yes to question 7, how would you rate your knowledge on this field?

Poor Modest Very good

9. Do you have practical experience using digital maps? Yes No

10. If you answered Yes to question 9, how would you rate your knowledge on this field?

Poor Modest Very good

11. Do you have practical experience using PDA devices? Yes No

12. If you answered Yes to question 11, how would you rate this experience?

Poor Modest Very good

13. Do you have experience with mobile navigation applications? Yes No

14. If you answered Yes to question 13, how would you rate this experience?

Poor Modest Very good

15. Are you familiar with the area of Lonneker North-East of Enschede? Yes No

16. Are you familiar with the area of Glanerbrug South-East of Enschede? Yes No

16. Do you agree to participate in my research field surveys? Yes No

17. If you answered Yes to question 16, please indicate a phone number or / and an e-mail address through which I can contact you in order to arrange a survey date / time comfortable for you:

____

____

* (If applicable)

In case you have any questions you can contact me at: delikostid13285@itc.nl or: (Address), (tel.)

Thank you very much in advance for your time and consideration.

Your participation will be much appreciated.

Regards,

Ioannis Delikostidis

Appendix C: Interview questions

1. Could you easily understand your position in the city through the map?
2. Is zooming function helpful?
3. Did you use zooming in order to better understand your position?
4. When you are zooming out, do you still understand where you are in the map?
5. Did you use zooming in order to better understand the POIs in the city?
6. Is information displayed clear in different zoom levels?
7. Is overall map view of your visiting points adequate?
8. Are the sizes of the symbols adequate in different zoom levels?
9. Do you get more info when you zoom-in and less when zooming out?
10. Would you like more info when you are zooming in?
11. Are you pleased with the zooming function?
12. Is orientation function helpful?
13. Are you pleased with the orientation function?
14. Is rotation function useful?
15. Are you pleased with the rotation function?
16. Is panning function helpful?
17. Are you pleased with the panning function?
18. Is map helpful in such a small screen?
19. Do you prefer 2-D, 3-d mapping or both?
20. What are the deficiencies of this map, if any, that includes zooming / panning / rotation / orientation functions?
21. What you didn't like in the functions that you tried on this application?
22. What other info you would like to be included in the map?
23. What other suggestions you have, if any, in order to express your needs better through this mobile map?