JOURNAL OF USABILITY STUDIES

Vol. 3, Issue 2, Feb 2008, pp. 90-98

A New Approach to Analyze Human-Mobile Computer Interaction

Jürgen Kawalek
University of Applied Sciences
Zittau/Görlitz
Brückenstraße 1
02628 Görlitz
Germany
j.kawalek@hs-zigr.de

Annegret Stark
University of Applied Sciences
Zittau/Görlitz
Brückenstraße 1
02628 Görlitz
Germany
astark@hs-zigr.de

Marcel Riebeck
University of Applied Sciences
Zittau/Görlitz
Brückenstraße 1
02628 Görlitz
Germany
mriebeck@hs-zigr.de

Abstract

This paper describes a tool for log file recording and a method for quickly and easily analysing human-computer interaction with mobile devices. The tool logs screenshots and quantitative interaction data, such as number of clicks and timestamps. The analysing tool provides the ability to evaluate the interaction sequences and to export an MS Excel®-sheet for statistical analysis. To evaluate the tool, a usability study was conducted comparing the effectiveness of this tool in the laboratory and in the mobile context. Findings show that the tool is the first step toward a very effective, unobtrusive analysing method for user interaction in the mobile context. Combined with debriefing methods, it would be an optimized way for usability testing with mobile devices.

Keywords

usability method, usability data analysis, laboratory study, field study, mobile device, think aloud

Introduction

Some research has been done on usability testing methods comparing user interaction with mobile devices in the laboratory or in the mobile context. Results show that it is appropriate to test mobile user interaction in the lab (e.g., Jambon, 2006; Kaikkonen et al., 2005). But, testing in the lab does not work for mobile applications like navigation applications or electronic tourist guides, as these applications do not work without contextual information.

Approaches such as following the user with a cameraman or letting her/him wear a capture vest with a camera on her/his shoulder are supposed to influence natural behaviour. Therefore, an unobtrusive way of testing in the field is needed.

Before testing the usability of a mobile application, several questions have to be answered. Is it appropriate to test in the lab, or does the context have an enormous effect on interacting with the mobile application? Which kinds of data are supposed to produce most of the usability problems and which method is appropriate to the development status of the application, respectively? These questions will be answered in this paper.

Lab versus field evaluation

The first decision to make is whether a usability test should take place in a laboratory or the field. Both approaches have advantages and disadvantages.

Conducting usability tests in the lab is organisationally much easier than testing in the field because almost all influencing factors can be controlled. Interaction and think-aloud data can be recorded with several cameras and screen capturing tools because of having easy access to all technical options. In addition, participants are comfortable with thinking aloud because there are no passers-by. All in all, the context, which is a big part of mobile interaction, is not considered. Thus, results of usability studies of mobile applications conducted in the lab cannot be assigned to mobile use.

In field studies, context factors are taken into consideration. Therefore, the evaluator has less control of the setting. Capturing interaction data might be difficult as participants might refuse to think aloud while walking along a street. In addition, following them with a camera is obtrusive and might change their normal behaviour. Nevertheless, in some cases testing in the field is indispensable, e.g., when testing a pedestrian navigation application (Kawalek, Hermann & Stark, 2006).

Subjective versus objective methods

The second question to think about is whether subjective or objective data/results might lead to the most efficient outcome. Both approaches again have advantages and disadvantages.

Subjective evaluation methods, such as questionnaires or thinking aloud protocols, provide data from which conclusions can be drawn about user acceptance, satisfaction, or problems of use. These data reflect verbal representations of factors of which the test subjects are aware. Facing subjective measures one main methodological problem arises: these data might easily be distorted by errors of judgement, e.g., halo effect, answering tendencies, primacy-recency effect, and so on. Another problem is related to inconsistencies in doing something and thinking about it. In other words, there is often a difference between what people say and what they actually do (Schweibenz & Thissen, 2003).

The think-aloud method, on the one hand, has proved to be very effective for learning about cognitive processes of participants solving a problem. It can provide rich data, which log data do not seem to be able to provide (Branch, 2000). On the other hand, thinking aloud also is said to cause reactivity as participants work differently from their normal behaviour as a result of thinking aloud. Van den Haak et al. (2003) found out that participants thinking aloud perform less successfully than those working silently. In their study, the workload, plus the think-aloud condition seem to affect the task performance in a negative way. Therefore, an alternative method has to be tested, since working on tasks silently is most comparable to the normal working context.

For obtaining valid usability data, it is necessary to gain objective data about what test participants really do. But, deploying a human observer influences the participants and leads to irrelevant results. To give an example, the participant may suffer from test anxiety (Bradner, 2004), or try to show behaviour that they expect to be socially desirable (Edwards, 1957).

To prevent these undesirable side effects, the human-computer interaction can be recorded technically. The most frequently used tool for technical recording of user interaction behaviour is log file recording. Using this technique, all user input and system messages are stored with a time stamp. The primary advantage of log file recording is the possibility of observing and storing operations that cannot be verbalized. Indeed, objective data show how the user interacts with the application, but it does not offer why it is this way.

To meet all discussed aspects, not only subjective but also objective measurements should be analysed. As mentioned above, subjective data can be influenced by individual experiences and abilities. That's why objective log data and screenshots of every interaction were captured in this study, in addition to think-aloud data. This approach became necessary to fully understand user interaction with the mobile application and to identify usability problems. To find out if usability problems can be identified only by analysing log data, this setting was compared with two think aloud settings - one in the lab and one in the field.

Methods

A usability study with 30 participants was conducted using the MDA III and a mobile application that provided information about local traffic, pharmacies, travelling, and shopping. The participants were asked to perform three scenario-based tasks (Rubin, 1994):

- Booking a flight
- Finding a pharmacy nearby that is open after 6 p.m.
- Finding a tram connection to the railway station

After successfully finishing a task, participants were asked to push a "finished"-button.

One third performed the tasks in the lab (group 1); one third tested the application out in the field, accompanied by a researcher (group 2); and the last third tested out in the field on their own (group 3). Groups 1 and 2 were asked to think aloud. In group 1, a video of the user interaction with the mobile application was recorded using Bluetooth and remote desktop control for mirroring the screen to a PC and using a screen capturing tool.

Log file recording tool

To capture the log data, a log file recording tool for mobile applications was developed, which records several events of the mobile device.

The following data are logged by the tool:

- Clicks with timestamp and the position on the screen.
- Item selection: For controls which hold a list of data (e.g., listbox, combobox, treeview etc.). Each time an item is selected by the user the data of the new active item is stored.
- Screen changes: As soon as the user manipulates data in the dialog, the time and action are recorded.
- Screenshots: In addition, with every change (manipulation of the user), a screenshot is taken to be able to retrace the user's interactions.
- "Finished"-Button: After finishing a task, the participant has to click this button. A timestamp is saved to identify the time a user needed to fulfil a task.

Analysis

Stored objective data (screenshots, screen events and relevant timestamps, number of clicks, and time needed to fulfil a task) can be imported to the analysing tool. The evaluator is able to browse through the screenshots analysing the sequence the user took to cope with the tasks. For every screenshot, it is possible to show all the clicks made by the user (Figure 1). Moreover, by reviewing the screenshots, the evaluator can rate the user's interaction in comparison to the ideal way an expert would have taken to solve the tasks.

In this case, the evaluator can give marks from 1 to 3 for every screen. Mark 1 characterizes a perfect step, 2 stands for an additional step not really needed but on the right way, and 3 indicates a wrong step in interaction as an error or a wrong menu choice. This rating might help to analyse big datasets more quickly without evaluating every user's sequence separately. After finishing the review process, the evaluator can export an Excel spreadsheet, including the interaction sequence of every participant with the ratings (1, 2 or 3) for every screen, a table of how often a screen has been shown to the user, as well as the total time and clicks a user needed, and for every task, respectively. Now, the evaluator can identify wrong steps and deviations at a glance because wrong steps are marked by 3, and deviations can be identified by how often a screen was shown compared to the ideal way.

The video and the thinking aloud data were analysed by an experienced evaluator. The identified usability problems were categorised and interpreted and resulted in suggestions for improvement.

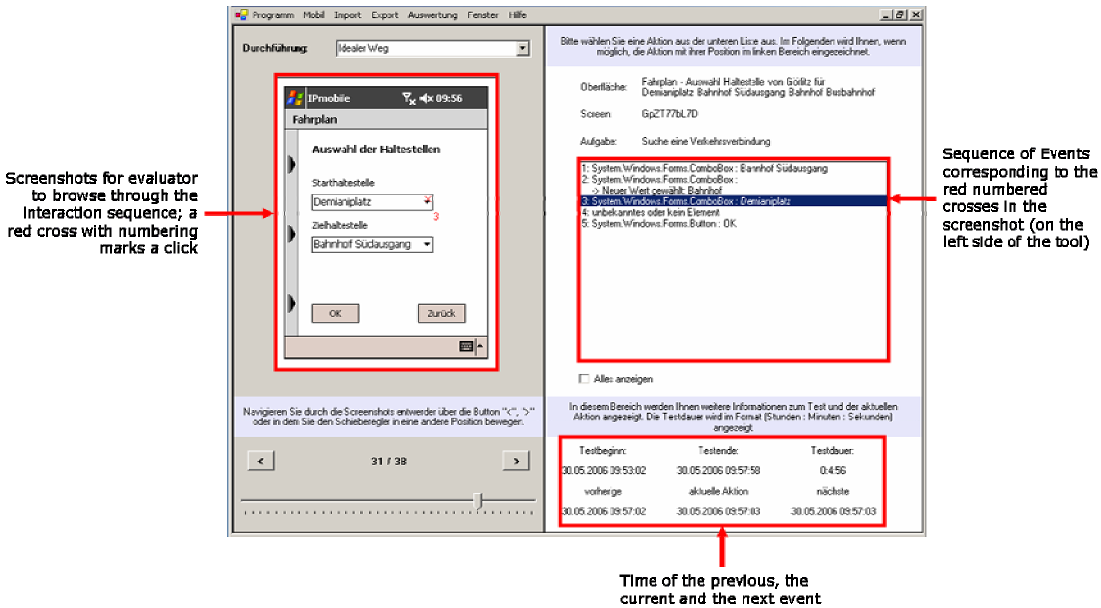

Figure 1. Screenshot of the analyzing tool

Results

Thirty students took part in this study. They were between 20 and 37 years old (mean: 27.5). Of these, 21 were female, 9 male. None of them had any experiences with handheld computers.

Overall, a total of N=21 different usability problems were identified.

Table 1 lists all of the issues and the corresponding method (L: Log file analysis, V: Video data analysis, T: Think aloud) with which they were detected.

Table 1. Identified usability problems with methods

No.	Description of usability problems	L	V	T
1	Data input object: Entering stage	x	x	x
2	Missing Back button	x	x	
3	Handling: dynamic menu function		x	x
4	Handling: calendar		x	x
5	Handling: keyboard		x	x
6	Link Naming	x		x
7	Data input object: Entering data	x		x
8	Information preparation: low-priced flight	x		x

No.	Description of usability problems	L	V	T
9	Display Number of Passengers	x		x
10	Data caching	x		x
11	Interaction structure: opening hours	x		x
12	Interaction structure: district	x		
13	Interaction structure: Schedule-Button	x		
14	Default setting number of passengers	x		
15	Entering airport	x		
16	Entering period	x		
17	Entering time			x
18	Entering name and post code of a city			x
19	Missing functionality: bookmark			x
20	Handling: keyboard case sensitivity			x
21	Missing information: Arrival time			x
SUM		**13**	**5**	**15**
Number of problems identified only with this method		5	0	5

Altogether, the analysis of the log files identified 13 usability problems, the video data detected 5, and the think aloud method identified 15 problems. It also can be seen that the log file analysis identified five usability problems that were not detected by any other method. The think-aloud method also found five problems not detected with the other two methods. The video analysis did not reveal any additional problems. To specify the quality of the different problems, they were categorized as seen in table 2 (L: Log file analysis, V: Video data analysis, T: Think aloud).

Table 2. Categorization of usability problems

Categories	No. of usability problems	L n/N	V n/N	T n/N
Inappropriate data input objects	1, 7, 15, 16, 17	4/5	1/5	1/5
Interaction structure problems	2, 11, 12, 13	4/4	1/4	
Handling: Interaction object	3, 4, 5, 20		3/4	4/4
Information preparation	8	1/1		1/1
Object settings	9, 14, 18	2/3		2/3
Data caching	10	1/1		1/1
Link Labelling	6	1/1		1/1
Missing information/ functionality	19, 21			2/2

The use of inappropriate data input objects was identified by log file analysis, as well as with video and think-aloud method. Problems within the interaction structure were identified most easily with log files because they cause significant discrepancies from the ideal interaction sequence. Problems in handling interaction objects, such as dynamic menus, could not be detected by log file analysis because a screenshot cannot capture the animation of objects and dynamic interactions on the screen. They could be found only by analysing the recorded video, or by being verbalized in the think-aloud setting.

Difficulties of the participants in preparing necessary information for the task can be detected with both thinking aloud and with log file analyses, if they cause deviations, such as long editing time. Object settings that cause errors, problems with caching, and problems in understanding link labelling can be found by analysing log files when the user deviates from the ideal way of solving the tasks. These issues, of course, can also be found with the think-aloud method.

Missing information on a screen or missing functionality cannot be identified by log file or video data. This can only be found with thinking aloud.

Table 3 shows the number of the identified usability problems per method and group separately; for Groups 1 and 2, the number of double usability problems per group (which were found with both methods); and finally the total number of identified usability problems per group without the double.

Table 3. Identified usability problems by method and group

Group	Method	Number of Problems Separately	Number of Problems Double	Problems Total
1 (lab)	Logs	10	2	13
	Video	5		
2 (field)	Logs	11	7	19
	Think aloud	15		
3 (field)	Logs	7	-	7

Table 3 shows that the thinking aloud method helped to find 75 % (n=15) of all usability problems detected (N=21). Analysis of the video data only found five usability problems. Considering the total number of usability problems identified per setting separately, 13 usability problems were found in the lab setting (group 1), 19 in the mobile setting with thinking aloud data (group 2) and 7 in the mobile setting with nothing but log file data (group 3). So, in group 2, most of the usability problems were found. The field setting with thinking aloud seemed to combine the advantages of testing in the field and capturing rich data with verbal protocols because, in group 2, most of the usability problems were identified. As the used test application is not relying on mobile context information, usability problems assessed in the lab did not differ from those that appeared in the field.

Figure 2 shows all usability problems (N=21) and the overlaps between the methods with which they were identified.

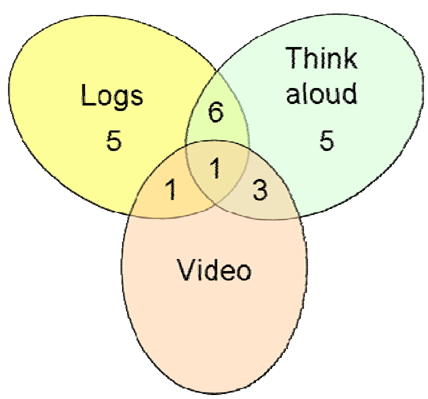

Figure 2. Identified usability problems per method

Five (24 %) usability problems each were found solely by analysing the think-aloud data and by analysing the log files. The usability problems identified only by the think-aloud method concern problems related to interaction object settings and handling, as well as missing functionality and missing information. The five usability problems that were identified by analyzing only the log file data were related to problems within the interaction structure of the application, such as missing back buttons and problems with interaction objects that cause subsequent errors. They could only be identified by analysing wrong steps in interaction that, for example, caused detours or increased editing time for solving a task. The video data did not detect any additional problem that did not also occur within the two other methods.

Six usability problems were found with log file analysis, as well as with analysing the thinking aloud data. These problems are mostly related to the interaction objects. The three problems identified with video and thinking aloud were problems related to the handling of interaction objects, e.g., problems in selecting a date with the calendar object, handling the keyboard, or dealing with the dynamic navigation menu.

Discussion

The log file recording and analysing method for mobile applications was developed to record human-computer interaction with mobile devices in an unobtrusive way and to analyse captured data more easily. The first results are promising.

Analysing log data in the presented way can easily detect usability problems according to the interaction structure of the application. These problems could be identified easily because the log files contained data such as number of clicks and timestamps. In addition, the user's interaction with the device could easily be compared with the ideal way of problem solving with the analysing tool. Those problems also could have been found by analysing the video data, but that would have cost much more time because the video does not contain timestamps or record clicks. With thinking aloud, those problems were not mentioned because they did not cause obvious error messages.

Moreover, it was found that participants not familiar with the system forgot to think aloud, possibly because they were cognitively overloaded by using a new system and solving unknown tasks. So, it can be concluded that log file analysis is an effective way to detect problems in the interaction structure and problems with interaction objects that cause subsequent errors. One disadvantage has to be mentioned here: Log file analysis can point out how problems occur, but it cannot find out why. User interactions have to be interpreted to get suggestions for improvement. Therefore, an experienced evaluator is needed.

Analysing only objective interaction data is also not appropriate for identifying problems in handling interaction objects, e.g., typing a word with keyboard and pen or selecting a date within a calendar-object. For detecting these kinds of usability problems, video or think-aloud data are necessary because those problems do not appear in the screenshots. In this study, the analysis of the video data identified no usability problems that were not also found in the log files or the think-aloud data. Therefore, it can be concluded that video data does not bring any added value compared with the other methods. Furthermore, it is difficult to capture videos of the user interaction in the mobile context because of the limited memory capacity of standard mobile devices.

Regarding technological aspects, one problem in capturing interaction data with mobile devices is that the code of the application has to be manipulated to get the application known to the log file recording tool and to get the desired log data. Another problem was that participants often forgot to click the "finished"-button. Therefore, the researcher had to control and complete the marks at the end of a task after the tests. It would be desirable for future tests if the analyzing tool offered an easier way to accomplish this. More analysis functionality would be helpful, too, e.g., the graphical representation of deviation from the ideal task solution.

Some methodological aspects have to be considered. To identify as many usability problems as possible by just analysing the log data in the presented way, the evaluator has to know the application really well. As the results of this study show, recording only screenshots and clicks is not sufficient to identify all problems. Therefore, a screen-capturing tool that produces a video of the screen of the mobile device would be a better solution. But, as long as there are restrictions in storage capabilities, capturing screenshots is the only alternative.

Log files also cannot help to find out whether the test person is missing any information on a screen or functionality within the application. Therefore, thinking aloud data would be helpful. However, it is also challenging to get good think-aloud data. Participants had to concentrate very strongly on interacting with the unfamiliar system, and therefore, often forgot to think aloud. It also could have influenced the interaction time and the way of interacting by reflecting upon it. Furthermore, to think aloud in the form of talking to yourself can be embarrassing in public places like museums. Thus, alternative usability testing methods under real conditions are necessary.

The recommended method would be to test without thinking aloud, and instead, complete the method with a retrospective thinking aloud (video confrontation). Showing the user the screenshots or video of the interaction with the mobile application will certainly help the participant to recall everything and verbalize the usability problems not appearing in the log files. The effectiveness of this method has to be tested in further studies.

Practitioner's Take Away

The following ideas for a practitioner came out of this research:

- Evaluate mobile usability in the field with unobtrusive methods.
- Avoid thinking aloud method for field testing.
- Use remote mobile usability testing for gaining practical relevant usability data.
- Implement log file analysis if it is possible for detecting problems in interaction structure.
- Combine objective data with subjective methods, e.g., confronting the user with captured screenshots (video confrontation).

Acknowledgements

This work as part of the IKAROS-project was founded by the Federal Ministry of Education and Research Germany.

References

Bradner, Erin (2004) Keeping your distance: remote usability testing or the lab – which is the best? *User Experience*, 6,11

Branch, J. L. (2000) Investigating the Information-Seeking Processes of Adolescents: The Value of Using Think Alouds and Think Afters, *Library & Information Science Research*.

Edwards, A. L. (1957) *The Social Desirability Variable in Personality Assessment and Research*. New York, NY: Holt, Rinehart, & Winston.

Jambon, F. (2006) Reality Testing of Mobile Device: How to Ensure Analysis Validity? *CHI 2006*, April 22-27 2006, Montreal, Canada.

Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. (2005). Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. *Journal of Usability Studies*, 1(1), 4-17.

Kawalek, J., Herrmann, C. & Stark, A. (2006) Mobile Anwendung aus ergonomischer Perspektive. In Sieck, J. & Herzog, M. A. (Eds.), *Wireless Communication and Information* (pp. 147-156). Aachen: Shaker.

Rubin, J. (1994). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New Jersey: John Wiley & Sons, Inc.

Schweibenz, W. & Thissen, F., (2003) *Qualität im Web. Benutzerfreundliche Webseiten durch Usability Evaluation*. Berlin: Springer.

Van den Haak, M.J., de Jong, M.D.T. (2003) Exploring two methods of usability testing: concurrent versus retrospective think-aloud protocols. *IEEE Professional Communication Conference, 2003*. IPCC 2003. Proceedings. IEEE International. 21-24 Sept. 2003.

About the Authors

Jürgen Kawalek
From 1994 to 1997, Professor Dr. Jürgen Kawalek was director of the European Teleservice evaluation Centre at the PRZ (TU Berlin). Since 1998, he has been a professor for multimedia- psychology at the University of Applied Sciences in Zittau/Görlitz. He is head of the IKAROS project, as well as the Center for eLearning.

Marcel Riebeck
Marcel Riebeck works as research associate at the University of Applied Sciences Zittau/Görlitz, Germany. He holds a graduate degree in Communication Psychology and has a special interest in Human Computer Interaction. Marcel is currently project manager of an interdisciplinary research project that focuses on log file analyses for understanding user needs.

Annegret Stark
Annegret Stark received her graduate degree in Communication Psychology in 2004. She worked as research associate at the University of Applied Sciences in Zittau/Görlitz focusing Human Computer Interaction with mobile devices. Currently, she works at the Center of eLearning.