

“In the pocket” : An empirical study of multimodal devices for mobile activities

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

*GRIC-IRIT, Université Paul Sabatier, 118, rte de Narbonne, 31062 Toulouse Cedex 7

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

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

Abstract

This paper presents some methodological issues and some of the first findings that have arisen from our work on the « Multimodal mobile devices » joint project at GRIC-IRIT, CLIPS-IMAG, and FT R&D department. Our main research and design interest is in the development of novel interactive devices for the mobile user that exploit classical (i.e. existing) and innovative input modalities. Two different studies are briefly presented and some results are commented. The first experimental study aims at providing some empirical results in order to discuss the functional roles play by multimodal interfaces in the context of a real task. The second study explores the role of coordinating artefacts in mobile cooperation, and addresses the problem of contextual constraints which may restrict the use of some modalities.

1 Introduction

Multimodal interfaces combine many simultaneous input modalities and may present information using different output modalities. The input modalities can be quite simple (for example two different pointing devices) ; but they can also include more sophisticated techniques such as speech and gestures recognition.

The growing interest in multimodal interface design is inspired largely by the goal of supporting more flexible, natural and robust performance in the context of real-world situations (ever changing factors, continuous shifting from acceptable to less than perfect conditions, social context of use,...). Unfortunately so far few studies have given an account of the way multimodal systems are actually used by people in the context of real tasks. Moreover most of the previous studies have considered a limited number of modalities.

In the same time, rapid and major technological advances lead to the use of many electronic devices (laptops, mobile phones, PDA,...) that allow users to interact with people and information in different geographical locations (Harper & Brown, 2002). Mobile usage of PDAs offers even more challenges, as not only do the problems arising from the diminishing size of the PDA screens but also the completely different user environments. The use of mobile devices gives opportunities to act in contextually enriched environments, but natural field environments also are challenging because of the large number of contextual sources of disturbance ; for example physical and social constraints imposed on the interaction with the device may vary rapidly. Multimodality can be seen as a candidate to help the user dealing with this variation (Oviatt, 2000); but so far few empirical data have been produced to explore the theme of whether multimodal interfaces can reliably support adverse usage environments.

2 Exploring multimodal interaction

This first study was focused on the assesment of a device integrating different input modalities such as gesture, speech and physical manipulation of the device itself (“embedded physicality”).

The main objectives were :

- To tackle the problem of multimodality in “meaningful” situations of interaction, i.e. contexts of tasks/activities the users are familiar with.
- To study the processes of appropriation of the multimodality along several sessions of interaction.
- To explore the use of innovative modalities in order to empirically check out the role they may play in the man-machine interaction, the challenges they pose individually (monomodality) and in relation to the use of other modalities (multimodality).

More precisely this study aimed at addressing some specific issues: how is multimodality carried out and what are the factors that determine the choice of modality ? Is there any shift from a modality to another ? Do the properties of modalities play any role in this choice? What are the criteria that lead a user to adopt or give up a modality of interaction ? How do the users combine the different modalities ?

2.1 Method

2.1.1 *Setting*

The strategy adopted for addressing those problems was first to use empirical Wizard-of-Oz (WOZ) simulation to test a partially simulated multimodal system¹. The purpose of this WoZ experiment was to enable the users to interact in a natural way while limiting the cost of development of the system and to avoid the current technological limitations at this exploratory step of the study of innovative modalities (gestual and tangible modalities).

2.1.2 *Subjects*

10 subjects (regular users of Internet but not computer specialists) took part in this experimental phase. In order to put the subjects in a familiar situation, they were asked to use an e-mail application implemented on a pocket PC which was connected to their personal electronic mailbox.

2.1.3 *Procedure*

At the beginning of each session, the subjects were briefly introduced to the task to be performed and to the setup including the messenger software, the input and output modalities. The actions performed by the subjects during the sessions (from 4 to 10 sessions of 15 minutes average duration each) were recorded by means of various sensors (video cameras, system events recorder). The experimental sessions were immediately followed with a self-confrontation session (interview of the user based on the video recording of his/her activity).

¹ WoZ setup has been used in other empirical studies of multimodality (see for example Bernsen & Dybkjær, 2001)

2.1.4 Data

Different data were collected. During the interaction, the commands and the windows manipulated by the subjects were indexed temporally and automatically stored in a file. The actions of the subjects, the modalities used, the duration and the contents of the commands were collected a posteriori from the analysis of the video recordings of the experimental sessions. All those data were encoded using the Actogram software² which offers quantitative analysis and graphical display functions.

Recording of the self-confrontation sessions was carried out in order to allow qualitative analysis focused on the identification of the strategies applied by the subjects when using the modalities.

2.2 The modalities : between innovation and classicism

Four modalities were implemented on the PDA (see figure 1) :

- the vocal modality (unconstrained natural language interaction).
- the tactile modality, (with a styllet to point the command buttons)
- the gestural modality, (which relies on a set of predefined gestures)
- the “tangible” modality (which associates commands with predefined handling modes of the artifact : changes of orientation, various movements, etc.)



Figure 1 : The tactile, gestural and tangible modalities

Only the tactile modality was really implemented ; the other modalities were simulated by the WoZ setup.

2.3 Results

2.3.1 Global use

One of the first results obtained about the use of the modalities, indicates that all the modalities were used by the subjects during the different sessions. In addition, one can observe an evolution in the use of multimodality during sessions:

- During the first sessions there was a homogeneous distribution in the use of the modalities which corresponds to a phase of an informal test of the system. This fact is corroborated by the data resulting from the self-confrontation sessions ;
- but quickly some preferences in the use of the modalities appear in the following sessions ;
- these preferences vary according to the subjects.

2.3.2 Intra-individual specialization

As mentioned above, many subjects exhibit a tendency to preferential uses of modalities. These tendencies may be local (they appear only during one session of interaction) or general (they

² <http://www.actogram.net/>

appear in other sessions), specialized (the same modality is used to carry out only one and only one command) or plural (the same modality is used to carry out several commands).

2.3.3 *Influencing factors in the choice of modality*

The activity graphs and the verbal reports put into evidence several factors which are likely to direct the choice and the change of use of the modalities. The four principal ones are listed below :

- Context of course of action (recurring operational procedure) :

We could identify that the tendencies to preferential use generally result in sequences of recurring actions done by the subjects. These sequences are strongly related to the context of the course of action.

- Change of goal :

The changes of modality are often associated with phases of local reorientation of the activity.

- Properties of the particular implementation of the modalities :

During the evaluation of a multimodal device, it is necessary to distinguish the inherent characteristics of a given modality from the particular implementation of this modality. The fact that a modality is not used does not mean that it is unsuited but that its technical realization can be inadequate or constraining.

- Error recovering

Many shift between modalities happens in situations perceived by the user as dysfunctioning ones ; the shift from a modality to another is then seen as a possible resource to recover from errors.

3 **Cooperative mobile activities in real-world settings**

The second study we briefly present here is a first attempt to identify the contextual constraints users may have to face when using a mobile device for coordination purpose in a real world situation.

Two subjects (teachers) were asked to explore the two levels of a museum in order to prepare a visit with young children. They had to select a set of objects and places of interest in respect with pedagogical issues and to define a convenient pathway for the visit.

As the subjects were not located at the same floor, they were given a simple communication device³ (a multimedia mobile phone with which they could send each other vocal messages, short-text messages, and pictures) in order to enable them to coordinate their joint activities. The task had to be fulfilled in real-time as the subjects were told there was no opportunity to exchange information outside of the museum. Each subject was briefly introduced to the scenario and to the MMS mobile phone. They were not encouraged to follow any particular strategy, such as to coordinate step-by-step while moving in the rooms or to wait until the end of the parallel exploration of the museum.

The activity of each subject was video-recorded all along the session, and used as a support for verbal reports gathering. In addition, one of the user was equipped with a wearable microcamera embedded in her spectacles to give a subjective view of the interaction with the environment and the device (see figure 2). Despite some identified limitations (Mark, Christensen, & Shafae, 2001), the microcamera technology satisfies many requirements in terms of capture of the mobile device usage from the viewpoint of the users and broader aspects of the surrounding environment.

³ For a presentation of a more sophisticated scenario of electronic tourist guide see for example the HIPS-I³ EC project



Figure 2 : Subjective view of the use of the MMS device

The collected data outline the different forms of interaction performed by the users with the mobile device and provide information on the acceptable practice in social environments. They give an account of the strategies the subjects use when facing dynamic contextual constraints.

4 Conclusion and proposals for new research

In this article we presented two exploratory studies focused on the use of a portable multimodal system in a experimental setting and on the use of a mobile device for coordination purpose in a real world environment. The results of the first study generalize earlier findings : effective use of multimodality (all the modalities were used), interindividual differences (existence of preferential tendencies), change of modalities in the dysfunctioning situations. Beyond this convergence, an original result of this study is that the subjects get used to nonusual modalities rather easily ("tangible" modality for example).

This kind of device is likely to be used in variable contexts. Thus difficulties may arise when the modalities of interaction will not be equally adapted. The multimodality takes a crucial importance then as it offers to the users a ressource to cope with the contextual variations of the environment (physical, social, etc...).

In the next step of this work we intend to link the multimodal and mobile dimensions in real world settings observation. We will made available to the users a multimodal device they will be able to use in different contexts. The purpose of this step is to situate our inquiry of innovative input modalities in the daily activities of the users over a longer period of time.

References

- Bernsen, N. O., & Dybkjær, L. (2001). Exploring Natural Interaction in the Car. In *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue* (pp. 75-79). Verona.
- Harper, R., & Brown, B. (2002). *The Wireless World*: Springer-Verlag Publishers.
- Mark, G., Christensen, U., & Shafae, M. (2001). *A Methodology Using a Microcamera for Studying Mobile IT Usage and Person Mobility*. Paper presented at the Workshop on Mobile communication, CHI'2001.
- Oviatt, S. L. (2000). Multimodal System Processing in Mobile Environments. In *Proceedings of the UIST '00 conference*. San Diego, CA: ACM.