

Considering Cost in Usability Evaluation of Mobile Applications: Who, Where and When

Georgios Fiotakis, Dimitrios Raptis, and Nikolaos Avouris

Human-Computer Interaction Group, E&CE Department, University of Patras,
Ypatias Str., GR-26500, Greece
{fiotakis, draptis}@ece.upatras.gr, avouris@upatras.gr

Abstract. As computing moves from desktop to outdoor everyday life activities, usability evaluation must take into account new aspects, related with mobility and space. In this paper the effectiveness of established usability evaluation methods is examined through an extensive case study. The usability of an educational mobile museum application was tested using three different methods: inspection by experts, use in lab setting and use in the field. Results indicate that each method uncovers different types of problems at different cost. We believe that a combination of these methods may be worthwhile since their produced results seem to be complementary.

Keywords: Usability evaluation methods, mobile applications, user studies, expert evaluation, lab, field, cost, severity, effort.

1 Introduction

In many cases mobile applications are designed to facilitate everyday activities, related to specific contexts of use. Such mobile applications inherently require from their users not to limit their interaction to the device screen but additionally to interact through the application with the typical environment of use. Since mobile applications support everyday activities that in many cases take place in diverse environments, their design needs to take into account and cope with situations where the user's attention is allocated to stimuli diffused in the environment. Designers need to understand all these aspects and design mobile applications aiming to be embedded smoothly both in the human activity and the environment. An additional challenge is to tackle the problems that inherently characterize mobile technology, such as ergonomics of devices, difficulty with data entry, low processing capabilities, unreliable connectivity, etc., [7]. All these design considerations force us to adapt the traditional methods used for formative evaluation of desktop applications in order to effectively evaluate the mobile ones.

In this paper, we discuss a case study of usability evaluation of a mobile application that involved three different approaches. The application is a collaborative mobile game for a cultural museum. Three different evaluation studies of this application were performed and their findings are discussed in the following sections.

2 Questions Related to the Cost of Mobile Usability Evaluation

Due to the continuous evolution of mobile technology and the extensive competition of new mobile services and applications, their design life-cycle tends to be short. By considering that evaluation is a fundamental part of design, it seems that the optimization of the evaluation process is of high importance. Optimization can be achieved by providing appropriate answers to three fundamental questions related to the *cost*. Cost is perceived in terms of time, effort and efficiency.

The first two questions are: *Who?* and *When?* We propose the introduction of expert inspection techniques in early design phases prior to methods that involve typical users. The hypothesis is that a few experts can discover a significant number of problems and require a minimal amount of human and time resources. Then, after correcting the usability problems that the experts have identified, representative users can be involved in order to evaluate the updated version of the application and identify additional possible usability problems, of different nature sometimes, relating, often, with unexpected use of the application. Although the idea of combining user and expert based methods has already been suggested, its effectiveness has not been extensively verified in empirical research of mobile evaluation studies [5]. Another issue is related to the types of the problems that expert-based and user techniques can track. For example, expert based methods have not received much attention in the literature of mobile usability evaluation due to their inadequacy to record contextual influences over use [6].

The third question is: *Where?* Due to the dynamic context in which mobile applications are used many researchers have noted the importance of field evaluation ([3], [6]). However, some claim that there is little added value of performing the evaluation in the field, considering the required resources [4]. A recent survey shows that the situation tends to change as field studies gain considerable ground [2]. We've tried to answer these questions through the empirical study discussed next.

3 Case Study: Evaluating an Educational Mobile Application

The study consisted of three different tests. Study S#1 took place in a usability laboratory where three usability experts inspected the interface by following a given scenario. The experts interacted with the interface through an emulator in a desktop computer. Each expert coded the problems found according to a predefined coding (heuristics) and a severity scheme. Study S#2 involved users in the laboratory. A team of 8 primary school students were asked to follow the same scenario using the mobile application on a Tablet PC that contained a 2D map of the museum and a simulator of the PDA screen. The students manipulated a virtual character on the map, moved across rooms, selected exhibits, and the consequences of their actions to the mobile application were presented on the simulator. Each student interacted alone with the application. Study S#3 took place in the typical environment, inside the museum. A group of 8 primary school students (different than S#2) were requested to accomplish the same scenario using the mobile application on PDAs. In these last two tests users were asked to follow the think aloud protocol. For each one of S#2 and S#3, three usability experts were asked to evaluate the usability of the mobile application, by

observing user activity recorded in the form of logfiles and video. Data analysis was facilitated by the ActivityLens tool [1]. All experts that used ActivityLens had prior experience with the tool.

The three experiments identified 32 usability problems overall. 20, 17 and 18 usability problems were identified in studies S#1, S#2, S#3 respectively. The contribution of each method is shown in figure 1b. Considering the severity, in relation to the users, of these usability problems, we see that the expert based method identified more critical and serious problems than the others (figure 1a). Moreover, most problems found solely by field evaluation were categorized as cosmetic (80%) and none of them as critical.

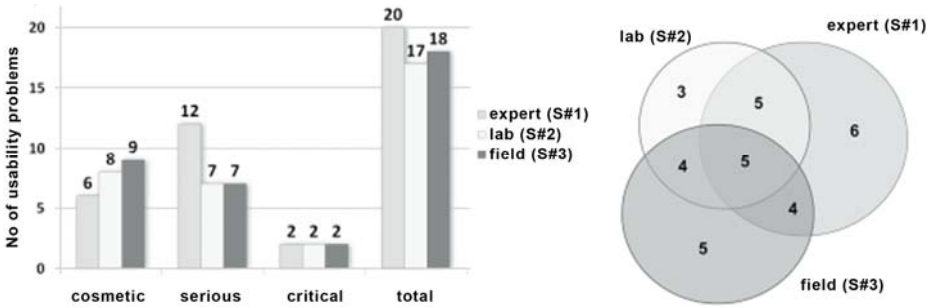


Fig. 1. (a) Severity ratings of usability problems per method, (b) Method performance

If we consider cost as the time consumed to perform the usability evaluation, then we found that the field evaluation (S#3) proved to be the most time consuming method requiring almost triple the time (310 min) than the lab evaluation (S#2) (110 min), while the first study (S#1) required just 32 min. Times mentioned include experiment preparation and analysis time. These differences in time is due to the fact that the experts had to deal with large amounts of observation data and that required considerable effort. The field evaluation required analysis of more sources of observation data (5 video files + logfiles) to capture the continuous movement of users in the field compared to the laboratory evaluation (2 video files + logfiles).

27 usability problems were found in relation to the user interface (84.3% of identified problems), 20 of them were located during S#1, 7 new ones were produced by S#2. From these figures, one might assume that field experiments are useless since they require a lot of effort and produce little results compared to other methods. But, if we consider that a mobile application extends outside the PDA's screen then there are going to be cases that these few and in our case characterized by evaluators as cosmetic problems, might have a significant impact on the design. The field experiment produced 3 problems related to environmental aspects (server delays, certain positions with low WiFi signal and problems in scanning the RFID tags), 1 problem characterized as “using the application in unexpected way” (a group of students developed a game strategy that allowed them to win by violating the purpose of the game) and 1 cosmetic problem related to the position of three RFID tags at a height that many children could not reach. The last, cosmetic for the evaluators but severe for the users, problem proved very significant since it changed many design decisions

concerning the use of RFID tags in the museum. Until then, all RFID tags were placed next to the exhibits. To cope with the identified usability problem some of them were moved lower and that introduced the problem of finding a mechanism to visually indicate to which exhibit each RFID tag corresponded. To solve this problem many different visual cues were placed on the RFID tags and at least three independent tests were performed in order to reach the final solution.

An in-depth inspection of the categories of problems revealed that the ones that are related to the PDA's interface (navigation and consistency problems) were mostly found by experts. On the contrary, the problems that concern the affordances of the interface objects, the visibility of services and the interaction with the surrounding space were found by methods that involved users.

4 Conclusions

In this paper we presented the results from evaluating a mobile application using three different methods: expert evaluation, laboratory and field user studies. The results indicate that the optimization of the evaluation process can be achieved through the sequential application of different evaluation methods throughout the design life cycle of a mobile application. Expert based techniques can locate, at low cost, various crucial usability problems related to the user interface in early development phases. Lab experiments can also identify at low cost many problems related to the user interface and some related to the activity as a whole. The most time consuming and hard-to-apply methods are the field experiments. However, they can deliver significant qualitative results concerning the real practices of users and the effects of the environment on the activity that cannot be found by other methods. As a conclusion, it seems that if all methods are combined they can give an explicit view of usability problems since their outcomes are in great extend complementary.

References

1. ActivityLens, <http://hci.ece.upatras.gr/activitylens>
2. Avouris, N., Fiotakis, G., Raptis, D.: On measuring usability of mobile applications. In: VUUM 2008 Workshop, pp. 38–44. IRIT Press, Toulouse (2008)
3. Goodman, J., Brewster, S., Gray, P.: Using Field Experiments to Evaluate Mobile Guides. In: HCI in Mobile Guides. Workshop at Mobile HCI 2004, Glasgow (2004)
4. Kjeldskov, J., Skov, M.B., Als, B.S., Høegh, R.T.: Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In: Brewster, S., Dunlop, M.D. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 61–73. Springer, Heidelberg (2004)
5. Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods, pp. 25–62. John Wiley & Sons, Chichester (1993)
6. Po, S., Howard, S., Vetere, F., Skov, M.: Heuristic Evaluation and Mobile Usability: Bridging the Realism Gap. In: Brewster, S., Dunlop, M.D. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 49–60. Springer, Heidelberg (2004)
7. Zhang, D., Adipat, B.: Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. *Int. Journal of Human-Computer Interaction* 18(3), 293–308 (2005)