Reality Testing of Mobile Devices: How to Ensure Analysis Validity?

Francis JAMBON
CLIPS-IMAG / MultiCom
BP 53, 38041 Grenoble cedex 9, France
e-mail: francis.jambon@imag.fr
phone: +33 4 76 51 45 10

Figure 1: The user (left) with the facilitator (right) and the cameraman (background) in one of the building corridors.

CHI 2006, April 22–27, 2006, Montreal, Canada.
ACM 1-xxxxxxxxxxxxxxxxx.

Abstract
This article describes, in the first part, the motivations of the MultiCom group at CLIPS-IMAG laboratory in reality testing of mobile devices. The second part of the article is dedicated to the methodological aspects of data collection and analysis when evaluating the usability of a mobile device in a quasi-realistic context.

Keywords
Remote Usability Testing, Data Collection, Evaluation Techniques, Mobile Devices.

ACM Classification Keywords
H.5.2 [User Interfaces]: Evaluation/Methodology

Introduction
The MultiCom group at CLIPS-IMAG Laboratory is involved in testing the usability of various kinds of interactive systems and environments, for instance context-aware personal digital assistants in intelligent environments. To do so, we usually simulate these specific environments in our usability laboratory which feature a large experiment room (100 m^2).

Nowadays, an increasingly number of projects involve usability evaluation "in the wild" [1, 2]. Furthermore, some of these projects must take place in the real

everyday life of the users and usually last more than one single day, for instance when context-aware mobile phones are used. The classic evaluation methods adapted to these experimental situations are diaries and questionnaires. Unfortunately, data extracted from diaries and questionnaires suffer from a lack of precision in usage statistics. Moreover, user usually note down usability problems in diaries at the end of the day, and so may forget some.

That is why some authors [3, 4] , create new mobile multimedia recording techniques. These techniques usually succeed in mimicking the usability laboratories facilities in the wild. But, in some situations, the use of cameras and microphones is not acceptable for ethical reasons. Demumieux et Losquin propose to guess real phone usage thanks to the events collected from the mobile phone applications and its operating system [5]. However, this technique has a severe drawback: without video, it is usually very difficult to explain the user behavior, because the context is missing in the events that can be recorded. So, the results of this type of data collection are limited to statistics on the mobile phone applications usage: date, frequency, duration… In other words, it is possible to answer "when" and "what" but usually not "why".

Motivations

Kjeldskov et al. explored the added value of evaluating the usability of context-aware mobile systems in the field, and concluded that there is little added value to carry out in-the-field experiments [6]. In fact, Kjeldskov et al. were using a lot of visible hardware in the related experiments (clip-on camera, etc.), so the situation cannot be considered as ecologic. Moreover, a more recent article of the same first author have more mixed conclusions [7]. Isomursu and al. detect similar outcomes experimenting the "Experience Clip" technique [8]. In the MultiCom group, we aim at carrying out reality testing experiments of mobile devices "in the wild" without any unobtrusive piece of hardware and, when it is compulsory, without video.

The methodological problem here is that is not possible to know whether users act in a different way in ecologic situations or not, because the widely used reference observation technique –video– maybe a severe bias in these ecologic situations. Alternatives to video exist, for instance diaries or traces of activity collected from the device. The limits of diaries are that this is not possible to trust records in every situation: user forgot things or not note down them regularly. That is why we focus on the analysis of traces collected from the devices only, but keeping in mind that other techniques may be useful too (i.e. diaries, critical incidents technique) as complementary sources of information. Our final objective is to obtain the same level of correctness in the analysis of traces as in the analysis of the video. To carry out these experiments in the wild, we first need to validate both technical and methodological aspects in quasi-realistic experiments.

Experiments

Context

The experiments we carried out focus on a context-aware personal digital assistant that could also give geo-localized information to the user. The chosen scenario was simulating the indoor guiding of an invited executive in a professional context. The experiments took place in a real professional building during working hours. A facilitator and a cameraman were coming along with the user (see figure 1). The facilitator was in

charge of asking specific questions during the scenario execution. The main recording systems were situated in the observation room, located in one of the central rooms of the building.

Methodology
Two different types of data collection methods were used during these experiments. We named them *analogical* and *digital*. The analogical data collection refers to classic usability laboratory data collection: video of the context (DV camcorder), voice of the user (wireless microphone) and PDA screen (VNC-like tool). The digital data collection refers to the recording of user actions on the device, system feedbacks and device localization in the building. Except for the DV camcorder, all these data were monitored and stored in the observation room thanks to wireless technologies (HF or WiFi). The digital data collection were possible thanks our specialized software bus –Usybus– that enable the real-time annotation of the events in Noldus™ The Observer® tool [9].

The analogical and digital data were monitored by an usability professional and a computer science professional respectively, and stored in two different portable computers. Then, at the first step, the analyses of these two data sets were performed separately. No communications were allowed between the professionals. In a second step, once the two reports achieved, another professional cross-checked and compared the two sets of results, and if necessary, verified their validity with the raw video records. This way, we wanted to identify precisely what is the "analysis power" of each kind of data.

Results
Three severe usability problems of the device mockup had been detected by both methods. To do so, the analogical analysis used user think aloud comments and the digital analysis used the time duration calculation (subtask too long) and the difference detection (actions not executed).

Some usability problems could not be detected by the digital analyses because these problems do not change the sequence of user interactions and systems feedbacks. In these situations, the user usually asked the facilitator for help, so the user interactions were, at last, seen as regular. The facilitator presence is clearly a bias when analyzing digital data without video reference.

A severe usability problem were suspected by the digital analysis with the obstruction specification function, because for three users, this subtask lasted more than expected. In fact, the user and the facilitator were talking about a previous usability problem! Here also, the facilitator presence creates a bias too.

Conclusion and future work

These experiments prove that it is possible to evaluate the usability of a mobile device in the wild, without video. The false positive and false negative results are mostly caused by the facilitator presence. However, we must be aware that the usability problems detected may remain unexplained without the help the video. We suggest in this case to inquiry from the user about it afterwards, for instance, by replaying the interaction sequences. Another solution is to detect in real time the usability problems and inquire from the user about it with a method similar to [10].

We also realize that Reality Testing is difficult and costly. Hardware software, network as well as logistics issues are severe. For instance, two of our twelve experiments must had been removed from the analysis set due to unpredictable technical problems. In real-world, the failure rate must be quite higher. We now plan to validate in the wild the methodology tested in this quasi-realistic experiments. To prepare it, we now are validating the GPRS networking issue and the long duration video backup and then will carry out in the wild usability tests in Q3 2006.

Acknowledgements

The author would like to acknowledge the partners and the founding organizations of the Franco-Finnish project Adamos (Mobile Adaptive Services – Design Parameters and User Experience Factors) which give us the opportunity to carry out these experiments.

Short Biography of the Author

Francis JAMBON is Assistant Professor of Computer Science at University Joseph Fourier (Grenoble, France). His research interests are remote usability testing, automated usability evaluation, mobile devices, and user interface verification and validation.

References

[1] Waterson, S., Landay, J.A. and Matthews, T. In the lab and out in the wild: remote web usability testing for mobile devices, in *Proc. Conference on Human Factors in Computing Systems (CHI'02) extended abstracts* (Minneapolis, Minnesota, USA, 2002), ACM Press, pp. 796-797.

[2] Goodman, J., Brewster, S. and Gray, P. Using Field Experiments to Evaluate Mobile Guides, in *Proc. HCI in Mobile Guides, workshop at MobileHCI'2004* (Glasgow, UK, 2004).

[3] Roto, V., Oulasvirta, A., Haikarainen, T., Lehmuskallio, H. and Nyyssönen, T. *Examining mobile phone use in the wild with quasi-experimentation*. Helsinky Institute for Information Technology (HIIT), Technical Report 2004-1, August 13 2004.

[4] Salembier, P., Kahn, J., Calvet, G., Zouinar, M. & Relieu, M. "Just follow me" - Examining the use of a multimodal mobile device in natural settings, in *Proc. HCI International Conference* (Las Vegas, USA, July 22-27, 2005), Lawrence Erlbaum Associates.

[5] Demumieux, R. and Losquin, P. Gather customer's real usage on mobile phones, in *Proc. 7th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI'05)* (Salzburg, Austria, 19-22 September, 2005).

[6] Kjeldskov, J., Skov, M.B., Als, B.S. and 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 *Proc. Mobile Human-Computer Interaction – MobileHCI'2004* (Glasgow, UK, September 13-16, 2004), Lecture Notes in Computer Science, Springer-Verlag GmbH, 3160, pp. 61-73.

[7] Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S. and Davies, J. Evaluating the Usability of a Mobile Guide: The Influence of Location, Participants and Resources. *Behaviour & Information Technology*. 24, 1 (2005), pp. 51-65.

[8] Isomursu, M., Kuutti, K. and Väinämö, S. Experience clip: method for user participation and evaluation of mobile concepts, in *Proc. Eighth conference on Participatory design: Artful integration: interweaving media, materials and practices* (Toronto, Ontario, Canada, 2004), ACM Press, pp. 83-92.

[9] Noldus. *Noldus™ The Observer® Video Pro v5*. http://www.noldus.com, 2005.

[10] Arhippainen, L., Rantakokko, T. and Tähti, M. Mobile Feedback Application for Emotion and User Experience Collection, in *Proc. Proactive computing workshop (PROW 2004)* (Helsinki, Finland, 25-26 November, 2004), pp. 77-81.