JUS
JOURNAL OF USABILITY STUDIES

Issue 1, Vol. 1, November 2005, pp. 4-16

Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing

Anne Kaikkonen
Nokia
P.O.Box 407,
00045 Nokia Group, Finland
+358504837400
anne.kaikkonen@nokia.com

Titti Kallio
TeliaSonera Finland
P.O. Box 585
00051 Sonera, Finland
+358405036957
titti.kallio@teliasonera.com

Aki Kekäläinen
TeliaSonera Finland
P.O. Box 585
00051 Sonera, Finland
+35840834113
aki.kekalainen@teliasonera.com

Anu Kankainen
Idean Research
Tekniikantie 12
02150 Espoo, Finland
+358405079259
anu.kankainen@ideanresearch.com

Mihael Cankar
TeliaSonera Finland
P.O. Box 585
00051 Sonera, Finland
+358405950880
mihael.cankar@teliasonera.com

Abstract
Usability testing a mobile application in the laboratory seems to be sufficient when studying user interface and navigation issues.

The usability of a consumer application was tested in two environments: in a laboratory and in a field with a total of 40 test users. The same problems were found in both environments, differences occurred in the frequency of findings between the contexts.

Results indicate that conducting a time-consuming field test may not be worthwhile when searching user interface flaws to improve user interaction. In spite of this, it is possible that field testing is worthwhile when combining usability tests with a field pilot or contextual study where user behavior is investigated in a natural context.

Keywords
Usability testing, laboratory testing, field-testing, method comparison

Introduction
Usability testing is a common tool used to evaluate the usability of a mobile application in a development

process. Usability tests are usually conducted using a think aloud protocol based on K. A. Ericsson and H. A. Simon's work (1980, 1984). Users are given tasks in a test environment and encouraged to think aloud while trying to accomplish the tasks. This gives us, usability practitioners, information we need on how the user interface matches the natural human way of thinking and acting and highlights the features and processes to be improved.

Severity of the usability problems is an important factor when defining the urgency of actions related to a problem. The most urgent actions are needed when the problem prevents completion of the task. Dumas and Redish (1993) use four point scale, where the first severity level represents the most severe problems and the last the least severe. Also Kallio et al. (2004) divide the severity of problems into categories; high (failure in task execution), medium (not so severe, task can be executed) and low (minor problems).

Laboratory versus field

Usability tests are traditionally conducted in usability test laboratories, consisting of e.g. a living room or office-like area connected to a monitoring area with a one-way mirror. The laboratory environment is a peaceful space, where a test user can concentrate on the given tasks.

Usability researchers and practitioners have been concerned that laboratory evaluations do not simulate the context where mobile phones are used (Johnson 1998) and lack the desired ecological validity. Interruptions, movement, noise, multitasking etc. (Tamminen et al. 2004) that could affect the users' performance are not present in laboratory tests. The surrounding environment and mobility are assumed to set special requirements for mobile applications. Usability testing should take these requirements into account.

Even if there seems to be a common concern about the adequateness of laboratory evaluations, field evaluations have been rather rare. A literature study by Kjeldskov and Graham (2003) revealed that most (71%) mobile device evaluations were done in laboratory settings. This may be due to data collection techniques such as think aloud, video recording or observations being difficult in the field.

As mobile video recording systems, like small video cameras, have rapidly developed during past few years and conducting user tests in the field has become easier. It is now possible to attach a small camera to record the screen of the mobile device and collect that information for later review (Kjeldskov et al. 2004a, Roto et al. 2004). This development allows similar test setting in the field as in the laboratory; it is possible for test leader to follow what is happening on the screen and hear users' comments. This also allows the usage of think aloud protocol in usability test in the field. Despite the development of suitable tools testing in the field is still likely to be more time consuming (Kjeldskov et al. 2004a) than in laboratory setting. It may also require extra effort from test users and the test leader.

Resources for application development are limited in the mobile industry, and usability activities such as user-centered design and usability testing must be made very efficiently. The goal in a product development process is to find the biggest and most fatal usability problems within the strict limitations of

project budgets and deadlines. The focus of the usability inspection is not on finding every possible detail. Decisions made by usability expert when planning usability tests are related to risk management; how to optimize the effort and the outcome (Nielsen and Landauer 1993). Choosing the right evaluation method is important; scientifically validated information on suitable testing methods is valuable for usability practitioners. Kjeldskov, Skov and Stage (2004b) presented a good example of information practitioners' need when making decision on the method in their article Instant Data Analysis: Evaluating Usability in a Day. Hertzum (1999) compared role of three different methods (laboratory tests, workshops and field tests) in a product development cycle. His goal was also to increase the efficiency of the tests. In our study, the main question is to find out whether field tests are critical when evaluating mobile application usability or can the sufficient ecological validity be simulated in laboratory environment?

Comparative study

To find out the effect of the environment in usability tests, we conducted a comparative study both in field and laboratory environments. All the other elements (tasks, think aloud method etc.) were identical, only the test environment was different. The test environments were realistic: the laboratory was the kind where usability professionals usually run tests with a low budget and the field test was conducted in an environment where users use mobile applications.

Kjeldskov et al. (2004a) conducted a similar comparison study for an expert application. Their study and ours differ in some aspects, however: the laboratory setting in the Kjeldskov et al. study was built to resemble a part of the physical space of a hospital department. We believe that the closer the laboratory setting is to the actual environment, the smaller is the difference between results achieved in the laboratory and in the field. We used a normal low budget laboratory environment with no modifications for this particular test. This kind of setting better reflects the reality where industry usability practitioners work: they do not have the time or the budget to build special spaces for each project. In most cases mobility, e.g. motion while performing the task, multitasking, potential interruptions and interactions with bystanders, is not simulated in laboratory tests.

We wanted to test an application designed for consumers whereas Kjeldskov et al. (2004a) used an application designed for professionals. It is different to test a consumer application than a tool for professionals. Where professionals usually have a clear picture of their tasks, consumer users do not necessarily have a clear idea of the possibilities of the application they are using. The users may not have any specific and clear goals during their free time, or the goal can be vague, e.g. time killing or entertainment. Users of consumer applications can also be more explorative and creative with their actions than professional users (Mäkelä et al. 2001). Being explorative and non-task oriented may require more from the user in real mobile contexts than in the laboratory settings.

How many users are needed?

The number of test users in an ordinary usability test in an iterative development process is often 5 to 10 users per test round. Our goal was to study two groups of

users that in one location in order to have very little variation in performance. This means that the groups should not only have same experience level, but they should be big enough.
When looking for statistically meaningful differences between two test locations the number of users should be bigger. In a paper by Faulkner (2003) a minimum of 95% of the usability problems were found with 20 users and variation between the groups was fairly small.

Kjeldskov et al. (2004a) used 6 users in the field and 6 more in the laboratory. Although this might be adequate when conducting usability tests as part of an iterative product development process with a tight schedule, a bigger number of users can increase the power of the test to find differences between two test settings.

Research questions and hypotheses
a. Are the same problems and phenomena found in the laboratory and in the field?

If the comparison tests are conducted with large enough user groups, the number of problems in the field will be bigger.

b. Are the problems and the phenomena the same in both environments? If not, what is the difference?

The problems will differ between the environments. For example the long download time would be more tolerable in the field.

c. If there are differences between the findings, is their severity different in laboratory or in the field?

The problems in the field will be more severe due to interruptions during the task execution.

d. Are task execution times different? What can we conclude from it from the test design point of view?

Task execution times will be longer in the field.

e. Does the environment affect test user performance?

There will be more interruptions during the task execution in the field. These executions will affect performance.

f. Is a laboratory or field test more suitable when evaluating usability of mobile applications?

Field test will be more suitable when evaluating usability of mobile applications, because the context affects usage and performance.

Process

Test Settings
Two usability evaluations were conducted to study the effects of the testing environment. The first test took place in a typical usability laboratory setting with 20 test users. The second test took place in the field, with other 20 test users. The popular usability testing method was used in both environments: the think aloud protocol and predefined series of test tasks. Think aloud was prompted to users by asking users to tell what they were doing, what they expected to happen when making selections and whether something unexpected happened after the selection.

To avoid the effect of a moderator in the test, the tests were conducted with four moderators. All moderators were experienced usability practitioners with 5 - 13 years of experience in usability testing. Each moderator conducted an equal number of tests both in the laboratory and in the field. The instructions given to test users were predefined and written on paper. Moderators were given detailed instructions concerning the interaction with test users: how to give instructions, and when and how to prompt in problem situations. All moderators participated in data gathering and qualitative analysis of the outcome. Statistical analysis was performed by one moderator.

Test Users
The users in the two test groups did not differ regarding their experience with mobile phones.

The test users in both tests were from 22 to 35 years old. In the laboratory test the average age of the users was 28,5 years with a standard deviation of 3,5 years. In the field test the average age of the users was 28,8 years with a standard deviation of 4,0 years. For practical reasons all users lived in the greater Helsinki area. An equal number of men and women took part in the tests. All users had used mobile phones for more than five years and had some experience using WAP. The application used in the tests requires a handset with the Nokia Series 60 (with 3 softkeys) Symbian platform. Recruiting users having similar experiences with that handset would have been difficult. Instead, the test users were recruited such that they did not have experience with this particular handset but were rather Nokia's Series 40 (with 2 softkeys) mobile phone owners.

Laboratory Tests
The laboratory tests were conducted in a typical usability test environment. The setting was well controlled: there were no unexpected external interruptions, disturbing noises, varying lighting conditions and so on. The test sessions were recorded with three video cameras and a microphone. The cameras recorded the display and keyboard of the mobile handset, the user's face and an overall picture of the user.

The users were given a brief introduction to the mobile handset (i.e. basics compared to their own phones) and were instructed to think aloud during the test. The tasks were given orally in a predefined order using agreed wording. The oral task giving was decided to be used also in laboratory environment as paper-based instructions would have been difficult in the field.

Field Tests
The field tests were conducted in Helsinki. The users could walk, stand still, sit or do whatever they would normally do while performing the tasks. As in the laboratory, the moderator gave the defined tasks orally to the user one at a time, using the agreed wording.

Test sessions started in an office district in Ruoholahti around the daily rush hour. The users were instructed to take the metro and to go to meet a friend in the Itäkeskus shopping centre. The users were given the first task while they walked to the metro station. The moderator followed the user a few steps behind and gave the tasks during the journey as the users finished the previous ones. The metro trip to Itäkeskus took about 16 minutes. Parts of the tasks were performed while moving around in the shopping centre. During the

test the users had to cross a busy street, use crowded escalators, travel in a metro and walk in a large shopping centre full of disturbances - while doing the tasks given by the moderator.

Field Testing Equipment
In the field test users had to wear special equipment to allow the recording of the test data while they were on the move. The equipment can be seen in figure 1. The equipment consisted of two parts: the test user unit and the moderator unit. The test user unit had three video cameras. As in the laboratory, the first one captured the display and the keyboard of the mobile handset, while the second camera recorded the user's face. The third camera was fixed to give a view of the user's surroundings. The unit had also a microphone, an audio and video recording device, a quad processor, a wireless video signal transceiver and batteries to provide power.

The moderator unit consisted of a 6" LCD service monitor, a video camera, a wireless video transceiver and a battery. The camera recorded the user's surroundings from the moderator's perspective. The monitor enabled the moderator to see what the user was doing with the mobile application when the user was walking or was otherwise in a position where it would have been impossible to see what was going on without additional equipment.

Application
The tests evaluated the usability of an application called Mobile Wire. With this application, the user can transfer files between computers and the mobile handset. The application was in a pilot stage and that made it suitable for this test, as consumers were not yet familiar with the application. The tested application is one that would realistically be used in a mobile context (e.g. transfer a document from the office PC to the home PC or show an image stored in a remote computer from the screen of the mobile handset).

Figure 1. The field testing equipment and test going on.

Tasks
Users performed ten (10) tasks during the test. The first test task was easy and the goal of it was to make the user acquainted with the controls of the mobile phone. Tasks 2 to 10 were the ones actually testing the download and use of the test application.

1. "Getting started" task: sending a SMS.

2. Opening the received SMS, opening a WAP page mentioned in the message.
3. Downloading the test application to the device.
4. Finding and opening the downloaded application.
5. Using the application to view images shared by a friend.
6. Downloading the shared picture to the phone.
7. Taking a picture using the camera of the phone, saving it to home computer.
8. Giving rights to a friend to view the picture saved on the home computer.
9. Closing the application.
10. Deleting the application.

Findings

a. Are the same problems and phenomena found in the laboratory and in the field?

An identical set of 46 distinct usability findings were observed both in the laboratory and in the field. This number includes both usability problems (such as problems in navigation) and observations of usage (such as information search patterns).

Twenty-two (22) of the problems occurred more than once in the laboratory and in the field. The 22 usability problems are listed below and their frequency is presented in figure 2. The first nine problems in the figure are considered to be most critical ones; the problem prevented users to complete the task. These are marked with number 1 in the list of problems. Problems from 10 to 19 were severe problems; they caused clear problems to users while performing the tasks. These are marked with number 2 in the list of problems. The last three were minor problems causing discomfort when using system for the first time, but users learned them during the test. For problems marked with star, statistically significant differences were found in the number of occurrences. (See figure 2)

1. Cannot find the download link from WAP (1)
2. Does not find the price (1)
3. Difficulties in exiting the browser (1)
4. Difficulties in finding the application after the download (1)
5. Difficulties in understanding the concept of giving access rights (1)
6. Added buddy but did not give access rights to computer (1)
7. Did not save the changes after giving the rights (1)
8. Gives up the deletion of the application (1)
9. Gives up at least one task during the test (1)
10. Adding recipient problematic (2)
11. Cursor move causes problems (2)
12. Unclear whether the download process has started (2)
13. Download dialogue problematic (2)
14. Reloads application, not sure if it was already downloaded (2)
15. Tries to send picture from camera apps (wrong place) (2)
16. Tries to send picture from list of devices (wrong place) (2)
17. Tries to send picture from tab (wrong place) (2)
18. Tries to throw the application to trash bin (2)
19. Tries to delete the application from wrong folder (2)
20. Download time problematic (3)

21. Different menus from joystick and softkey cause problems (3)
22. Reads help at least once during the test (3)

b. Are the problems and the phenomena the same in both environments? If not, what is the difference?

In three cases, there was a statistically significant difference in number of times a finding was observed in the two settings. All of these were more common in the field.

1. Difficulties in understanding the concept of giving access rights to their own PCs to allow their friends to search for pictures stored in it (task 8, problem 5 in figure 2). While the same number of users in the field and in the laboratory eventually succeeded in giving the rights, the users in the field did so only after trying in several of the wrong places.
   Laboratory: 1, Field: 8
   Chi-square = 6.62, df = 1, $p = 0.01$
2. Unclear whether the download process has started (task 3, problem 12 in figure 2).
   Laboratory: 5, Field: 12
   Chi-square = 4.49, df = 1, $p < 0.05$
3. User tries to delete the application from wrong folder (task 10, problem 19 in figure 2).
   Laboratory: 8, Field: 17
   Chi-square = 7.79, df = 1, $p < 0.05$

c. Severity of the problems

On average the problems in the field were not more severe than the ones coming in laboratory. One of more frequent problems in the field was critical (1), the other two were severe (2).

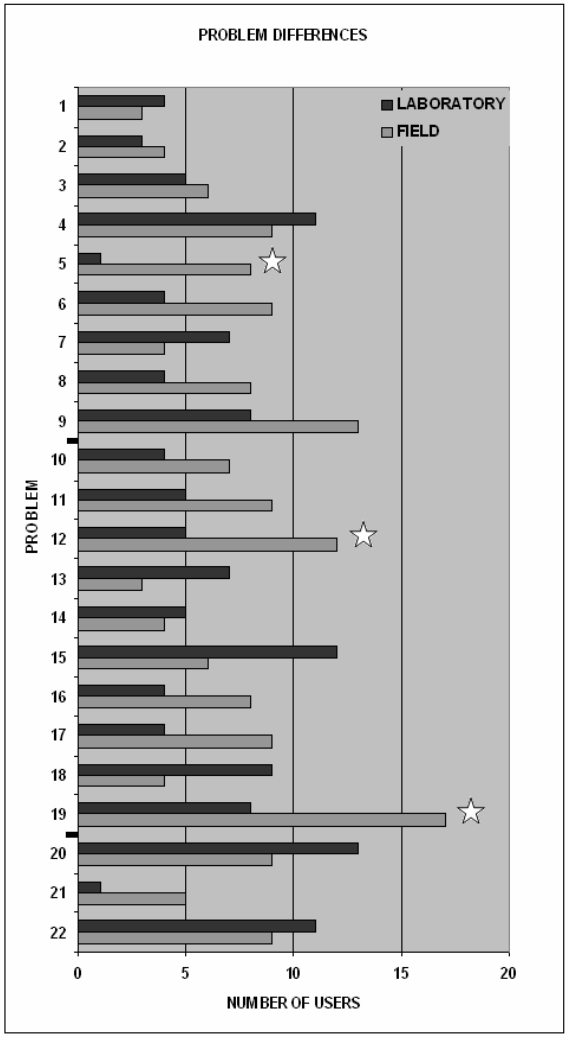

Figure 2. Problems that were observed more than twice in the laboratory and the field

An interesting issue is related to problem 20 in figure 2; in the laboratory environment 13 users complained about the download time, in the field 9 users commented about same issue. The wording used in laboratory showed more frustration than in field environment. Based on the reaction of the test users, the experts would have rated the problem severity differently in two locations: In laboratory problem was considered to be severe (2), but in the field the rating was minor problem (3).

d. Are task execution times different?
There were no significant differences in the execution times of individual tasks between the two test settings, as can be seen from figure 3.
Tasks in figure 3:

1. The WAP page is opened from the SMS
2. Application download starts after opening the WAP page
3. The application opened after installation
4. First friend's picture is opened
5. The location of the picture checked from phone
6. A picture is taken and saved on home PC
7. The friend is added to the buddy list
8. Access rights are given to the friend
9. The application is removed from the handset

Execution time measuring started when user started the task and ended when the task was completed. . Times were not measured for tasks that include significant, varying network delays, such as downloading WAP page.

Figure 3. Task execution times in the laboratory and in the field. There were no statistically significant differences between the environments

e. Does the environment affect test user performance?
The test location did not significantly affect speed and success of the task execution of the users. However the location seemed to have a greater impact on qualitative findings of the test.

The users' concentration in cognitively loading tasks was obvious in some cases during the field tests: when having a cognitively loading task, users behave differently than when having less loading task. When the task given was not familiar to them, the users might have stared at the phone display while walking on and barely avoiding the crowd on their way, or they might have stepped aside to a quieter spot to finish the task, ignoring the other people passing by. This

behavior gives insight on the level of difficulty of the tasks, and did not come out in a laboratory setting.

Potential interruptions did not seem to bother users during the test: even loud bystanders that captured the moderator's attention did not break the users' concentration. They kept on performing the task without significant interruptions. This behavior is actually quite normal in public transportation in Finland: people usually mind their own business without paying too much attention to the other passengers if they are strangers. There were only few curious passengers in the metro, mostly drunken people, that came to talk to the moderator, but even they left the test user in peace.

In the field, the test users seemed to talk more freely about the application, and whether they might use it themselves.

While downloading something or waiting for the downloading to finish the users did not always pay attention to the task at hand, but might pick up a free newspaper or check the missed calls on their personal phone. This was also the situation when they looked around and observed the other passengers. This behavior was only observed in the field and may have implications on how the users notice the various progress indicators in the application: In the field they may miss the indications that disappear with timing.

f. Is a laboratory or field test more suitable when evaluating usability of mobile application?
Even if there were no significant performance differences between the places, the total test time was longer in the field (on average 45 minutes) than in the laboratory environment (on average 35 minutes). The difference of the total test time is statistically significant. ($F_{1,0}$= 5.41; $p<0.05$). The difference is explained by the test setting in the field: certain tasks were synchronized to take place at specific locations. This means that many field test users spent some extra waiting time between these tasks.

Also, the time used for the test arrangements was longer in the field setting than in the laboratory. In the field tests more time was spent setting up the equipment, for example the backpack with the recording devices needed to be adjusted for each user. The field tests always started in the same place and followed the same route, meaning the test moderators spent more time travelling between places.

Table 1 summarizes the times needed for tasks, preparations and travel during this test. In the laboratory the tests could be arranged to start every hour, but in the field tests the interval between the tests had to be two hours.

Table1. Test times

Place	Total testing time (average)	Instructions and preparations (estimate)	Travel time for test moderator (estimate)
Laboratory	35 min	10 min	-
Field	45 min	20 min	40 min

Discussion or Implications or Recommendations

The comparative study we performed surprised us, as results did not support much the hypothesis we had placed. The following discussion is organized according to the research questions set in the introduction:

a. Are there differences in number of usability problems found in the laboratory and the field environment when testing a consumer application?
According to our study there was no difference in the number of problems that occurred in the two test settings. Our hypothesis that more problems would be found in the field was not supported.

b. Are the problems and the phenomena the same in both environments? If not, what is the difference?
Despite the fact that the same problems were observed, some differences emerged between the two settings. The problems that came out more frequently in the field seem to be related to understanding the application's logic, but, on the other hand, there were also complex issues with no differences between the two test settings.

c. Are there differences in severity of the problems between laboratory and the field?
The hypothesis that more severe problems would be found in the field test was not supported. Also in terms of problem severity, there were no differences between the test environments.

d Are there differences in task execution times? What can we conclude from that from an test design point of view?
No, the task performance times in individual tasks were no longer in the field than in the laboratory. However, the total time needed for the testing was longer in the field than in the laboratory, which indicates that the field-testing is a more time consuming method than the laboratory testing.

e. Does the environment affect test user performance?
In the field test there were potential interruptions, but these did not seem to affect the performance much. When having more complex task, users sought for a safe haven when performing it. Only few users were able to perform all the tasks while walking. In the field test the users concentrated heavily on the test. The test users kept on working on the task when entering and exiting the metro, for example. They did not seem to be bothered by the other metro passengers, even if these came to talk to the moderator. Only few people came and asked about the test from the moderator and even they left the test user in peace. It could be interpreted that users were performing the tasks inside a bubble, which in fact is normal behaviour in public places in Helsinki. People are concentrating on their own activities and ignoring what the others do unless the activity is noisy or threatening.

Although the moderator actions in the tests were defined to be same in the field and in the laboratory, the field test seemed to be more casual. Users tended to comment more freely about the application.

f. Is a laboratory or field test more suitable when evaluating usability of mobile applications?
When performing a user interface evaluation of mobile applications and devices, field-testing may not add significantly to the validity and thoroughness of the

test. Not because some problems might not be identified, but simply because the time needed in the field is longer and the effort required in the field is higher. This outcome differs from the results of Hertzum (1999); in his study the time required in the field tests was significantly smaller than in the laboratory test. In the study of Hertzum, the field test was conducted by users, without supervision of the test leader. Based on our study, laboratory tests seem to give sufficient information to improve the user interface and interaction of the system. Since the same problems are found, field-testing can be cost efficient if it can be combined with a commencing pilot study.

Communication in the field situation between the moderator and the test user was more casual after the test, and it seemed to be easier for the test user to tell about feelings related to the concept in general. According to Roto et al. (2004) the field test method is suitable for situations where not only interaction with a system is tested, but also user behaviour and environment is examined. In addition, confidentiality of the application or device in the industry often drives the decision towards the laboratory testing; especially in the beginning of the development cycle.

In the field environment the users seemed to search for a peaceful area to interact with the device and application. Similar behaviour has been observed when people use mobile phones in person-to-person communication. This search for a safe haven has been thought to be related to the need to seek peace in person-to-person communication (Kopomaa 2000). The need for a personal space does not seem to be related only to communication with others, but is more general: in public places people need to have privacy when doing their own business.

Impact of the study and future work
As in industry the goal is to find the biggest and most fatal usability problems within the strict limitations of project budgets and deadlines, this study helps in making decision on the test place. Laboratory test gives sufficient information when testing usability of a mobile application.

The results presented in this paper apply to cases where a mobile consumer application is tested in the field to find usability problems to serve product development. More tests with different kinds of applications are needed to validate and generalize these results. Especially evaluating applications with location information the laboratory setting may not give the best outcome. In this study we concentrated on user interaction with the mobile application and the mobile device, but the data gathered in this study can be further analyzed to better investigate the behaviour of the test users and the effect of the environment.

Practitioner's Take Away

- When testing a user interface of a mobile application, field testing may not necessarily be the best place; mostly because it is more time consuming than the lab test.
- Testing in the field requires double the time in comparison to the laboratory. In the field you can run half of the tests per day you run in the laboratory.
- In a field test, running a pre-test or a pilot is critical: there are so many details that can go wrong, and you really need to check that everything is working correctly.

- When testing in the field, be prepared that things will not go as planned: there may be interruptions and unexpected events more than in lab.

Acknowledgements

We want to thank Tuomo Nyyssönen from Nokia for his valuable help with the test equipment.

References

Dumas, J.R., Redish J.C (1993) A Practical Guide to Usability Testing. Ablex Publishing Corporation Norwood NJ

Ericsson, K. A., & Simon, H. A. (1980) Verbal Reports as Data. Psychological Review, 1980, 87, 215-251.

Ericsson, K. A., & Simon, H. A. (1984) Protocol Analysis: Verbal Reports as Data. Cambridge, MA: Bradford Books / MIT Press.

Faulkner, L. (2003) Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments and Computers 2003, 35 (3), 379- 383.

Hertzum, M. (1999) User Testing in Industry: A case study of laboratory, Workshop and Field Tests. Proceedings of 5th ERCIM WORKSHOP ON "USER INTERFACES FOR ALL" http://ui4all.ics.forth.gr/UI4ALL-99/Hertzum.pdf

Johnson, P. (1998) Usability and Mobility; Interactions on the move. In Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices, Glasgow, Scotland, GIST Technical Report G98-1.

Kallio, T. , Kekäläinen A (2004) Improving the Effectiveness of Mobile Application Design: User-Pairs Testing by Non-professionals. Proceedings MobileHCI 2004Conference, Glascow,UK Springer-Verlag 2004 315-319.

Kjeldskov, J., and Graham, C. (2003) A Review of Mobile HCI Research methods. In Proceedings of the 5th International Mobile HCI 2003 Conference, Udine, Italy, Sringer-Verlag,.

Kjeldskov J., Skov M. B., Als B. S. and Høegh R. T. (2004a) Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field . In Proceedings MobileHCI 2004 conference, Glasgow, UK. Springer- Verlag, 2004. 61-73.

Kjeldskov J., Skov M. B. and Stage J. (2004b) Instant Data Analysis: Evaluating Usability in a Day. In Proceedings of NordiCHI 2004, Tampere, Finland. ACM, 2004. 233-240

Kopomaa, T. (2000) City in your pocket. Birth of the Mobile Information Society. Gaudeamus, Helsinki 2000.

Mäkelä, A., and Fulton Suri, J. (2001) Supporting Users' Creativity: Design to Induce Pleasurable Experiences. In M.G. Helander, H.M. Khalid, T. Ming Po (Eds.), Proceedings of International Conference on Affective Human Factors Design. ASEAN Academic Press, 2001. 387- 394.

Nielsen J. and Landauer T.K. (1993) A Mathemetical Model of Finding Usability Problems. In proceedings of INTERCHI '93. 206-213.

Roto, V. Oulasvirta, A. Haikarainen, T. Lehmuskallio, H. & Nyyssönen, T. (2004) Examining mobile phone use in the wild with quasi-experimentation. HIIT Technical Report 2004-1, August 13.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2004). Understanding mobile contexts. Special Issue of Journal of Personal and Ubiquitous Computing. 8, 135-143.

Anne Kaikkonen is senior user researcher in Nokia Technology Platforms. She has 15 years experience in using psychological knowledge in design and development; first in traffic psychology, later in technology area. She has been working in Nokia for seven years, focusing mainly in usability of mobile Internet and mobile applications.

Titti Kallio (M.A. in psychology) is working as Director of Design and Usability for TeliaSonera. She has 20 years of experience in the usability field.

Aki Kekäläinen has proudly tackled usability and user experience puzzles offered by a variety of devices and services for 7 years in both national and international projects. Aki's main focus has been on the mobile field, but all attempts to block him from getting involved with web, Windows, voice and interactive TV areas have also been rather futile. Currently he holds a BSc degree in computer engineering and is studying for a MSc in computer science. He is working as a Usability Specialist at TeliaSonera.

Mihael Cankar Currently a usability specialist at TeliaSonera, Mihael has eight years of experience in designing interactive services. He has worked as the users' advocate in a variety of projects both for the web and for mobile devices. He also strives to finish his Master's degree in Human-centered Information Systems at the Helsinki University of Technology really soon. Honest.

Anu Kankainen, PhD, is psychologist who has 12 years of experience in working together with engineers, designers and marketing people in several international and national projects focusing on mobile technologies. She did her doctoral thesis on user-centred product concept design of mobile products at Helsinki University of Technology. Currently, she is Senior User Experience Specialist at Idean Research (www.ideanresearch.com) in Espoo, Finland.