ANALYSING MOBILE AND VOICE SERVICES USAGES IN CONTEXT

RACHEL DEMUMIEUX, RAPHAELLE HAUTIN
France Telecom Research and Development. 2 avenue Pierre Marzin. 22300 Lannion, France.
{rachel.demumieux,raphaelle.hautin}@orange-ftgroup.com

Abstract. This paper presents two techniques for evaluating usability and usages of cell phones and interactive voice services. Some tools have been developed to collect user interactions in context. The purpose of these techniques is to carry out remote field studies and observing the interactions over a long period without disturbing the customers. The paper first addresses the description of two tools and the protocols. Then, we illustrate some results.

1. Introduction

Cell phones and voice services are becoming increasingly popular and used by millions of people all over the world. They have changed the way we communicate with friends, family and machines (text, voice or tactile mode inputs). Observation in urban environment is fundamental to better understand usages and usability issues encountered by customers. But mobility makes it difficult to carry out field studies while controlling conditions and environment.

We illustrate some techniques and tools adopted for evaluations in laboratory and natural settings in order to highlight our practice.

Laboratory tests are performed in order to observe usability issues with the mobile and voice services. These methods (i.e. user tests, interviews, expert reviews) enable us to identify the main usability problems (Nielsen, 1993). But they are limited to understand the usage and system impacts in various environments like home, street, work, etc. (Kellar, 2005).

An approach consists in creating artificial mobility situations (Beck, 2003; Kjeldskov, 2004). They conducted some experiments based on different configurations (i.e. sitting on a chair, walking on a treadmill, walking in a pedestrian street…). The main limit noted by authors is the number of users necessary in each experimental condition to compare the observed performances.

Another approach observes the usages in urban environment with equipment such as cameras (Roto, 2004; Kaikonnen, 2005) enabling to record user interactions. This technique is based on the best possible control over nuisance variables in the mobile test environment (multiple trips with different conditions). The main limits are technical (quality of microphones, weight of cameras, participants interrupted by people during the experimentations and time spent to code data).

Another method consists in employing a "simple" tool: the diary. The users note their activities (Ozkan, 2002) by using a system. The main weakness is to add a task for the users. And the obtained data are only declarative.

In response to these limits, our challenge was to conduct some tools allowing to track day-to-day user interactions without constraints for them and to conducting remote tests over period of several weeks even months.

This paper describes our experiences conducting such research in two kinds of applications. We illustrate several tools and the methods. Finally, a few benefits and drawbacks are detailed.

2. Context

For the evaluations of the mobile and voice services in natural settings, we perform some field studies. Our goals are described as follows:

- Complete traffic data (useful for the assessment of the systems performance, for the marketing department and for the billing) by adding usages and usability data.
- Apply a User Centric Design process (UCD) by performing some field studies at the different steps.
- Assess the impact of the systems in natural settings.
- Provide some recommendations for the designers.
- Use the results to improve current services and to specify new services.

3. Tool for the cell phones

Cell phones with open Operating System (OS) are sold (Windows Mobile or Symbian) for a few years. Most of them are Smartphones which provide some features such as email access, video player, Bluetooth connection, etc. Furthermore, some "extra" applications can be created thanks to open source platforms and tools provided by manufacturers. These open systems gave us the opportunity to design a tool collecting user interactions in mobility.

3.1 DESCRIPTION

The software called "ACIDU" which stands for Usage Data Collect Application (Demumieux and Losquin, 2005) is embedded on Windows Mobile handled devices (i.e. SPV C500, C100, E650…). In a nutshell, the tool based on automatic loggings allows the phone activity tracing.

The download is achieved by a SMS which encloses an URL. A dedicated menu is added on the mobile with some options enabling to stop or delete the application. The log files are saved and sent automatically to a server by GSM or GPRS. For security considerations, no private data like conversations or message contents are collected.

The application collects time stamped and heterogeneous data: frequency and duration of the application used (directory, messages, agenda, etc.), phone events (number of incoming and outcoming calls, SMS…), procedures required to achieve tasks (menus and keys pressed), Cell-id changes (location information which is implemented but not yet evaluate).

3.2 STUDY

In this section, we describe the method adopted (see below). We performed a field study with 110 customers using the SPV C500. They participated over a period from one to four months by using their phone as usual.

TABLE 1. The protocol.

Steps	Actions
Defining representative criteria required for the recruitment	We made a statistic study based on traffic data of the SPV C500 subscribers
Drawing up documents for the participants	We wrote the study modalities, the learning guide to install application, the mails, etc.
Pre testing the tool and the protocol	We requested 10 internal participants to evaluate the protocol
Recruiting the participants	We contacted 1000 SPV C500 customers (by phone). 180 customers accepted the test
Obtaining the participants agreement	They returned the modalities document signed by mail or email
Processing the field study	They downloaded ACIDU et we gathered data during four months.
Getting declarative data	We performed two surveys by phone: focused on customers'

	profile, current habits, opinion (110 participants answered)
Closing the study	We rewarded with some coupons the participants and they deleted ACIDU
Analyzing the results	We sorted quantitative and qualitative data and we compared the gap between declarative data (surveys) and objective data

3.3 ILLUSTRATION OF SOME RESULTS

In this paragraph, we emphasize two examples of the objective results. The first one illustrates the global usages and the second one the favorite procedure to access to the directory.

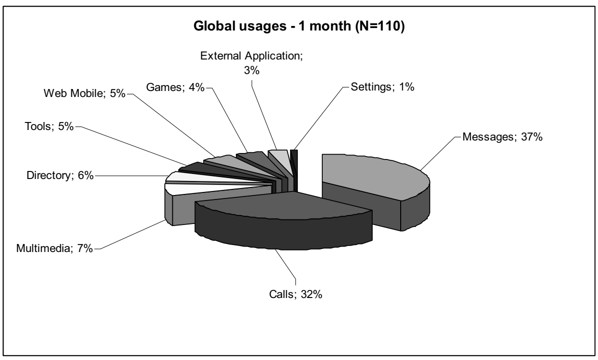

Figure 1. The used features over one month period.

The results show that the main features used are messages (37%) and calls (32%). The other features represent only 30% of the usages despite the wide functions offered by this mobile.

Three procedures enable to access "contacts" menu; either selecting "start" menu or from the Orange homescreen via the icon "contacts" or from the right softkey "contacts".

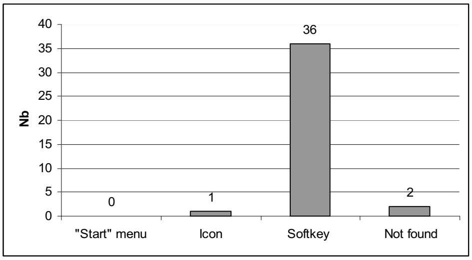

Figure 2. The procedures to access to the contact menu.

The results show that most of the participants having Orange Homescreen (N=39) had one favorite procedure which is quicker to access to the directory in selecting the softkey "contacts".

4. Tools for the voice services

In this part, we present the voice services using the natural language technology. The customer speaks with his own words; he is not constrained by few choices in a tree structure. The system interprets the user's query, and directs him to the operator which deals with his request. The design of these systems is based on the development of different elements such as: the speech recognition, the semantic interpretation of the requests, and the man-machine dialog. One must pay attention to each of these parts in order to reach an optimal user experience. The technological performances are necessary but not sufficient to ensure the quality of a voice service, it also depends on the dialog between the user and the machine.

4.1 DESCRIPTION

Several tools are designed to improve the man-machine dialog. The first one is used to record the whole call: the system messages and the user's answers. It also collects side data like date and duration of the call, the dialog phases followed by the user, the output text of the recognition system, etc. The audio recording is useful to explain some failures, for instance due to a disturbing noise or to another person talking in the room. A second tool is used to store, listen, transcribe, and analyze these recorded calls.

4.2 CALL ANALYSIS METHOD

The main stages of dialog usability studies are detailed below:

- Define the objectives of the study.
- Pre test the tools: collect, import and listen to few calls.
- Propose a first set of analysis criteria.
- Choose a sample of recorded calls (hundreds).
- Validate the analysis grid (see table 2) on a subset of the sample. It may lead to the addition of a new criterion, or new options for an existing one, etc.
- Write a document describing precisely each criterion, and each option, to avoid evaluation mistake. Rules may also be defined to properly fill in the grid.
- Listen to the sample of calls, and fill in the grid for each of them.
- Obtain statistics.
- Propose recommendations to improve the system. The obvious goal is to increase the quality of service and the users' acceptance. For that, we also perform some surveys with the customers.

TABLE 2. Example of analysis grid.

Number of expression of the reason for the call: ⊙ 1 expression ○ 2 ○ 3 ○ Over 3 ○ None	**Number of understanding of the reason for the call**: ⊙ 1 good understanding ○ 2 ○ 3 ○ Over 3 ○ None	**System global understanding:** ⊙ OK ○ NOK ○ Not applicable ○ Don't know
System request for more information: ⊙ Yes ○ No	**Hang up:** ⊙ No ○ At welcome message ○ At request for the reason for the call ○ After the request for the reason for the call ○ Upon system invitation	**DTMF guidance:** ○ Yes ⊙ No
Perfect call: ⊙ Yes ○ No	**Comment:**	

4.3 SOME RESULTS FOLLOWING THIS METHOD

In this section, we highlight two examples of results. The first one describes global uses of a helpline service (speech and natural dialog solution). The second one presents a comparison between different dialog strategies.

The efficiency and usability of this service are regularly assessed to improve the dialog. Currently, the understanding success rate (speech recognition + semantic interpretation) of what customers are saying, reaches 89 %. The results obtained from the grid detailed above are as follows:

- In a sample set of calls (893) in which users spoke in a comprehensible way, the efficiency rate of vocal service was 97%.
- 68 % of the dialogs were considered “ideal”. "Ideal" is defined here as a customer-system dialogs that followed a “perfect” course: no “off-topic” query, no silence from the user, no misunderstanding of the system, etc.
- The hang-up rate in the middle of a call is estimated at 6.5 %.

This analysis is focused on one specific step of the dialog. The user gives an input, the system expresses its understanding and the user can rectify if necessary. We compare two kinds of dialogs to evaluate the impact of messages on the customers' behavior. The first message only gives a feedback based on the user's answer (dialog #1). The second one provides the same feedback and in addition a message to correct a potential system error (dialog #2). We compare the users' behaviors (1000 calls have been interpreted for each dialog).

TABLE 3. Distribution of reactions during an incorrect feedback.

User's behavior upon incorrect feedback	Dialog #1	Dialog #2
He says nothing	225	95
He cancels/corrects	**14**	**93**
He repeats	13	0
He rewords	5	1
He validates	0	1
Total	**257**	**190**

The results show that the users tend to correct the mistake when they are invited to by the system.

5. Interests and limits

According to our practice, the benefits of these methods and tools are as follows:

- These techniques monitor users' activities remotely over a period of several months.

- These works enable to enhance the knowledge of the main usability issues in natural settings. Furthermore, the features utilities are also evaluated.
- These tools do not disturb the users' task.
- All the features of the services can be analyzed (as opposed to laboratory conditions where a few predefined tasks had to be selected).
- The tools collect a lot of users' expressions and navigation strategies not expected by the project team.
- These methods that gather objective and declarative data allow to measure the gap between both ("real activity"/users' perception).

Nevertheless, some drawbacks are also observed such as:

- These studies require a long time for the tools development, the protocol, the running of the experiments and the analysis.
- Some data are not understandable without obtaining some additional explanations and opinions provided by the users.
- For ACIDU, the recruitment of the participants was long and difficult.
- ACIDU is currently available on Windows Mobile phones which are rather targeted for the techno-addicts. Thus the results can not be generalized for the mass-market.

To conclude, further research and experiments are considered in order to improve our tools and methods (i.e. requesting the users' opinions closer to the execution of the task).

References

Beck, E. and al.: 2003, Experimental evaluation of techniques for usability testing of mobile systems in laboratory settings, *Proceedings of OzCHI2003*, Brisbane.

Demumieux, R. and Losquin, P.: 2005, Gather customer's real usage on mobile phone, *Proceedings of MobileHCI'05*, Salzbourg, pp. 267-270.

Kaikonnen, A and al.: 2005, Usability testing of mobile applications: a comparison between laboratory and field testing, *Journal of Usability Studies*, **1**(1), 4-16.

Kellar, M. et. al.: 2005, It's a jungle out there: practical Considerations for evaluation in the city, *in HCI 2005,* Portland, pp. 1533-1536.

Kjeldskov, J. and Stage, J.: 2004, New techniques for usability evaluation of mobile systems, *International Journal of Human-Computer Studies,* **60**.

Nielsen, J.: 1994, *Usability Engineering, in* Morgan (eds), San Francisco.

Ozkan, N.: 2002, Exploring the use of mobile phone, *in Pproceedings of IFIP World Computer Congress,* pp. 89-102.

Roto, V. Oulasvirta, A. Haikarainen, T.: 2004, Examining mobile phone use in the wild with quasi-experimentation, *in* HIIT Technical Report, 2004-1.