CO-128

COMPARISON OF EVALUATION METHODS FOR FIELD-BASED USABILITY STUDIES OF MOBILE MAP APPLICATIONS

BURGHARDT D., WIRTH K.
Dresden University of Technology, DRESDEN, GERMANY

ABSTRACT

Empirical usability studies are crucial for an improved understanding of cartographic communication processes. User oriented studies of mobile applications has high demands on the evaluation methods. The user context plays an important role during the evaluation of user friendliness of mobile applications, thus it is important to provide realistic conditions during the usability evaluation. Because of multitasking requirements on the test persons during the field test, a much higher error rate occurs with the task fulfilment. Therefore also a higher probability exists of identifying dedicated usability problems.

The aim of this research was to compare several evaluation methods for the test of mobile map applications in a natural environment. During the test the users had to solve different orientation task in an unknown area. The test person were categorised into three different age groups (12-17 y., 18-60 y., > 60 y.). The evaluation methods investigated with this research were selected from observation methods, such as thinking aloud, audio and video recording and from survey methods such as interview or questionnaire.

The results show, that on one side field test require much more time for planning and execution than laboratory usability testing. On the other side very different requirements exist on the technique of data capturing. For example video and audio recording have to be much more sensible, to guarantee sufficient quality and they have to be robust against changing weather conditions. The analysis and interpretation of the data is more complex as disturbance from background noise and sun reflexion exacerbate the evaluation. Data acquisition becomes harder during field tests in cases where the evaluator could not see the evaluated object. For example it has occurred that the test person obliterated the screen with his hand or has turned away from the camera. Parallel recording from different sides is normally not practicable. With all these influence the efficiency of the field test suffers, because more time and more sensible techniques causes higher costs.

The statistical analysis was carried out for the different age groups. Thereby the time was considered to fulfil the task and several error categories were distinguished. Critical events could be identified and can be displayed through videos. The paper finishes with a comparison of field based evaluation method and recommendations for usability studies to test mobile map applications.

BACKGROUND AND OBJECTIVES

The newest technology achievements in the area of mobile devices represent a modern platform for cartography to provide geospatial information. To put on the market a high-quality product, usability examinations has to be carried out. The characterisation of the usability evaluation methods has to consider a number of different criteria. According to the moment of the evaluation an "ex-ante", "on-going" or "ex-post" perspective can be taken, which can be further differentiate in either summative (result oriented) or formative (process oriented) evaluation. If experts execute an evaluation on the bases of directives, it is called an analytical evaluation. In contrast empirical evaluations are carried out with the help of potential users. If the focus of the evaluation is put on separate system components, one calls this a micro evaluation, while the evaluation of the whole system is termed as macro evaluation.

Taking advantage of mobile applications the consideration of context plays an important role (Tamminen, 2004; Gartner et al., 2004). Context models are presented for example by Nivala and Sarjakoski (2003), Reichenbacher (2004) and Kjeldskov et al. (2005). On account of the new demands with context consideration of mobile applications the usability evaluation methods must be adapted as well. Usability evaluations of mobile applications are described for example by Gorlenko and Merrick (2003), Kaikkonenn et al. (2005) and Fiotakis et al. (2009). Specific field-based evaluations methods are investigated by van Elzakker et al. (2008). A comparison of laboratory and field test for the usability evaluation of mobile applications can be found in Duh et al. (2006) and Hoai (2009).

The field-based usability methods analysed within in our empirical study are subdivided in observation methods and survey methods. The observation methods are audio and video recording both by eyewear glasses and camcorder, while the applied survey methods are the questionnaire, the unstructured and the half-structured interview. The empirical study is carried out with 18 test persons from three different age groups (12-17 y., 18-60 y. and older than 60 y.).

APPROACH & METHODS

The empirical field study requires thorough planning with definition of aims and methods (Barnum, 2002). The compliance of test rules guarantees that the test runs will be comparable. The candidates had to solve different navigation tasks with a GARMIN Oregon 400t device (see Fig. 1). The mobile device contained a base map Europe 1: 100.000 with overlay information through Open-Street-Map data for the provision of further interactive functionality.

Fig. 1: GARMIN Oregon 400t and test route on the campus of the university

In a first step the test person should be made themselves familiar with the mobile device. The second task was to insert coordinates into the device and find a specific building on the campus of the university. In a third step the candidates had to identify a point of interest by using the name of a restaurant. The forth task was the interview with the supervisor and to fill out the questionnaire.

The decision about required number of test persons was influenced by the findings of Nielsen (1989) and Virzi (1992) that between five and six persons found approximately 80% of usability problems. Therefore the field test was carried out with six candidates per age group, in total 18 test runs. The test equipment can be seen in Fig. 2. Video and audio information are recorded with camcorder from Panasonic SDR-S100 and a mobile eyewear recorder.

Fig. 2: Camcorder Panasonic SDR-S100 (left) and Mobile Eyewear Recorder (right) for video and audio recording.

A pilot test was carried out with a complete run through. Thereby the functionality and completeness of the used equipment has been checked. The time for the test execution were measured as well. As result of the pilot test the test route and the tasks for the tests were slightly optimised.

Fig. 3: Field-based usability study with test person, supervisor and camera operator.

The analysis of data captured during the field test is based on the following error categories (Duh et al., 2006):

- Cosmetic problems: The user is disturbed during task execution. It is still possible to work further. Small delays might occur, but the task can be solved finally.
- Serious problems: The user has difficulties to proceed with test execution. Much more time is required to solve the task. The error recurred frequently across test persons (at least in 50% of the cases).
- Critical problems: The user is not able to complete the task without interference from supervisor. The error recurred across all test series.

Quantitative measures were achieved by counting the occurred errors in dependence of age group and applied evaluation method. With the t-test the significance of the results are proved. The recorded errors were aggregated to “identified errors”, as one identified error could be recorded several times.

RESULTS

The evaluation of the field-based usability study focuses on three aspects:

- Comparison of data capturing methods
- Findings of the test candidates
- Findings of the supervisor

The results of the quantitative comparison of data capturing methods are based on the number of identified errors. Fig. 4 shows the distribution over error categories for the different field-based evaluation methods. For the detection of critical and serious problems the “thinking aloud” method is most effective. The highest number of cosmetic problems was identified by the half-structured interview. Therefore it could be concluded that a combination of “thinking aloud” method and video recording with eyewear are most suitable for the evaluation of mobile devices in the field. For a fast and easy evaluation it might be possible to replace the half-structured interview with open and closed questionnaire. An ineffective method is the un-structured interview and should be avoided.

Survey methods provide only an indication on possible errors. Further inspection and weighting of errors could be carried out by the evaluator with the observation methods. Advantages of the survey methods are the suitability of showing an average opinion, because the test persons are able to give a summary of subjective feeling.

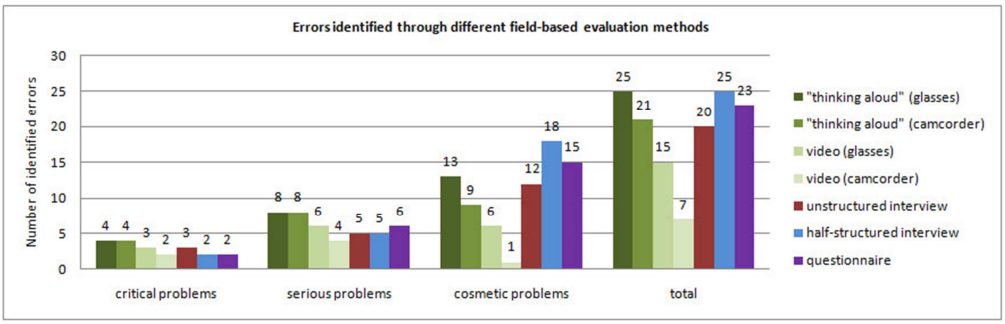

Fig. 4: Number of identified errors for the different error categories and in dependence of data capturing methods

Fig. 5 presents the results of quantitative comparison for the different age groups. The senior test persons identify in an empiric usability analysis the most critical problems so to say the high-quality mistakes. The middle age group is suitable for the refinement of a product. Young test person seems to be inappropriate for this kind of evaluation, because they are difficult to recruit and they tolerate a lot of the errors. Young persons are often not self-confident enough to "blame" the tested device.

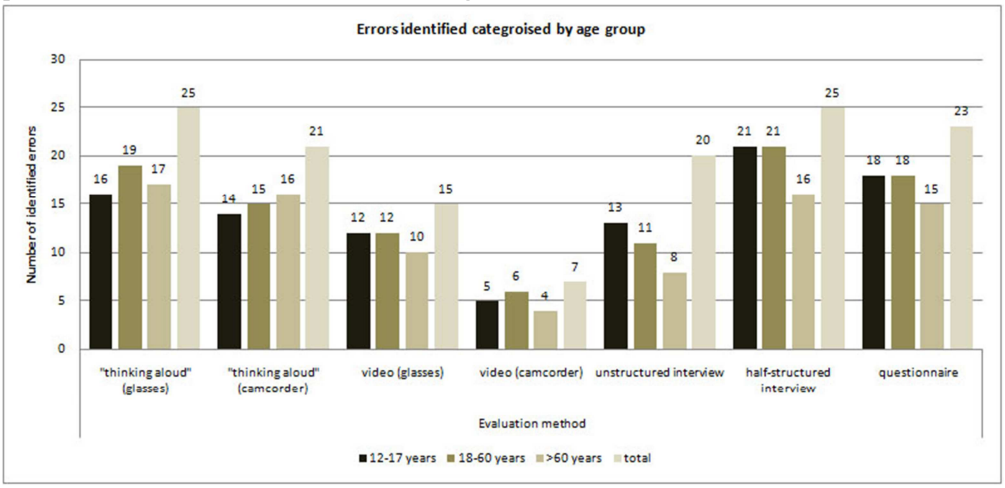

Fig. 5: Number of identified errors for the different age groups in dependence of data capturing methods

Fig. 6 gives an indication of time required for test execution and data analysis for the different evaluation methods. In total around 12-13 hours are required for data capturing and analysis per person. Most time consuming are the analysis of video recording. For the "thinking aloud" method the analysis with the eyewear recorder requires two times of the field test, while the camcorder requires two and half times. The reason is the good audio quality of the eyewear recordings, whereas the camcorder captures much more background noise. The surveys could be analysed easier. The test person speaks direct into the microphone and articulates more selective.

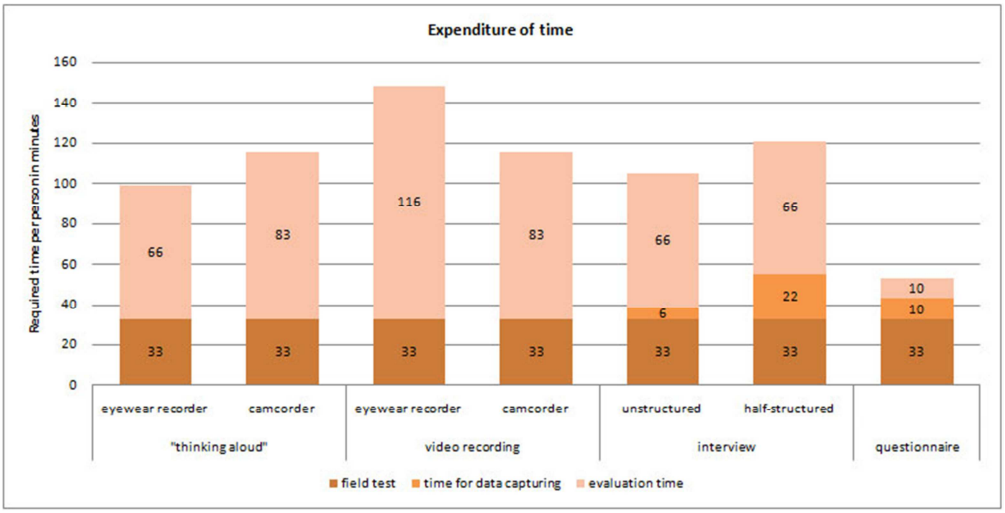

Fig. 6: Required time for the field-based usability methods

		device	result type	advantages	disadvantages	identified errors: critical problems	identified errors: serious problems	identified errors: cosmetic problems	identified errors: total
observation methods	„thinking aloud"	eyewear recorder	❖ weighting of errors through evaluator ❖ metric data: * time measurement * number of errors ❖ qualitative data: * error * positive comments * negative comments	❖ directly on body ❖ good speech recording	❖ very sensible technique	4	8	13	25
		camcorder		-	❖ re cording person ❖ background noise ❖ recording person is disturbing factor	4	8	9	21
	video recording	eyewear reorder		❖ view on display ❖ test person recorded	❖ not always complete usable ❖ unusual posture	3	6	6	15
		camcorder		❖ whole situation ❖ always equally useable	❖ no view on the display ❖ Barriers on the way	2	4	1	7
survey methods	interview	unstructured	❖ qualitative data ❖ no weighting of errors ❖ comments on errors ❖ general opinion	❖ shows emotional main point	❖ danger of useless information	3	5	12	20
		half-structured		❖ dialog ❖ detailed answers	❖ evaluator necessary ❖ needs interview guideline	2	5	18	25
	questionnaire	open questions	see interview	❖ personal explanations	❖ test person writes only short answers	2	6	15	23
		Likert-scale	❖ average tendency	❖ easy evaluation	❖ identifies no errors	-	-	-	-

Tab. 1: Quantitative and qualitative comparison of field-based evaluation methods

CONCLUSION AND FUTURE PLANS

The presented article gives an overview of advantages and disadvantages for the applied field-based evaluation methods. During this empirical evaluation only one type of questionnaire (Likert scale and open questions) were applied. Thus an open question refers to the suitability of questionnaires with a different structure. It might be worth to evaluate different video recording devices as well, for example a camera which is installed directly on the mobile device. All test persons were observed individually. An alternative evaluation scenario would be the evaluation of a group of people. This way more communication between participants could be expected. In consequence test execution might be easier for the participants and further spoken information also about the usability of the mobile applications could be captured.

REFERENCES

Barnum, C. M. (2002): Usability testing and research. Longman, New York, München.

Chan, S. & Fang, X. (2003): Mobile commerce and usability. In: Siau K., Lim E. Advances in Mobile Commerce Technologies. In: Idea Group Publishing, Hershey, PA, pp. 235-257.

Duh, H. B.-L.; Tan, G. C. B. & Chen, V. H.-h. (2006): Usability Evaluation for Mobile Device: A Comparison of Laboratory and Field Tests. In: ACM International Conference Proceeding Series, Proceedings of the 8th conference on Human-computer interaction with mobile devices and services; Volume 159, pp. 181–186.

Fiotakis, G.; Raptis, D. & Avouris, N. (2009): Considering Cost in Usability Evaluation of Mobile Applications: Who, Where and When. In: Lecture Notes In Computer Science. Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Volume 5726, Issue 1, pp. 231-234.

Gartner, G.; Frank, A. & Retscher, G. (2004): Fußgängernavigation im städtischen Umfeld-Das NAVIO-Projekt. In Kartographische Schriften: Der X Faktor-Mehrwert für Geodaten und Karten, Volume 9, Issue 14, pp. 82-89. Kirschbaum Verlag GmbH, Bonn.

Gorlenko, L. & Merrick, R. (2003): No Wires Attached: Usability Challenges in the Connected Mobile World. In: IBM Systems Journal, Volume 42, Issue 4, pp. 639-651.

Hermann, F.; Bieber, G. & Duesterhoeft, A. (2003): Egocentric Maps on Mobile Devices. In: G. Bieber and T.Kirste (Eds.) Proceesing of the 4th International Workshop on Mobile Computing, IMC 2003, IRB Verlag, Stuttgart, pp. 32-37.

Hoai Anh, P. (2009): Remote Test vs. Labor Test. Diplomarbeit. Technische Universität Dresden. Fakultät für Informatik. Institut für angewandte Informatik, Dresden.

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

Kjeldskov, J.; Skov, M. B.; Als, B. S. & Hoegh, R. T. (2004): Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In: Proceedings of the 6th International Mobile HCI 2004 conference, Springer-Verlag, Berlin, pp. 61-73.

Kjeldskov, J.; Graham, C.; Pedell, S.; Vetere, F.; Howard, S.; Balbo, S. & Davies, J. (2005): Evaluating the Usability of a Mobile Guide: The Influence of Location, Participants and Resources. In: Behavior & information Technology, Volume 24, Issue 1, pp. 51-65.

Kjeldskov, J. & Stage, J. (2004): New techniques for usability evaluation of mobile systems. In: International Journal of Human-Computer Studies. Volume 60, Issues 5-6, pp. 599-620.

Nivala , A.M. & Sarjakoski, L.T. (2003): Need for Context-Aware Topographic Maps in Mobile Devices. In: Virrantaus, K., H. Tveite (Eds.); Scan GIS´2003-Proceedings of the 9th Scandinavian Research Conference on Geographical Information Science, Espoo, Finnland, pp. 15-29.

Po, S.; Howard, S.; Vetere, F. & Skov, M. B. (2004): Heuristic Evaluation and Mobile Usability: Bridging the Realism Gap. Mobile Human-Computer Interaction – MobileHCI 2004, Lecture Notes in Computer Science. Volume 3160, pp. 591-592.

Reichenbacher, T. (2001): The World in Your Pocket-Toward a Mobile Cartography. Proceedings of the 20th Cartographic Conference, August 6-10, Beijing, China, CD-Rom.

Reichenbacher T. (2004): Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices. Doctoral thesis, Institute for Photogrametry and Cartography, Technical University of Munich, Verlag Dr. Hut.

Ryu, Y.S. (2005): Development of Usability Questionnaires for Electronic Mobile Products and Decision Making Methods. Doctoral thesis, Blacksburg: Virginia State University.

Schneidermeier, T. (2009): Usability of Mobile Devices: Untersuchungen zur Gebrauchstauglichkeit mobiler Endgeräte am Beispiel Apple iPhone. Magisterarbeit. Universität Regensburg, Information und Medien, Sprache und Kultur, Regensburg.

Tamminen, S., Oulasvirta, A., Toiskallio, K., Kankainen, A. (2004). Understanding mobile contexts. Special Issue of Journal of Personal and Ubiquitous Computing. Volume 8, Issues 2, pp. 135-143.

Umlauf, M.; Pospischil, G.; Niklfeld G. & Michlmayer, E. (2003): LoL@, a Mobile Tourist Guide for UMTS. In: Information Technology & Tourism, Volume 5, Issue 3, pp. 151-164.

van Elzakker, C.P.J.M.; Delikostidis, I. and van Oosterom, P.J.M. (2008). Field - based usability evaluation methodology for mobile geo - application. The Cartographic Journal, Volume 45, Issue 2, pp. 139-149.

Virzi, R. (1992). Refining the test phase of usability evaluation: How many subjects is enough? Human Factors Volume 34, pp. 457-486.

Weiss, S. (2002): Handheld usability. Hoboken, NJ: John Wiley & Sons.