The Cartographic Journal *Vol. 45 No. 2 pp. 139–149 Use and Users Special Issue 2008*

REFEREED PAPER

Field-Based Usability Evaluation Methodology for Mobile Geo-Applications

Corné P. J. M. van Elzakker[1], Ioannis Delikostidis[1] and Peter J. M. van Oosterom[2]

[1] Department of Geo-Information Processing, ITC, P.O. Box 6, 7500 AA Enschede, The Netherlands, [2]Section GIS Technology, OTB, Delft University of Technology, P.O. Box 5030, 2600 GA Delft, The Netherlands
Email: elzakker@itc.nl

An important aspect of a Dutch research project into usable (and well scaled) mobile maps for consumers is presented: the development of an appropriate field-based usability evaluation methodology for the prototype of a geo-mobile application that will be the result of a user centred design approach. Automatic generalisation, required for the user's orientation in space, but also for progressive data transfer, will be an important aspect of the prototype. What is reported here is an effective and technically unique user research methodology, based on a combination of video observation, thinking aloud and semi-structured interviewing. The experiments that have led to this outcome revealed some interesting usability issues that deserve further investigation.

Keywords: usability evaluation, mobile geo-application, generalisation, user centred design, user research method

INTRODUCTION

The use of geo-information in mobile devices (smartphones, PDAs) is constantly increasing. Map displays play a prominent part in this, although they suffer from limitations set by the small screens, local (outdated) map copies, storage capacity and processing power of the devices. With the availability of high bandwidth wireless connections it should be possible to overcome some of these limitations and up-to-date map displays may be generated at the right level of detail and adjusted to the needs of the users. In this context, a Dutch research project started in 2006 with User Centred Design (UCD), automatic generalisation and mobile geo-applications as keywords. This paper starts with a short description of the nature and objectives of this research project and its achievements so far. Generalisation is particularly relevant for mobile geo-applications. Not only because zooming is a very important method of interaction by users with the map interface, but also because the progressive transfer of geo-information from server to device may be beneficial to usability. Therefore, part of the research deals with the application of the topological Generalised Area Partitioning (tGAP) structure in variable scale data structures suitable for progressive data transfer, touched upon in another section of this paper. This generalisation concept is implemented in a prototype of a mobile geo-application, developed through UCD techniques. The usability of that prototype will be evaluated later in several iterative steps and the main objective of this paper is to present a methodology that may be used for this evaluation. First, a brief overview will be given of some research methods and techniques that have already been used in testing the usability of existing mobile geo-applications as well as those that have not been used before, but demonstrate great potential. An analysis of these methods resulted in several test scenarios which have been put into practice with a mobile geo-application that is comparable with the prototype that is under development in the main research project. The results of the practical testing are presented, but the main outcome is a proposed methodology for field-based usability evaluation of mobile geo-applications, and particularly the interactive map displays thereof.

RESEARCH PROJECT

The research project on 'Usable (and well scaled) mobile maps for consumers' (UWSM2, URL1) is executed by a consortium of research and development organisations (Delft University of Technology, ITC and TNO Defense, Security & Safety), software companies (ESRI and 1Spatial) and end-user organisations (Municipality of Amsterdam and ANWB, the Dutch Automobile Association). The project is partly funded by the Space for Geo-Information innovation programme (RGI, URL2) of the Netherlands' government. The overall objective of the project is to find solutions for multiple/vario-scale representations of geographic information and interaction with the consumers via

DOI: 10.1179/174327708X305139

mobile devices with context-aware, easy to use interfaces in a connected (wireless Internet) client–server setting. As such, two scientific challenges are addressed: automatic generalisation and the human factors aspect of mobile applications. The challenge is to dynamically tailor the human–computer interaction to the user and momentary usage context by, for example, allowing quick access to specific details and supporting adequate browsing behaviour when needed.

The research was triggered by recent mobile map experiences of the end-user organisations in the consortium. For example, the ANWB wants to support several categories of tourists, holiday-makers and day-trippers (pedestrians, cyclists, car drivers, sailors, etc.) with dynamic access to geo-information. Such as: up-to-date information on the travelling infrastructure, relevant landmarks and related descriptive information, route planning and navigation support, etc. The ANWB is already active in the market for mobile publications through a mobile Internet site, but recent experiences with mapping software are disappointing. Thousands of mobile map devices were sold to ANWB members, but an evaluation showed that these devices were not considered very useful as their users 'get lost with it' (quote from the ANWB representative in the first consortium meeting; minutes of 10-04-2006). There appear to be orientation problems because of the small screens and a lack of clear overview. People do want to use these devices, but improvement is required. Therefore, the ANWB is looking for better solutions that – among other things – resemble the experience of using a printed map more. Important factors for such experience are seamless and rapid zooming and panning (and not jumping from one fixed scale to the next and thereby 'losing the user'. In order to publish real-time tourist information it is also essential that the foreseen solution is based on a client–server environment.

For these reasons, the whole research project is driven by sample use cases as suggested by the municipality of Amsterdam and the ANWB, involving user questions like: 'Where can I park my car and what will it cost?' 'Where is the museum?' 'How can I get there?' and 'How can I get back to my car?' These were combined into a scenario: a verbal description of possible uses of the system to be designed. The scenario describes the user and his/her tasks, including his/her goals, desires, and the context of use. It was the starting-point for further specification of the desired functionality through use case and interaction modelling with the help of UML, the Unified Modeling Language.

UWSM2 is organised into a number of Work Packages (phases), following a sound, iterative user centred design methodology (Figure 1). In the course of 2008, successive versions of prototypes will be designed in this iterative process and these prototypes will be evaluated with representative users. This paper proposes a methodology for the usability evaluation of the first prototype.

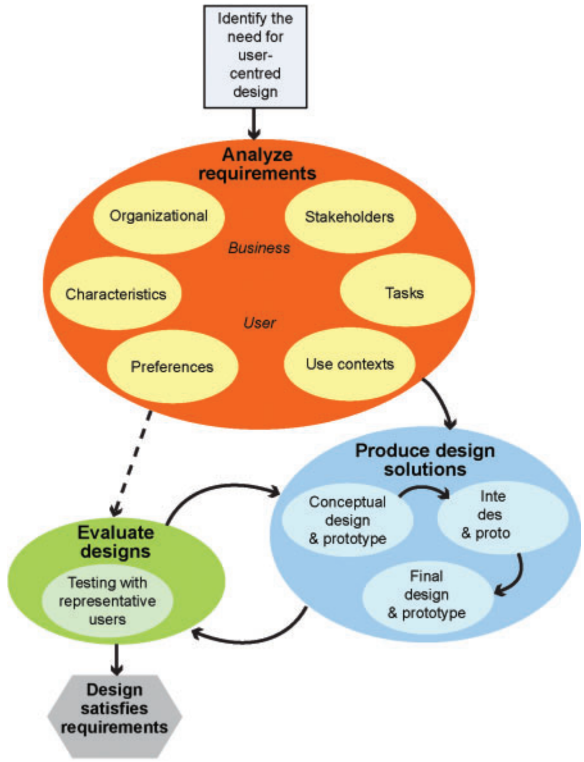

Figure 1. The user centred design process (after Van Elzakker and Wealands, 2007)

GENERALISATION FOR MOBILE GEO-APPLICATIONS

Owing to the small screens of mobile devices, which cannot show much of the map, a user needs to pan and zoom a lot to get an overall spatial understanding of an area. That is, a feeling for the sizes, directions and distances between the relevant objects and their context. However, after a zoom or pan action, a complete redraw is performed in nearly all mobile geo-applications. When this happens the user often loses the 'mental' link between the two map displays. Initial experiences show that users get lost (in the map display) and do not build up a good mental map in order to support their task. Therefore, users may not fully appreciate new applications based on mobile maps. One of the key solutions to this problem of user disorientation is the concept of 'vario-scale maps'. However, data structures supporting variable scale data sets are still very rare. There are a number of data structures available for multi-scale databases based on multiple representations (MRDBs), i.e. data to be used for a fixed number of scale (or resolution) intervals. These multiple representation data structures try to explicitly relate the corresponding objects at the different scale levels, in order to offer consistency during the use of the data. Drawbacks of the multiple representation data structures are that they store redundant data (same coordinates, originating from the same source) and that they support only a limited number of scales. Another drawback of the multiple representation data structures is that they are not suitable for progressive data transfer, as each scale interval requires its own (independent) graphic representation to be transferred.

The proposed solution is the 'topological Generalised Area Partitioning' (tGAP) structure. This data structure can offer vario-scale vector data to mobile users. In earlier research, both the theoretical and practical (implementation) aspects of the tGAP structure have been described (Van Oosterom, 2005; Van Oosterom *et al.*, 2006). The purpose of the tGAP structure is to store the data only once, with no redundancy of the geometry, and to derive different representations of the same data on the fly according to the level of detail required. One way of using the tGAP structure is to produce a continuous range from rough to detailed representations (vario-scale map). With this, a smooth zooming functionality can be offered, which can be realised through progressive transfer. This appears to be a promising solution for mobile maps.

The tGAP structure is first of all intended for (polygon) maps based on a hierarchical area partitioning structure, such as topographic, soil and land use maps. When zooming in, the tGAP is used by going deeper into the tree structure to retrieve more detailed area features. This is also the basic idea behind the progressive transfer support: the top-levels of the structure contain the rough overview representation and are sent first. This will then be refined by the lower levels of the structure which will be sent next. The ordered data from the progressive transfer are the input at the client side to support a smooth zoom visualisation.

Alternative map generalisation procedures try to go from one fixed scale to the next (e.g. from 1:10 000 to 1:25 000) in either a semi-automatic way (using a combination of generalisation tools and a human operator) or through a fully automated approach. Haunert and Wolff (2006) and Haunert (2007) present a combinatorial optimisation method to aggregate area partitions from one specific scale to the next. The quality of these generalised maps is better that standard tGAP maps. However, their approach does not yet result in a vario-scale map as required (and the fixed scales will make users lose their orientation again). Therefore, the two generalisation approaches are combined: first, a fixed-scale generalisation is computed and this generalisation is then used as a constraint when starting again from the largest scale to build the tGAP structure (Haunert *et al.*, forthcoming). The difference from the original tGAP creation procedure is that the smooth generalisation is now constrained to the features that are all part of the same smaller scale feature (previously computed in the preprocessing) that also overlaps with the least important feature. In this manner, the grouping of the tGAP will finally be the same as the combinatorial optimised fixed-scale generalisation. However, the added value of the tGAP structure is that all scales in between are also covered in a variable manner. As an alternative to using a computed fixed scale generalised map as constraint, it is also possible to use an existing smaller scale map as constraint. In the latter case, one only has to take care of sliver processing when overlaying the large-scale feature with the small-scale feature. This new method is called the constrained tGAP. But in the end, the result is a high-quality vario-scale structure enabling the realisation of usable mobile maps. The remainder of this paper is devoted to the way in which such mobile geo-applications could be tested.

PREVIOUS USABILITY TESTING OF MOBILE GEO-APPLICATIONS

Disregarding, for the time being, the distinction between quantitative and qualitative research, two basic usability testing methodologies may be distinguished for the determination of the usability of mobile geo-applications: laboratory-based and field-based. They each have their advantages and disadvantages depending on the functions and aspects of the applications that have to be assessed and the context of use. Considering the special nature of mobile geo-applications, where the user interacts with the device/application and the natural environment at the same time, usability cannot be properly checked by means of controlled laboratory experiments alone. In the laboratory, a big part of the contextual information cannot be investigated and real users' behaviour and activities may not be sufficiently understood (Mennecke and Strader, 2003; Leitner *et al.*, 2006; Kaikokken *et al.*, 2005). Despite this, most studies on the usability of mobile geo-applications are executed in the laboratory, while only 19% is done in the field (Kjeldskov and Graham, 2003; Kaikokken *et al.*, 2005).

For our research, we investigated some earlier characteristic examples of usability testing of mobile geo-applications: the GiMoDig project (Nivala and Sarjakoski, 2005; Sarjakoski and Sarjakoski, 2005), the TIP system (Hinze and Buchanan, 2005) and the Trammate project (Kjeldskov *et al.*, 2005). An UCD approach was followed in the GiMoDig project, involving the assessment of context-aware mobile map-based prototypes. Heuristic and expert evaluations were among the testing methods. Observation techniques were used in the TIP project, but no reference is made to user testing in the real context of use. The motivation behind the Trammate project was the comparison of different usability testing and data collection methods, executed both in the field and in the laboratory. Among the evaluation methods used were heuristic walkthrough, rapid reflection, thinking aloud, video and audio recording, observation and data logging.

Our conclusions were that each of the research methods and techniques applied had the ability to exclusively address specific usability problems; consequently, usually a combination of methods gives the best results. In addition, a combination of laboratory and field testing seems to be the best solution as it provides a deeper investigation of different usability problems than can be found with each approach alone. Laboratory-based testing can thoroughly assess interface and representation usability, whereas field-based testing can uncover user-application interaction in real/natural mobile environments. However, another conclusion was that such usability evaluation of mobile geo-applications requires a large amount of human resources.

POTENTIAL RESEARCH METHODS AND TECHNIQUES FOR USABILITY EVALUATION

Despite this need for significant human resources, for the UWSM2 project, where satisfactory scalability and usability of mobile maps in real contexts of use with real users is the main object of research, a combination of laboratory- and field-based testing should be performed. In view of the fact that so far there has not been much experience with field-based usability evaluation we decided to pay special attention to the usability evaluation methodology to be used in the field and we tested several alternative combinations of research techniques.

For organisational reasons, this testing of research techniques was done before the first prototypes were actually available for testing with representative users. Therefore, first of all, we explored the 'use case' and 'interaction' models and the future prototype functions and we extracted several user tasks from them. These were: the understanding of the geographic location of the user him/herself through the mobile map (we like to call this 'personal geo-identification'); searching for the required points of interest (POIs); receiving online location-based information; planning of routes to the POIs; and navigating to the POIs. Converting these user tasks into questions that the system is aimed at answering, in this way supporting the decision-making procedure of the user, was the next step. Given the aims of the UWSM2 project, we focused on questions related to map scaling features, so that we could assess the clearness, consistency and comprehensibility of the map representation at different scales, the user's easiness of understanding his/her position on the map in relation to the real surroundings at all zoom levels and the helpfulness of the zooming/panning/rotation/orientation functions for better understanding of the map/route (Delikostidis, 2007).

All this means that, with the evaluation of the prototypes in the UWSM2 project, we are aiming at both qualitative and quantitative results. This combination of quantitative and qualitative research is a common characteristic of the user centred design process, whereby qualitative research is implemented to test the first prototype designs and more quantitative research to evaluate the final designs. So far, hardly anything is known about how the cognitive process works for users moving around in the real world with the aid of mobile geo-applications. To help understand this, we first of all require more qualitative user research. But, on the other hand, we also want to have quantitative measures telling us, for instance, how long it takes for a subject to execute a particular use task and how many subjects arrive at the 'correct' answer.

Several tools are available to collect the required primary qualitative and quantitative data from representative users. Van Elzakker (2004) distinguishes the following four main categories: interviews, questionnaires, observation and product analysis. An interview with subjects may take two different forms: structured and unstructured. During an unstructured or in-depth interview – which may also be executed with a focus group of respondents collectively – questions are formulated spontaneously, albeit within an interview framework. The advantage of unstructured interviews is that a lot of in-depth information can be gathered when little is known of the problem-solving process under investigation. But their disadvantage is that it is difficult to compare the answers obtained from different respondents. Comparability of data is more assured with structured interviews and written questionnaires.

In order to obtain useful and comparable results from investigating cognitive processes in complex problem-solving through questionnaires and interviews, it is also required to have the subjects actually execute one or more tasks with the help of the designed prototype. This may lead to 'products', or answers to geographical questions, that can be subject to product analysis. In such a context, it is important to consider the moment at which the data are collected. In the case of a research method called retrospection, the investigator interviews the subject (or lets him/her complete a questionnaire) after the completion of the task. Possible recollection problems may be overcome by questioning and prompting subjects during the problem-solving process. The method is called introspection if the subjects themselves choose the moments at which they report on their cognitive process. The biggest source of distortion and invalidity of the data obtained through retrospection, questioning and prompting during problem-solving, and introspection is the danger that subjects feel inclined to interpret and rationalise their problem-solving behaviour. Another possible problem with interviews and questionnaires is that subjects may be steered too much into directions predicted by the investigator.

These problems do not occur if information is obtained by observing the problem-solving behaviour while it takes place. In simple observation, the investigator may watch the subjects during their execution of tasks and make notes. However, in this way the research results would not be verifiable and, therefore, it is better to make use of special equipment to record the observations. This also makes it possible to carefully analyse the observations at a later stage. Possible recording techniques are: video, eye movements, brain activities and screen logging. One problem with these techniques is that the analysis of the outcomes, just like the analysis of video recordings and computer logs, cannot provide complete in-depth information on the cognitive processes that take place in the minds of the subjects. After all, especially when testing a first prototype, we want to know not only what the test persons are actually doing, but also why they are doing it and what they are thinking. For this purpose, the so-called 'think aloud method', in which subjects are asked to voice their thoughts when executing particular problem-solving tasks, is very appropriate (Van Elzakker, 2004).

A METHODOLOGY FOR TESTING THE FIRST UWSM2 PROTOTYPE WITH REPRESENTATIVE USERS

In our experimental approach, in which we were looking for an appropriate methodology to test the expected first prototype of the UWSM2 project, we first of all considered mainly qualitative testing methods and techniques. The testing would have to be implemented in field-based user surveys to obtain primary data on efficiency, effectiveness and user satisfaction as the fundamental attributes of usability (ISO-9241-11, 1998). The methods that we selected after assessing their advantages and disadvantages against the research objectives were observation, thinking aloud, video/audio recording (including screen logging) and semi-structured interviews. Although all of these methods have been used in the past for the evaluation of

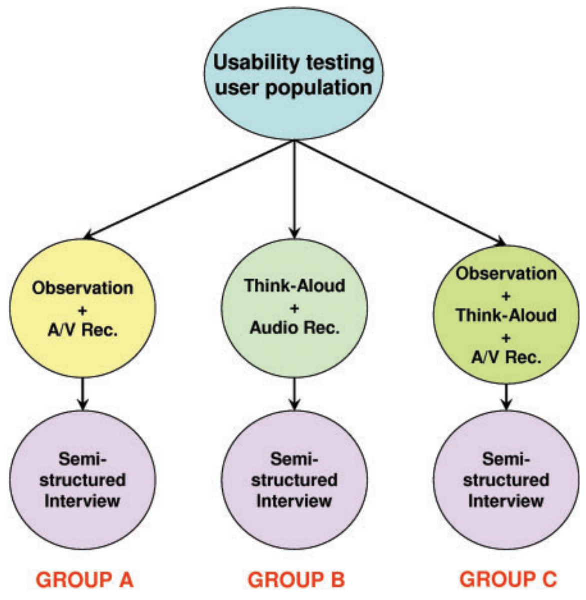

Figure 2. Three different combinations of techniques applied in three test user groups

mobile as well as other geo-applications, there was a need for improvement so that they would better fit into a field-based testing environment, benefiting from the use of advanced technical solutions. Three different combinations of research techniques were created, which were later applied in three test user groups (see Figure 2).

Two main issues were considered during the design of the research experiment. The first was the possible influence of the researcher (observer) on the behaviour of the subject during the test and the second the possible need for more than one researcher to be involved in the test execution. The solution was the technical design and construction of a special field survey system that would enable one researcher to carry out the testing. We were looking for a solution better than just video recording with a camcorder, as the analysis of the test data in that form would be cumbersome. Besides, the researcher would have to be dedicated to using the camcorder, continuously aligning it in order to have a good view of the mobile device display, alternating with views of the environment and the subject. Most likely, this would also pose a significant bias, as the observer would have to stay very close to the user during testing.

The system implemented enables the observer to stay at a fair distance, 10 to 25 m from the test subject, remotely observing, capturing audio and video data and communicating with him/her during the test sessions. The system consists of three black and white wide-angle mini cameras, two pairs of video transceivers, a video quad processor, a handheld video/audio recorder, a laptop, two pairs of modified PMR audio transceivers, a Wi-Fi and Bluetooth-enabled PDA, a Bluetooth GPS receiver and several lithium-ion and Ni–MH batteries (Figure 3).

Two of the cameras were installed on a hat worn by the user, capturing the environment in front of him/her and his/her hands interacting with the mobile device (a HP iPAQ hx4700 PDA on which the mobile geo-application was running). The video signals from the two cameras were transmitted wirelessly to the observer through the pair of video transceivers and another video signal was obtained from the third camera that was carried by the observer on his chest which was capturing the user from a distance (Figure 4).

The laptop with video output capability, carried by the observer in a backpack, was remotely capturing the screen of the PDA through a Wi-Fi connection in near real-time. The three video signals from the cameras together with the PDA screen capture through the video output of the laptop were connected to the quad processor and the output of the quad was connected to the handheld audio/video recorder. The output of the quad processor is a cross-like frame including all four video signals in one screen (Figure 5).

The PMR audio transceivers were used for the remote communication of the observer with the user and for the audio recording of their thinking aloud through the handheld audio/video recorder. Being modified, these transceivers allowed the continuous full-duplex communication between the observer and the user without the need to press the transmitting button of the corresponding transceivers.

Perhaps the biggest advantage of this unique technical solution is the instantaneous synchronisation of the various audio and video recordings, which makes later analysis much easier.

METHODOLOGY PUT TO THE TEST

After the construction of the field survey system, the next step was the selection of test subjects. As this part of the research project was executed in another part of the Netherlands, the target population of the UWSM2 prototype (citizens of Amsterdam and tourists to the city) could not be used. Considering that the main aim of this stage was the comparison of different usability testing methodologies, students from an international educational institute (coming from all parts of the World) were found to be a good alternative. From this pool of test subjects who, like tourists, do not know the survey area, homogeneous and comparable user groups could easily be formed. The selection was made through pre-questionnaires, asking for age, gender and knowledge/experience in different fields such as paper and digital maps, GPS and navigation systems and mobile devices. Based on their answers three similar user groups of six test subjects each were formed, including four men and two women in each group, aging from 25 to 40 (Table 1). It was ensured that the test persons were unfamiliar with the survey area, the village of Lonneker, near Enschede.

In order to assess the three proposed usability testing methodologies for a prototype that had not yet been developed, an existing mobile geo-application with generally comparable functionalities had to be selected. A series of criteria was created, assessing different aspects of the candidate applications, including e.g. smooth zooming functionalities, different amount of detail/information in different zoom levels, zooming/panning/rotation functionalities, availability, and detailed coverage of the study area. Although UWSM2 aims at using advanced online map

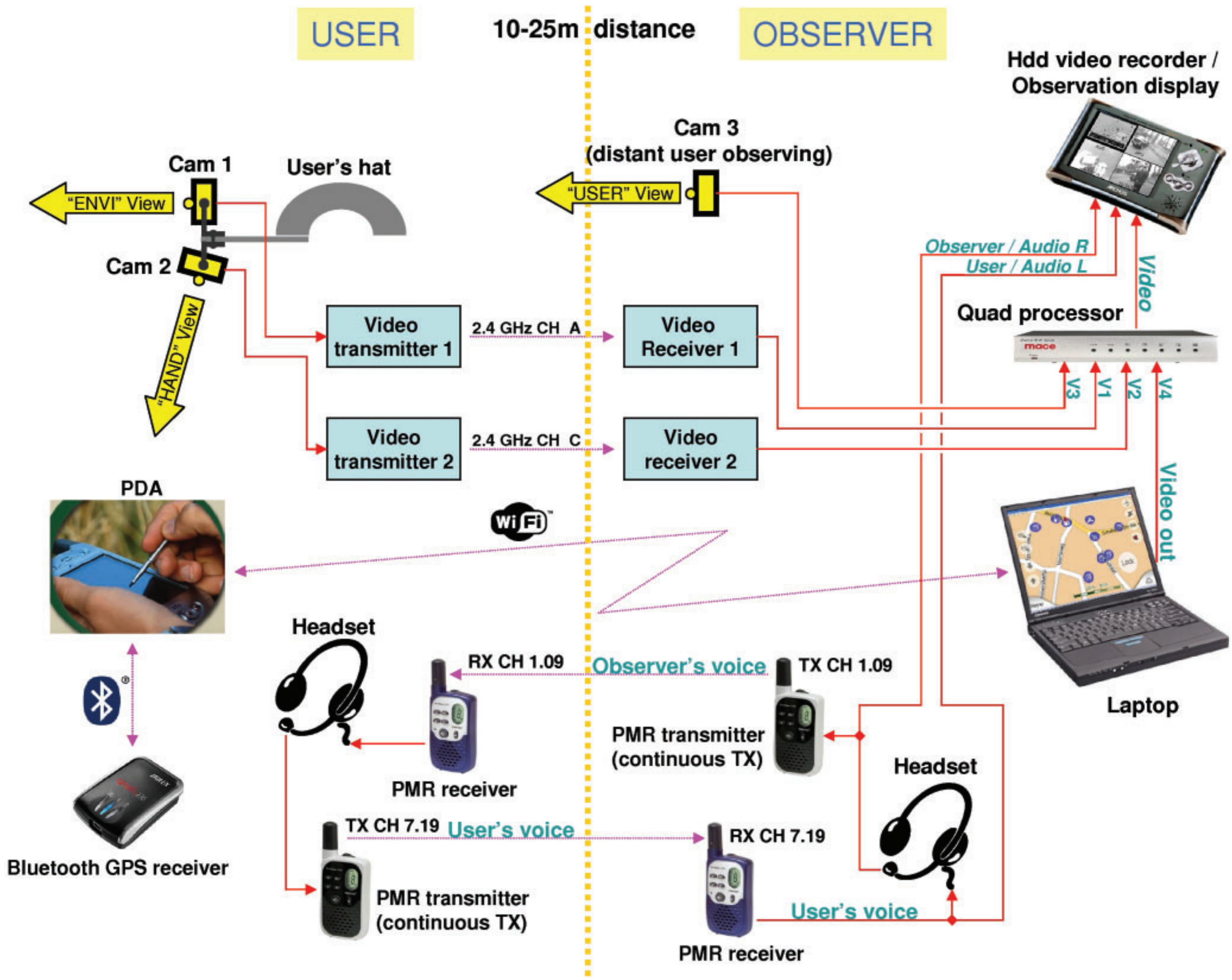

Figure 3. Overview of the field observation system implemented

Figure 4. The researcher carrying the observation/recording equipment and the test subject wearing the hat with the cameras on it

Figure 5. A screenshot of the video recordings using the quad processor output signal

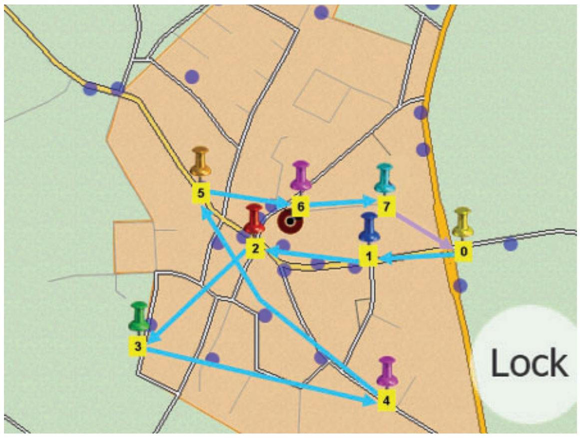

Figure 6. The seven navigation POIs as shown in the mobile map

retrieval techniques and real-time generalisation, in this part of the research we focused on the map scaling functionalities only, using an off-line existing mobile geo-application.

Several popular applications, implemented for car and/or pedestrian navigation were assessed based on the above criteria (see Table 2). In the end, the iGO My way 2006 application was selected for our testing (URL3).

In setting up the tests, efforts were made to keep the dynamic context variables within acceptable limits, including rules for the test execution times, amount of light, weather conditions and environmental noise. The test sessions were scenario-based, including seven navigation tasks in a predefined order to corresponding POIs (see Figure 6).

Differently coloured and clickable pin-points and corresponding icons were used at larger scales to discriminate between and characterise the POIs as well (see Figure 7).

The subjects, representing visitors to an unfamiliar city, were supposed to use the functions of zooming, panning, rotation and orientation in order to successfully navigate to these POIs.

Table 1. The composition of the three test groups

Test subject groups			Knowledge/experience				
Group A	Group B	Group C	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 2. Comparison of existing geo-mobile navigation applications

	TomTom v.6 Western Europe	Route 66 Navigate v.7 Europa	Destinator v.6 Western Europe	Marco Polo v.3 Europe	iGO My way 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 capabilities	NO	NO	NO	NO	NO
Indicative cost	156 EURO	209 EURO	193 EURO	170 EURO	170 EURO
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

Figure 7. POI icons, pin-pointed POIs and information pop-ups

OVERALL USABILITY EVALUATION RESULTS FOR THE TESTED APPLICATION

Although evaluating the mobile geo-application iGO My way 2006 was not the main objective of this research, the overall outcomes, in terms of effectiveness, efficiency and user satisfaction, are worth mentioning.

For (indirectly) measuring the usability attribute of efficiency of the mobile map interface, the time consumed for each navigation task was recorded (see Table 3 and Figure 8).

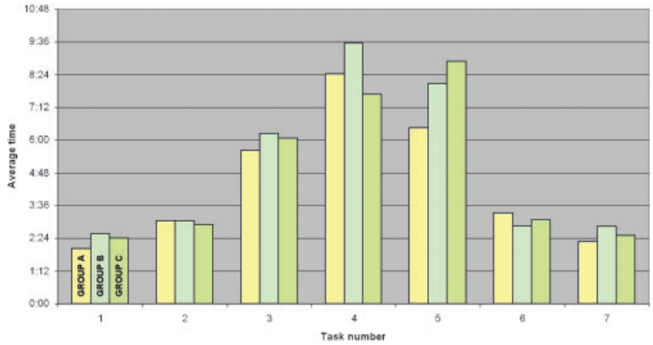

Figure 8. The three test groups compared with respect to efficiency

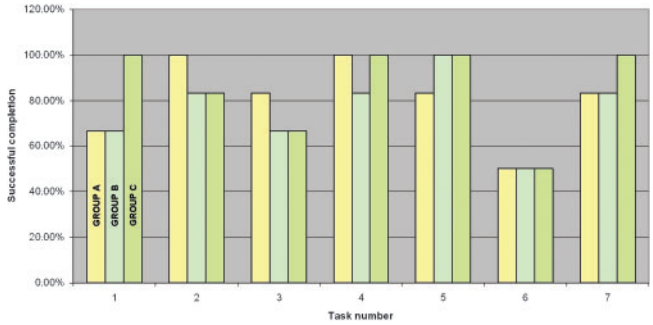

Figure 9. The three test groups compared with respect to effectiveness

Table 3. Timings of each subject in task execution

Group	User	Task 1 time	Task 2 time	Task 3 time	Task 4 time	Task 5 time	Task 6 time	Task 7 time	Tasks total time	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

Notes
Timings expressed in min:s
x = POI not found within the time limit

Table 4. Navigation task parameters (distance between POIs and set time limits)

Task	From	To	Distance (m)	Time (min)
1	Point 0 (car)	Point 1 (Rabobank ATM)	140	4
2	Point 1 (Rabobank ATM)	Point 2 (Café Sprakel)	190	4
3	Point 2 (Café Sprakel)	Point 3 (lake)	340	8
4	Point 3 (lake)	Point 4 (church)	600	14
5	Point 4 (church)	Point 5 (Vakdrogist – post box)	440	12
6	Point 5 (Vakdrogist – post box)	Point 6 (map orientation panel)	220	4
7	Point 6 (map orientation panel)	Point 7 (kindergarten)	120	3
	Total		2050	49

Table 5. Comparison of the three usability evaluation methodologies (note: audio recordings and semi-structured interviews were executed with every method)

Group A		Group B		Group C	
Methods: Observation + A/V recordings		Methods: Thinking aloud + Audio recordings		Methods: Observation + Thinking aloud + A/V recordings	
Advantages	Disadvantages	Advantages	Disadvantages	Advantages	Disadvantages
With direct observation researcher gets immediate impression of problems in task execution	Results of direct observation are not verifiable	Thinking aloud gives more insight into cognitive processes of the subjects during task execution	Audio recordings of the thinking aloud alone (i.e. without video) do not give the best impression of the problems with task execution	Methodology gives the best and most complete information about the cognitive processes of the subjects during task execution	From a technical point of view: the most complex methodology. Possible problems with battery power. System requires thorough testing before and during every test session
Video recordings allow for verification, confirmation and thorough analysis of user actions afterwards (also by other researchers)	Technically complex (as complex as the methods applied with Group C) and problems with battery power	No direct observation; user not influenced by observer	Observer may not prompt subject to think aloud	A/V recordings allow for thorough and verifiable analysis of research outcomes afterwards	Analysis of research outcomes is cumbersome and time-consuming
Audio recordings of remarks made by subjects or observer may help in the analysis of research outcomes	If subjects are not asked to think aloud, audio recordings do not give complete insight in users' thoughts during problem-solving process	Audio recordings allow for verifiable analysis of research outcomes afterwards	Interpretation of thinking aloud difficult without video recordings of subjects' interaction with device and the environment, screen logging and 'body language'	With direct observation researcher gets immediate impression of problems in task execution. Observer can also immediately deal with (technical) problems occurring during test execution and may prompt subjects to think aloud	Subjects may be influenced by observer
		From a technical point of view: easier to implement. Long battery power	In the absence of an observer: technical problems may only be discovered afterwards		

The total session times in the last column of Table 3 are not just the net sum of the execution of the seven tasks, but also include the time required for initial briefing/ debriefing, transportation to and from the testing area, preparing equipment and interviewing. The high values of total session times in test groups A and C can be explained by Wi-Fi disconnections, which were repaired in the course of the research by installing a longer range Wi-Fi card in the laptop and some other system improvements. The absence of high total session duration values in test group B can be explained, of course, by the fact that no video recordings were used with this test group. After the improvement of the field observation system there appeared to be no significant differences in the total session time required by the three test groups. Another interesting conclusion from studying Table 3 is the difficulty subjects had in finding POI 6. This was due to an unclear task description.

Effectiveness was measured as the percentage of the successful completion of each task by each user within predefined time limits, which were experimentally determined by pre-testing pilot user surveys (see Table 4 and Figure 9). Among other things, Figure 9 shows that there is no statistically significant difference in the success with which the three test groups have completed the seven tasks. Efficiency (see Figure 8) was measured by the mean time needed for the completion of a task, related to the success percentage achieved. Also here there are no statistically significant deviations between the three test groups. It shows that the selection of subjects for the three different test groups was appropriate and the results of the comparison valid.

User satisfaction was established through post-survey semi-structured interviews, for which almost 10 min were required on average. Generally, the subjects were satisfied with the map scaling of the mobile geo-application tested, the information available in different zoom levels and the functions of zooming, panning rotation and orientation of the map. However, reasons for confusion were the inability of the mobile map to be oriented towards the actual viewpoint of the user while he/she is standing still and the absence of properly placed landmarks on the map. Indeed, one of the most interesting findings was that more than one third of the test subjects navigated incorrectly because of initial misunderstanding of their location. These same users, while having failed or spent a lot of time executing the first task, performed better in the remaining ones. This probably demonstrates an ineffective interaction between the cognitive mapping of the users and the geo-communication provided by the mobile application. This aspect will be investigated further in a separate research project.

In addition, the natural settings of the user surveys in this research project led to the discovery of several practical difficulties in the execution of field-based studies. Inconvenient weather conditions, local residents' disturbance and shortcomings of the electronic equipment were among the difficulties we were confronted with during the experiment. These can be minimised by careful time planning, notification of local communities and thorough pilot testing of the equipment. Despite these difficulties, field-based surveys proved to be an effective tool for the investigation of mobile user problems in real contexts.

PROPOSED METHODOLOGY FOR TESTING THE USABILITY OF WELL-SCALED MOBILE MAPS

The main objective of the research reported on in this paper was to find an appropriate field-based research methodology to evaluate the first prototype of the UWSM2 project. The results of the comparison of three combinations of techniques with three comparable test groups are summarised in Table 5. Not surprisingly, it was found that the combination of observation through video recordings, screen logging, synchronous audio recording of the thinking aloud and semi-structured interviewing with the developed remote field observation system, as applied with test group C, appeared to be the most promising methodology to be used for the evaluation of the UWSM2 project prototype. It investigates several usability problems, potentially gives most information about the use and the user and allows for a deeper analysis of the results. At the same time, the influence on the subject is kept low, as the observer remains relatively invisible to the subjects during the tests. Our research has found that technical and practical issues need not stand in the way of obtaining the research outcomes required.

The methodology proposed can be further extended with more advanced functions and techniques like automated analysis of the results, or alternative techniques such as post survey questionnaires, GPS data logging (time and location, implying speed, turning, slowing down, etc.) and remote PDA data logging.

CONCLUSION

This paper presents the findings from the investigation into a feasible field-based usability evaluation methodology for a mobile geo-application in which generalisation, as a consequence of smooth zooming, plays a prominent role. The proposed methodology consists of a combination of common research techniques and includes a new technical solution for field-based qualitative usability research, which can be used to evaluate the prospective UWSM2 prototype but, no doubt, other mobile geo-applications as well.

Further research is required, within or outside the UWSM2 project. For instance, the findings of the actual user tests in this project brought up the relevance of further investigating the personal geo-identification problem, related to the users' question 'Where am I?' and the possible ways to support answering it, before continuing to the exploration of further geographic questions and their corresponding tasks. Whether the exploitation and application of cognitive map models, use of more efficient context-based overview/focused viewing transitioning techniques and provision of different types of information to the user, such as video, photos, text, virtual reality representations, or combinations of them could be the solution, is clearly something to be investigated.

BIOGRAPHICAL NOTES

Dr Corné P. J. M. van Elzakker is an Assistant Professor in the Department of Geo-Information Processing at the International Institute for Geo-Information Science and Earth Observation (ITC). He studied Human Geography with cartography as main subject at Utrecht University. His main research interests are in the field of use and user issues in geo-information processing and dissemination. He did his PhD research into the use of maps in the exploration of geographical data, chairs the ICA Commission on Use and User Issues as well as the Project Group on the Use of Geo-Information of the Section of Cartography and Geovisualisation of GIN (Geo-Information Netherlands).

URLs CITED

URL 1: UWSM2, the research project's website <http://www.gdmc.nl/uwsm2/> (last accessed 14/03/2008).
URL 2: RGI, Space for Geo-Information innovation programme <http://www.rgi.nl/?l=eng> (last accessed 14/03/2008).
URL 3: iGO My Way navigation software <http://www.i-go.com/en/> (last accessed 14/03/2008).

REFERENCES

Delikostidis, I. (2007) Methods and techniques for field-based usability testing of mobile geo-applications, MSc thesis, International Institute for Geo-Information Science & Earth Observation (ITC) Enschede, the Netherlands. Available online at: <http://www.itc.nl/library/Papers_2007/msc/gfm/delikostidis.pdf> (last accessed 14/03/2008).

Haunert, J-H, Dilo, A. and Van Oosterom, P. (forthcoming) Using the optimised constrained tGAP for progressive vector data transfer. Submitted to **Computers & Geosciences**.

Haunert, J.-H. (2007) Efficient area aggregation by combination of different techniques. In: Proceedings of 10th ICA Workshop on Generalisation and Multiple Representation, 2–3 August 2007, Moscow, Russia.

Haunert, J.-H. and Wolff , A. (2006) Generalisation of land cover maps by mixed integer programming. In: Proceedings 14th International ACM Symposium. Advances in Geographic Information Systems (ACM-GIS'06), pp. 75–82.

Hinze, A. and 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. <http://www.cs.waikato.ac.nz/~hinze/isdb/publications/hinze_buchanan_CONTEXTmHCI.pdf> (last accessed 02/11/2006)

ISO-9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDTs) – Part II, Guidance on usability. International Organisation for Standardisation.

Kaikokken A. *et al.* (2005) Usability testing of mobile applications: a comparison between laboratory and field testing. **Journal of Usability Studies**, 1(1), 4–16.

Kjeldskov, J. and 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: Springer–Verlag, pp. 317–335.

Kjeldskov et al. (2005) Evaluating the usability of a mobile guide: the influence of location, participants and resources. **Behaviour and Information Technology (BIT)**, 24, 1, 51–65.

Leitner, G., Plattner, S. and 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.

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

Nivala, A-M. and Sarjakoski L. T. (2005) Preventing interruptions in mobile map by personalisation. Department of Geoinformatics and Cartography, Finnish Geodetic Institute, Masala, Finland.

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

Van Elzakker, C. P. J. M. (2004) **The use of maps in the exploration of geographic data.** Utrecht / Enschede: KNAG / Faculteit Geowetenschappen Universiteit Utrecht / ITC. Netherlands Geographical Studies 326.

Van Elzakker, C. P. J. M. and Wealands, K. (2007) Use and users of multimedia cartography. In: **Multimedia Cartography,** Second Edition, ed. by Cartwright, W.; Peterson, M.P. and Gartner, G., Chapter 34, pp. 487–504. Berlin etc.: Springer.

Van Oosterom, P. (2005) Variable-scale topological data structures suitable for progressive data transfer: the GAP-face tree and GAP-edge forest. **Cartography and Geographic Information Science**, 32(4), 331–346.

Van Oosterom, P., de Vries, M. and Meijers, M. (2006), Vario-scale data server in a web service context. In the Workshop of the ICA Commission on Generalisation and Multiple Representation, June 25th 2006.