

“Just follow me”. Examining the use of a multimodal mobile device in natural settings

*Pascal Salembier**, *Julien Kahn***, *Guillaume Calvet**, *Moustapha Zouinar***,
*Marc Relieu***

*GRIC-IRIT, Université Paul Sabatier, 118, rte de Narbonne, 31062 Toulouse Cedex

**France Télécom R&D-DIH/UCE, 38-40 rue Gal Leclerc, 92794 Issy-les-Moulineaux

salembier, calvet@irit.fr; julien.kahn, moustapha.zouinar,
marc.relieu@rd.francetelecom.com

Abstract

This paper describes a method and a piece of equipment for capturing data about the use of a mobile multimodal device in natural settings. The equipment includes an instrument worn by the user to capture his or her perspective while moving, a means of capturing a wider view of the context and a multimodal data logger.

1 Introduction

The growth of mobile systems (e.g. Personal Data Assistants, mobile handsets, etc.) sets new methodological and technical problems for empirically studying their use and their usability in natural settings. The mobility of a user implies not only a change of environment, but also physical motion and rapid shifts of contextual features that make both the interaction with the system more complex for the user and the data collection process harder for the analyst. Among the empirical studies devoted to the use of mobile systems, there is a well-established distinction between two main approaches (laboratory studies and field studies), that are based on different epistemological assumptions (see Kjeldskov, Skov, Als et al., 2004).

The first approach consists of studying the interaction in a laboratory. This focuses on controlling the variables which can affect the interaction, and allows the construction of artificial settings to put the user in a mobility situation. For example, in some studies, users are asked to walk on a running machine (Beck, Christiansen, Kjeldskov et al., 2003; Pirhonen, Brewster, & Holguin, 2002). In general, with this experimental approach, the assumption is that it is possible to recreate in a laboratory a “natural” situation of user mobility. From this point of view, mobility is primarily considered under its physical and awareness aspects. Such reductionism must address the traditional problem of the ecological validity of the collected data. To an actor, the act of movement implies more or less important context changes, not only from a physical point of view (lights, sounds, objects, etc.) but also from a socio-cultural one (places, people, situations, etc.). Even if they do not have an automatic effect on the activity, these variations are potentially relevant and may have unpredictable effects on the users’ course of action and consequently on the usage of the device. Moreover, the human-machine interaction itself generates its own context of achievement by a moment by moment dynamical management, focused on the achievement of a given goal (e.g. a search for information, writing messages, etc.)

To address the limitations of laboratory experiments when evaluating mobile devices a second approach has been explored that focuses on naturalistic studies. With this approach, the natural character of mobility, and consequently its context, is preserved as much as possible and what is lost in the control of the variables is gained in terms of ecological validity.

Some recent research has questioned the added value of naturalistic studies in the identification of usability problems compared to laboratory experiments, and has managed to address this issue in a systematic way (Kjeldskov et al., 2004). Some authors have found a trade-off between the two approaches, using the notions of “field experiments” (Goodman, Brewster, & Gray, 2004) and “quasi-experiments” (Roto, Oulasvirta, Haikarainen et al., 2004). They attempt to keep a strict experimental control of the factors whilst at the same time preserving the realism of the usage. Their approach is essentially quantitative and aims at explaining the causal relationships between factors. This follows the orthodox tradition of experimental psychology applied to HCI (testing hypotheses, controlling variables, and performing statistical tests).

However, one of the main challenges of the naturalistic approach is to define methods and equipment that enables to study empirically the use and the usability of mobile technologies in the wild. The purpose of this paper is to describe the approach that we adopted in a study held during summer 2004 at La Cité des Sciences of Paris-La Villette, which aimed to evaluate the usage of a mobile multimodal device¹. A PDA with a multimodal interface and a 802.11 wireless network card was used in a collaborative context where two users had to play a kind of quiz game (the goal was to discover clues in order to answer to a set of questions connected with different thematic spaces of the museum).

2 General principles

Our objective is to perform empirical studies of interaction with mobile multimodal devices, in natural settings, with an emphasis on the observation of real practices and on the analysis of meaningful action. Different relevant features of the situation have to be recorded: the user’s actions and reactions, the interaction with the multimodal device, and the relevant aspects of the surrounding environment. Apart from the physical effects of motion on the activity, one of the main consequences of mobility is to induce rapid shifts in the context, that may have direct effects on the usage of a device. Thus, it is necessary to collect data related to the evolution of context, be it physical or social. However, all of the different contextual factors cannot be defined a priori, especially when users operate in an open environment. The collected data must therefore enable the experimenter to perform the analysis afterwards, based on a rich recording of activities.

This methodological orientation requires the equipment to be portable and to be sufficiently autonomous to enable the experimenter to follow the user in the different environments. In addition, the equipment should be as unobtrusive as possible (physically, psychologically, socially) for the user. A possible solution is to follow the user as closely as possible with a video recorder. Unfortunately, this method has many shortcomings. First, it may induce a feeling of discomfort, as it is not always socially pleasant to be followed and video-recorded in a public space. A second problem comes from the difficulties associated with recording simultaneously the user’s actions and the multimodal device; the user can rapidly shift from one input modality to another, and in the same time it is necessary to keep a trace of the output modality. Therefore, the

¹ This study was conducted in the context of the LUTIN project (<http://www.lutin.utc.fr>)

experimenter would need to switch continuously from the user to the display, selecting in real-time what is the meaningful information at a given moment. A third problem comes from the fact that it is not possible to define a priori the important factors in the dynamic surrounding context that can explain the user's actions. The experimenter has to make a choice between focusing on the manipulation of the device and losing the contextual aspects, or taking a broader view, which for instance could include interactions with other people (vom Lehn, Heath, & Hindmarsh, 2001), and missing the fine-grained sequences of actions on the device.

This particular issue exemplifies from a technical point of view the difference between two general traditions of observation: in one hand an approach drawing inspiration from sociology and ethnoscience which focuses on interactions with other actors in the same physical environment; and in another one the classical HCI approach, which mainly focuses on the interaction between a single user and a device.

3 Description of the recording set up

These different constraints led us to design a mobile recording unit which was physically unobtrusive and discreet and that enabled us to collect data on the user's actions and interaction with the device, and the context. This equipment includes three modules (figure 1):

- A micro-camera fixed on spectacles providing a so-called "subjective" view since it captures part of user's visual field. This view also captures a part of the interaction with the device and provides some information on the current context.
- A contextual view captured by an experimenter which accounts for the user's global context.
- A automatic logging tool, which captures all of the system's events and thus avoids a tiresome manual coding.

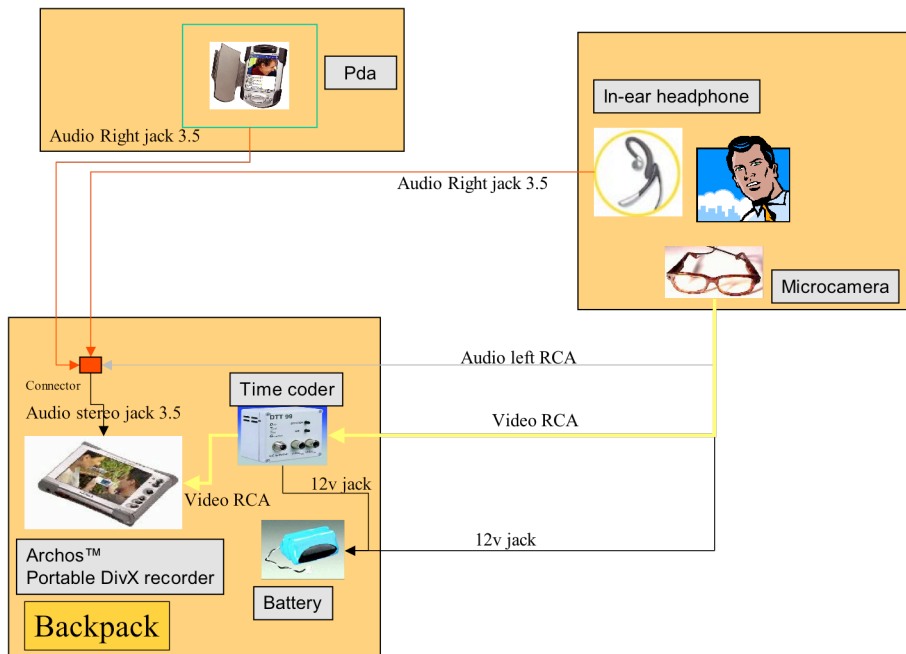


Figure 1. Mobile recording set up.

These data are then enriched with the verbalisations of the users from the self-confrontation interviews (see section 3.4.); this arrangement of data allows a fine-grained analysis of the user's activity.

3.1 Contextual view

In order to identify afterwards the contextual features that make the user's actions meaningful to the analyst, it is crucial to capture the contextually relevant information from the user's point of view. This can be achieved by using cameras distributed in the environment at judicious locations (when the user moves in a restricted or known environment) or by using a video camera held by an experimenter.

A limitation with the fixed cameras is their lack of flexibility: there is one single viewpoint on the situation and it cannot be dynamically changed. On the other hand an experimenter following the user with a video camera can adjust in real-time the centring in order to have the more relevant viewpoint at a given time. A drawback with this approach is that different experimenters will not behave uniformly when following and recording the users. Ideally, it would be better to have the same experimenter recording all of the experimental sessions throughout the study.



Figure 2. Unobtrusively shadowing the user : the experimenter follows the user with a camcorder.

We selected the option of having an experimenter who follows the user at a distance in order to have a wide viewpoint of the situation. Contrary to other approaches (for example, Kjeldskov & Stage, 2004) the presence of the experimenter remains discreet and unobtrusive since there is no need to record the user's interaction with the device as these data are available through the subjective view and the automatic logging application (see section 3.3).

3.2 Subjective view

The capture of contextual information in the surrounding environment of the user is not sufficient, since it does not provide access to the user's viewpoint, noticeably when he/she pays attention to artefacts or manipulates them. Therefore we used a wearable micro-camera attached to spectacles (Calvet, Kahn, Salembier et al., 2003) that enabled the collection of data about the immediate environment in the user's visual field, the interaction with the multimodal device, and more

especially the manipulation of this device. This information completes the data collected via the automatic logger, and facilitates the analysis of the raw data in the log files.



Figure 3. User's "subjective" viewpoint.

The micro-camera contains a microphone that provides sound information about the surrounding environment, specifically: the vocal interaction with the speech recognition module of the multimodal application, the interaction with other agents, and the ambient environment in which the activity takes place. The audio-video data are directly recorded and compressed in DivX format in a blow on a portable hard-disk (ArchosTM). The user carries the hard disk and a set of batteries in a backpack. In a variation of this basic configuration, the audio-video stream is directly transmitted via a wireless emitter to a remote receiver (figure 4).



Figure 4. The wireless video-recording system (which actually still requires some wires between the handset and the wireless emitter...).

This set-up was actually not used during the LUTIN experiments because of the ubiquitous metallic structures in the environment that generates interferences with the wireless system.

3.3 Logger

The subjective view is not always sufficient to capture all the events that happen on the display of the device. This may be because of mundane causes such as light reflections on the screen or because the user is looking at something or someone in the environment and subsequently the device is out of the scope of the camera. Moreover, coding the data related to the interaction with the device from the video-tapes is difficult and time-consuming. To cope with those limitations a logging application has been designed that automatically records the traces of the interaction with the device on a server in a nearby room over a 802.11 wireless local area network (Merlin, 2004).

The system also includes filtering and export functions that facilitate further data processing, facilities for selecting and formatting data, and functions for combining logged data with the audio and video data.

3.4 Replay and self-confrontation sessions

The contextual and subjective views provide an audiovisual support in order to conduct self-confrontation interviews with the users (figure 2).



Figure 5. A self-confrontation session

The general idea behind the confrontation approach is to provide a subject or a group of subjects with traces of an activity (e.g. writings, schemas, annotations, data from the automatic logging of actions on a computer system, and more usually audio and video recordings) in order to collect verbal comments. In the same time external traces enable the analyst to match the verbal report with the traces of activity. The so-called self-confrontation term applies when a subject is shown a recording of his/her own activity in order to put him/her in the context of a past experience.

In the LUTIN experiment, the digital videotape was shown to the subjects on a desktop computer running a customized version of the ActogramTM data analysis software² (Kerguelen, 1998) which features video editing facilities.

4 Analysis

The audio-video data were coded by hand on the ActogramTM tool which enables the simultaneous viewing of multiple synchronized videotapes. The software provides a dedicated window that supports on-the-fly coding of information based on the categories of contextual information previously identified to perform the analysis (figure 6).

Different taxonomies of contextual information have been used in recent works on usability testing for mobile and handheld devices (see for example (Hoyoung, Jinwoo, Yeonsoo et al., 2002; Roto et al., 2004). Most of them are essentially empirical and are not grounded in a core of theoretically defined concepts.

² <http://www.actogram.net/>



Figure 6. The ActogramTM data analysis tool.

The different events (including the users' actions) were coded in three categories in order to describe the contextual dimensions of the situated activities. We draw inspiration here from the distinction introduced by Quéré between “environment”, “context” and “situation” (Quéré, 1997). In our study the “environment” refers to the surrounding physical properties (light, noise, physical arrangement of the space,) that may play a role in the selection and use of the available interactive modalities. These properties are identified and coded by the experimenter when repeatedly viewing the videotapes (contextual view). The “context” refers to the features that can give sense to a particular action: the type of activity that the user is engaged in at a given moment (e.g. interaction with the device, information acquisition, motion management, etc.), the historical dynamics of the user's course of action (e.g. incidental sequence, interaction episode with the other player, interaction with other actors, etc;). The viewing of the tapes, combined with information from the log files, is used to code these contextual elements. The “situation” addresses directly the experience of the user and subsequently the dynamics of his/her commitment in action at a given moment. This level of analysis, which requires a description of the activity from the user's point of view, is based on the self-confrontation reports.

5 Conclusion

A big challenge for empirical studies of mobile activities is to go beyond the punctual usability evaluation of a technological device conducted during a limited period of time. Even though some usability problems can be quickly identified in a one hour experiment, a real understanding of the usage of a technological device and how it reshapes an activity requires covering a wide range of user experiences in real world situations. However, in order to address this appropriation process i.e. the process through which the artefact is adopted and transformed (Carroll, Howard, Vetere et al., 2001), it is necessary to conduct longitudinal studies. Therefore, as it becomes impossible in practice to track and observe users continuously during long periods of time, we need to extend the scope of methods that we use to study mobile activities. Methods such as diary reports (Palen & Salzman, 2002) and interviews supported by the traces of the user's activity (provided for example by automatic logging) are natural candidates. Nevertheless it is still important to combine these approaches with the direct observation of the real activities of users.

6 Acknowledgments

This work was funded by RNRT through the LUTIN project (usability lab at the Cité des Sciences) and France Telecom R&D through the MOTIVE project (multimodal interaction and mobility). Thanks to Julie Dugdale for helpful comments.

7 References

- Beck, E., Christiansen, M., Kjeldskov, J., Kolbe, N., & Stage, J. (2003). Experimental evaluation of techniques for usability testing of mobile systems in a laboratory setting. In *Proceedings of ozCHI 2003*. Brisbane, Australia.
- Calvet, G., Kahn, J., Salembier, P., & Zouinar, M. (2003). "In the pocket": An empirical study of multimodal devices for mobile activities. In J. A. Jacko & C. Stephanidis (Eds.), *Human-Computer Interaction - Proceedings of the HCI International Conference, June 22-27, Creta*: Lawrence Erlbaum Associates.
- Carroll, J., Howard, S., Vetere, F., Peck, J., & Murphy, J. (2001). *Identity, power and fragmentation in cyberspace: technology appropriation by young people* (Working paper No. 01/IDG/2001): Department of Information Systems, University of Melbourne.
- Goodman, J., Brewster, S., & Gray, P. (2004). *Using field experiments to evaluate mobile guides*.
- Hoyoung, K., Jinwoo, K., Yeonsoo, L., Minhee, C., & Youngwan, C. (2002). An empirical study of use contexts and usability problems in mobile internet. In *HICSS, Proceedings of the 35th annual Hawai International Conference on System Sciences*.
- Kerguelen, A. (1998). Quels outils concevoir pour aider au relevé d'observation sur le terrain? In *Actes du XXXIIIe Congrès de la SELF, "Temps et Travail"*. Paris.
- Kjeldskov, J., Skov, M., Als, B., & Hoegh, R. (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, Glasgow, Scotland* (pp. 61-73): Lecture Notes in Computer Science, Berlin, Springer-Verlag.
- Kjeldskov, J., & Stage, J. (2004). New techniques for usability evaluation of mobile systems. *Int. J. of Human-Computer Studies*(60), 599-620.
- Merlin, B. (2004). *Plateforme générique d'observation des interactions* (Rapport Technique IRIT-DIAMANT). Toulouse: IRIT.
- Palen, L., & Salzman, M. (2002). Welcome to the Wireless World: Problems Using & Understanding Mobile Telephony. In R. Harper & B. Brown (Eds.), *The Wireless World*: Springer-Verlag Publishers.
- Pirhonen, A., Brewster, S. A., & Holguin, C. (2002). Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of CHI'2002* (pp. 291-298). Minneapolis, MN: ACM Press Addison-Wesley.
- Quéré, L. (1997). La situation toujours négligée? *Réseaux*(85).
- Roto, V., Oulasvirta, A., Haikarainen, T., Kuorelahti, J., Lehmuskallio, H., & Nyysönen, T. (2004). *Examining mobile phone use in the wild with quasi-experimentation* (HIIT Technical Report No. 2004-1). Helsinki: HIIT.
- vom Lehn, D., Heath, C., & Hindmarsh, J. (2001). Exhibiting interaction: conduct and collaboration in museums and galleries. *Symbolic Interaction*, 24(2), 189-216.