SPATIAL-AWARE MOBILE SYSTEMS: EXPLORING USABILITY EVALUATION ISSUES

Tasos Spiliotopoulos, Panagiota Papadopoulou, Drakoulis Martakos

National and Kapodistrian University of Athens,
Department of Informatics and Telecommunications,
Panepistimiopolis, Ktiria Pliroforikis,
Athens, 157 84, Greece
{tasos, peggy, martakos}@di.uoa.gr

Abstract
This paper attempts to explore the issues related to the usability evaluation of spatial-aware mobile systems used for technologically mediating interpersonal and group communication. In this direction, it investigates the applicability of existing usability evaluation methods and principles on spatial-aware mobile systems and presents the results of a multi-method usability evaluation of such a system. The evaluation methodology comprises questionnaires, a word list, focus groups and actual logging of use, in order to evaluate the user experience during field tests with the system.

Résumé
Cet article essaye d'explorer les issues liées à l'évaluation d'utilisabilité des systèmes mobiles spatial-avertis utilisés pour technologiquement la médiation et la communication interpersonnelle et de groupe. Dans cette direction, il étudie l'applicabilité des méthodes et des principes existants d'évaluation d'utilisabilité sur les systèmes mobiles spatial-avertis et présente les résultats d' une évaluation d'utilisabilité étendue et de multi-méthode d'un tel système. La méthodologie d'évaluation comporte des questionnaires, une liste de mots, des focus groupes et la notation réelle de l'utilisation, afin d'évaluer l'expérience d'utilisateur pendant une série d'essais pratiques avec le système.

Keywords: usability, evaluation, field testing, spatial-aware, mobile systems

Introduction

As the computational power of mobile devices increases, it is possible to shift more functionality from traditional computers to such devices, as well as to enable new paradigms for communicating and disseminating information. These new paradigms require that mobile computing devices act as a link between the physical and the digital world. This requirement has raised the issue of combining virtual information with the user's direct surroundings, thus leading to the development of spatial-aware mobile systems. Location-based applications and services take advantage of the spatial-awareness of mobile devices, in order to adapt to the user context and provide more customized and efficient functionality to the users. The three main categories of use that spatial-aware

mobile systems focus on are orientation and wayfinding, access and creation of spatial data, and augmented reality (Froelich et al., 2008).
Mobile devices have special characteristics which pose challenges in designing usable applications and services for mobile systems. These characteristics may include the small screen size, low computational power, small interaction elements (e.g. buttons, joystick), poor communication and networking ability, limitations regarding graphical representation capabilities, and, most importantly, the diverse and unsteady context of use, since mobile systems can be used by people on the move, outdoors or while engaged in other activities.
In addition, when designing mobile systems which are spatial-aware, several other issues have to be taken into consideration, such as security, social disclosure and privacy. The design of spatial-aware mobile systems is a new and emerging research field, and as such researchers and practitioners have been more concerned with exploring the capabilities of such systems than focusing on their ease of use. System usability, defined as "*The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*" (ISO 9241-11, 1998), is regarded as one of the aspects of software quality, and as such it can be described alternatively as "*quality in use*". (Bevan, 2001). Usability evaluation has proven to be an invaluable tool for ensuring the quality of computer systems, in general.
This paper attempts to explore the issues related to the usability evaluation of spatial-aware mobile systems and presents the results of such an evaluation on a system that combines all of the three aforementioned categories of use of spatial-aware mobile systems, but focuses on mediating interpersonal and group communication and social interaction. The contribution of this paper for both researchers and practitioners lies in the provision of important issues that need to be considered for the usability evaluation of spatial-aware mobile systems.

The Prototype System

LOCUNET (LOCation-based Urban NETwork) is a research project that aims to investigate the social dimension of mobile locative media use in a modern urban setting by studying the properties and characteristics of communication during locative media use.
A functional spatial-aware mobile system prototype has been designed and developed for this purpose and used on a structured game-like activity that took place in the historical center of Athens, Greece. The prototype was designed and developed on the java platform and ran on Nokia N95 mobile phones. It acquires its position via GPS and depicts the position and state of the user, as well as other users, places and objects, on a specifically designed map of the area. It uses GPRS and 3G, where available, data connection to connect to a server that manages all the information of the activity in real time.

Selecting appropriate evaluation methods

The needs and particular characteristics of the project and the corresponding system were evaluated and taken into consideration, in order to select the most appropriate usability evaluation methodology. The contributing factors in specifying the evaluation methodology are described below.

- The goal of the evaluation, besides assessing the usability of the system, was to specifically address issues regarding a future extension of the system, as well as to gain valuable usability insights for the design and development of similar systems. This suggested an inclination towards formative usability evaluation and qualitative results.
- The application had to run on a mobile device. This sets several limitations regarding the hardware, screen size, user interaction capabilities, communication and networking ability, representation capabilities and software.
- The context of use of the system can be fairly complex. The system would be used outdoors, in an urban environment, while the users are mobile. Besides technical issues that need to be addressed, such as screen visibility outdoors, there could be times that the user would be distracted. Moreover, the system has to faithfully represent the physical space, while the fact that it is used in urban context may add considerable complexity. The user should be continuously aware of their position and state, as well as other users and objects in the area. In addition to the challenges posed on a technical level by this, such as spatial representation and GPS signal accuracy, there are important issues that have to be addressed regarding the interaction design.
- The communication requirements of the system are especially intricate. It should support interpersonal and group communication, digital content creation and users with different goals, while taking into account the position and the state of the users in the physical world.

An important and interesting debate is taking place, both in academia and industry, regarding the tradeoffs between field and lab usability evaluation of mobile systems (Kjeldskov et al., 2004). Proponents of the first approach consider evaluation in the field significantly more effective, since they believe that it is possible to obtain a higher degree of realism and a deeper understanding of the user experience, while taking into account the context of use (Abowd and Mynatt, 2000; Barnard et al., 2007; Nielsen et al., 2006). On the other hand, evaluation methods that take place in the lab are certainly less expensive, thus possibly allowing for more iterations, and at the same time certain attempts are being made to bridge the realism gap by reproducing or emulating the context of use, with mixed results (Kjeldskov and Stage, 2004).

Our conclusion was that the context of use of the LOCUNET system was particularly complicated and practically impossible to simulate in the lab. In addition, we observed empirically during internal testing that the user experience can be dramatically different between lab tests and a real-world outdoor activity with multiple users and a specific scenario. Therefore, it was decided that the evaluation should be based on usability field tests with real users under normal conditions of use, and that inquiry, instead of inspection methods, would be primarily used.

The field testing procedure consisted of two field tests, taking place on different days, under normal conditions of use. There were several improvements made to the system used in the second test and a number of changes to the scenario, according to the feedback obtained during the first test. The goal was not merely to compare the results of the two tests or to aggregate the results in order to provide statistically important summative evaluation results, but to provide a diverse user experience, so that different approaches to

the system would be tried and evaluated, and the results obtained would prove more useful and insightful. Besides, usability evaluation can be significantly more effective when it comprises more iterations with fewer users (Nielsen, 2000).
The ultimate goal of the evaluation, the available resources and the innovative aspect of the system and the project, both on a technical and a social level, meant that the evaluation should focus mainly on methods providing qualitative results.
The following usability evaluation methods were used in order to appraise the user experience during the field testing procedure.

- The System Usability Scale (Brooke, 1996), for measuring the subjective satisfaction of the users. The SUS was translated to the Greek language very carefully, since the terminology on the scale could be easily misunderstood or misinterpreted by the users (Finstad, 2006).
- An additional 6-item questionnaire focusing on the user-perceived quality of the system and on technical details.
- A word list, comprising 105 adjectives. The word list that was used was an adapted and translated version of the word list on UserFocus (UserFocus, 2008), with the sequence of the words being randomized for each user. This version is, in turn, based on the Microsoft Desirability Toolkit (Benedek and Miner, 2002). It was expected that this method would provide objective, useful and insightful results, regarding the desirability and usability of the system on many levels.
- A short informal focus group, emphasizing on the user experience with the system.
- Analysis of the server logs from the actual use of the system.
- Analysis of audio files from the actual use of the system. The users were asked to keep a small audio recording device on them during the activity.

The Field Test

The participants in each test were grouped in two opposing teams, each team consisting of 4 users and having its own headquarters on opposite sides on the map. As the users, equipped with their mobile GPS-enabled devices, walked around the area, certain digital objects would appear on their screen depending on their actual geographic location. The objective of the test was to collect as many items as possible and carry them to the team's headquarters. The users could also create and access "infopacks", which were essentially containers of digital information (text, sound, image, or video) associated with a physical location by means of geo-tagging. In addition to the infopacks, the participants could communicate via text messaging.

Evaluation Results

In order to present a concise and high-level view of the usability of the system, the results of the two tests have been aggregated where deemed necessary.
The subjective satisfaction from the use of the system, as measured from the SUS, can be regarded as rather usable. The scale yields a single number representing a composite measure of the overall usability of the system being studied. In the first test, the system scored 65.9 (out of 100) with a standard deviation of exactly 10, while only one participant out of 8 rated the system under 50. In the second test, the system scored 74.9 with a

standard deviation of 14.2, while all the participants had a positive view of the system (i.e. rated the system over 50). The difference in the ratings was expected, since the system was more stable in the second test. The relatively low value of the standard deviation in both tests confirms the validity and reliability, both of the SUS and its translation, and the evaluation procedure that was followed. According to the scoring instructions for SUS, scores on individual items should not be considered meaningful on their own, therefore no further analysis was required.

The rest of the evaluation results are qualitative and are presented aggregated, regardless of the methods that were employed. This is, both because of the need for a concise, holistic appraisal of the usability of the system, and because there was a level of interplay between the evaluation methods - for instance, a selection on the word list could act as feedback to the focus group.

The system was regarded as fairly simple and easy to use by the users. The user interface was considered effective and the application aesthetically pleasant. However, the system was not deemed easy to learn by the users.

The map and the artifacts on the map were clearly visible and the visibility of the screen was very good in outdoor conditions. However, the screen on the device was considered small for this application, while the majority of the users agreed that a touchscreen would be easier to use for this purpose.

The users' experience with the system was described as fun and addictive, and a certain level of immersion in the game was achieved, especially in the second field test. Most users complained that it was difficult and time consuming to use the messaging capability effectively for communication with the particular user interface and keyboard, and suggested that a predefined set of messages should be available to them.

On the more technical side, the response time of the application and the map refresh rate were judged as satisfactory by the users, while improved on the second field test. In addition, the reliability and stability of the system had greatly improved in the second test and considered satisfactory, as well.

Discussion

The usability evaluation of spatial-aware mobile systems involves several issues that need to be addressed, pertaining to the system under evaluation and the selection of appropriate evaluation methods. On the study presented, we have attempted to identify and provide some useful insights regarding such issues.

The issues that have to be considered regarding specific systems under evaluation may include the goal of the evaluation, the degree of mobility, the context of use of the system, the degree of context-awareness provided by the system and the available resources, in terms of users, time and technical equipment.

In addition, while selecting the appropriate evaluation method, there are several issues that have to be taken into account, such as:

- The type of method (inquiry, inspection, testing).
- The results provided by a method, whether they are quantitative or qualitative, objective or subjective.
- The cost of a method, in terms of users, equipment and time available.

- The degree of intrusiveness of a method.

Furthermore, we have supported the popular view in favour of usability evaluation of mobile systems in the field compared to usability evaluation in the lab, observing that usability field tests with real users in a typical context of use can provide a level of realism which can not be easily reproduced in the lab.
Finally, we have confirmed the software engineering approach that dictates more usability evaluation iterations with fewer users. We observed that the system that was employed in the second field test was significantly more functional and usable. We attribute this to the fact that the evaluation results from the first field test were fed to the design process for the system that was used in the second test. Furthermore, partly because of a more stable system and partly because of slight amendments to the evaluation methodology and the test scenario, we observed that the usability evaluation during the second field test yielded some different results and evaluation insights. Since the focus of the evaluation was formative, we argue that the evaluation procedure based on two field tests under slightly different conditions provided significantly more useful results than we would have hoped to obtain from one field test with more users.
We expect that this study can serve as a starting point for both researchers and practitioners in gaining a better understanding of the issues related to usability evaluation of spatial-aware mobile systems.

Acknowledgements
The research project titled “LOCUNET” is supported by the Greek General Secretariat of Research and Technology under the framework of the Operational Program PEP Attikis, Measure 1.2. The Programme is co-financed (70%) by the European Fund of Regional Development (EFRD), which aims to facilitate the reduction of the inequalities within the European Union regions.

References
Abowd, G., and Mynatt, E., “Charting past, present and future research in ubiquitous computing”, *ACM Transactions on Computer-Human Interaction*. Vol 7, no 1, 2000, pp 29-58.

Barnard, L., Yi, J., Jacko, J., and Sears, A., “Capturing the effects of context on human performance in mobile computing systems”, *Personal and Ubiquitous Computing*. Vol 11, no 2, 2007, pp 81-96.

Bevan, N., “International Standards for HCI and Usability”, *International Journal of Human-Computer Studies.* Vol 55, no 4, 2001, pp 533-552.

Benedek, J. and Miner, T., “Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting”, *Proceedings of Usability Professionals Association 2002,* Orlando, 2002.

Brooke, J. “SUS: a "quick and dirty" usability scale”. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland (Ed.), *Usability Evaluation in Industry.* London: Taylor and Francis, 1996.

Finstad, K., “The System Usability Scale and Non-Native English Speakers”, *Journal of Usability Studies*. Vol 1, no 4, 2006, pp 185-188.

Froelich, P., Baillie, L., and Simon, R., “Realizing the Vision of Mobile Spatial Interaction”, *ACM Interactions*. Vol 15, no 1, 2008, pp 15-18.

ISO 9241-11:1998. *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability*, 1998.

Kjeldskov, J., Skov, M. B., Als, B. S., and Hoegh, R. T., “Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field”, *Proceedings of the 5th International Mobile HCI 2004 Conference*, 2004, pp 61-73.

Kjeldskov, J., and Stage, J., “New Techniques for Usability Evaluation of Mobile Systems”, *International Journal of Human-Computer Studies*. Vol 60, no 5-6, 2004, pp 599-620.

Nielsen, J. *Jakob Nielsen's Alertbox, March 19, 2000: Why You Only Need to Test With 5 Users*, 2000. http://www.useit.com/alertbox/20000319.html

Nielsen, C., Overgaard, M., Pedersen, M., Stage, J., and Stenild, S. , “It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field”, *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, Oslo, Norway, 2006, pp 272 – 280.

UserFocus. *Measuring satisfaction: Beyond the usability questionnaire*, 2008. http://www.userfocus.co.uk/articles/satisfaction.html