How inconsistencies between speech and graphical user interfaces may be detrimental… or desirable

Valérie Botherel

France Telecom Research and Development
Department of Services Enrichment for a Simple Access to Information
2, rue Pierre Marzin, 22300 Lannion, France
valerie.botherel@francetelecom.com

Laurent Karsenty

IntuiLab
Prologue 1, La Pyrénéenne, 31672 Labège Cedex, France
karsenty@intuilab.com

Abstract

This paper presents the results of an empirical study of a multi-device service which is accessible both with a graphical user interface on a PC and a speech user interface over a telephone. The objective of this study was to deepen our understanding of inter-usability issues related to multi-device services. Inter-usability is defined as the easiness with which users transfer and adapt the knowledge they have acquired from previous uses of their service while they access it with a new device. The results of this study, obtained from a sample of 20 users achieving a series of scenarios, highlights the need to keep data, functions and use procedures as consistent as possible between the various devices but reveals, at the same time, inter-device inconsistencies that are not detrimental to the dialogue quality and even, in some cases, desirable.

Keywords: Applied systems for Human-Computer Interaction, Spoken dialog system, Multi-device service, Usability

1. Introduction

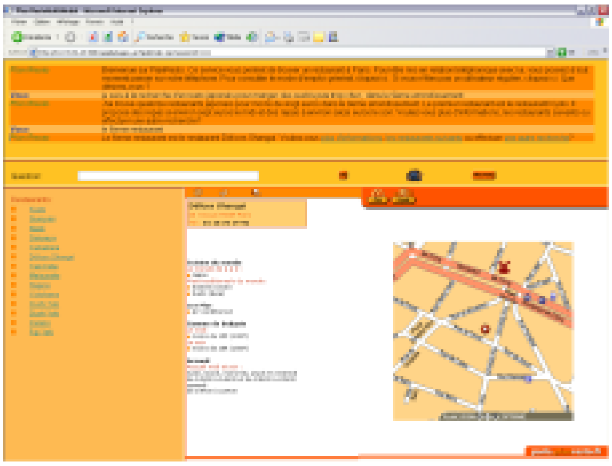
The concept of information ubiquity gives rise to a new generation of services accessible with fixed and mobile devices. Typically, each device has some physical, software and interactive characteristics (e.g., processing power, memory size, screen size, input devices, etc.) and, as a consequence, is associated with a specific user interface. Using a multi-device service implies, for this reason, an adaptation effort from all users as *inter-device inconsistencies* are encountered. It goes without saying that any multi-device service based on a speech user interface (SUI) on one side and a graphical user interface (GUI) on the other side inevitably presents inconsistencies. Information presentation and command input cannot be done in the same way, for instance. If the user's adaptation effort is too demanding, *interaction continuity* may be jeopardised. Interaction continuity rests on what we call inter-usability [1], or others, horizontal usability [4] to refer to the ease with which users can reuse their knowledge and skills for a given service when switching to other devices. To create inter-usable multi-device service, we need to better understand the circumstances in which users feel difficulties in order to define design rules that can help users to seamlessly switch from one device to another.

This is the goal of our research. This paper reports on an empirical study aimed at observing how users react to inconsistencies between two user interfaces allowing to access a new service, PlanResto. This service is developed by France Telecom Research and Development based on Artimis technology for intelligent dialog in natural language [3]. It allows searching for restaurants in Paris from 3 criteria: speciality, place and price. In this study, PlanResto is accessed from a PC with a GUI and from a telephone with a SUI. In both cases, a request can be made in natural language, typing it on the PC keyboard or speaking over the telephone (see fig. 1 for an example of spoken dialogue). With a PC, it is also possible to exploit the screen display and interact by direct manipulation (e.g., click on buttons or links as shown in fig 2).

Figure 1: Extract of a spoken dialogue with PlanResto.

User:	Find me an Italian restaurant please
System:	You want an Italian restaurant. There are more than 10 solutions. To continue, you may specify a price or location, consult the solutions or proceed with a new search.
User:	between 15 to 30 euros close to subway Sentier
System:	I found three Italian restaurants between fifteen and thirty euros close to Sentier. The first restaurant is the restaurant L'Uva Del Sole.
User:	the next restaurant please
System:	The second restaurant is the restaurant Il Buco. It offers menus around twenty two euros for lunch and menus around twenty two euros for dinner. Do you want more information, hear about the next restaurant or make a new search?
User:	more information
System:	The restaurant Il Buco is located 18 rue Léopold Bellan. Its phone number is <phone number>. Do you want more information, to be connected to this restaurant or make another search?
User:	more information
etc.	

Figure 2: Graphical User Interface of PlanResto.



2. Method

2.1. Test specific objectives

One of our goals was to investigate how expertise with a first device – called the referent device – may impact the discovery of the service with another device. We assumed that knowledge transfer would be reinforced in such a situation, facilitating interaction whenever the dialog context is consistent across both devices but possibly impeding it in case of inter-device inconsistencies. To study the effects of a referent device, we compared two conditions: in the first one, users had no referent device and accessed PlanResto by alternating both devices, the telephone with the PC; in the second condition, users used several consecutive times the same device before accessing PlanResto with the other device.

Actually, we anticipated that consistent interactions may be, in some contexts, less effective that inconsistent ones or, even, not desired by users. Grudin [2] already stressed the primacy of activity and context over consistency, which can lead users to prefer less consistent but more effective solutions. We also noticed in an earlier study [1] that some inter-device inconsistencies can be easily recovered by users or, sometimes, remain unnoticed without damageable consequence. This issue led us to distinguish different dialog context according to whether they are consistent or inconsistent across both devices.

2.2. Procedure

Twenty subjects were recruited and paid for this study, 9 women and 11 men, whose ages ranged from 28 to 61 years. They had varied occupations. All of them were web users, knew Paris, and had never used neither PlanResto, nor a SUI before. Each user achieved a series of 10 scenarios over 2 consecutive days. In this paper, we only report on the analysis of the 5 first scenarios. Each scenario prescribed request parameters which varied in terms of number and nature (2 or 3 specified parameters, adding or not the name of a restaurant).

Users were divided into 4 groups (see table 1). Groups 1 and 2 included users without expertise in a given device. They discovered the service alternating the SUI and the GUI. In Group 3, users acquired expertise with the GUI before discovering PlanResto over the telephone. In Group 4, the telephone was the referent device before users discovered PlanResto on the PC.

Table 1: Experiment protocol
(SC: scenario; SUI: speech user interface over the phone; GUI: graphical user interface of a PC)

Scenario		*SC1*	*SC2*	*SC3*	*SC4*	*SC5*
Alternating devices	Group 1	GUI	SUI	GUI	SUI	GUI
	Group 2	SUI	GUI	SUI	GUI	SUI
Referent device	Group 3	GUI	GUI	GUI	GUI	SUI
	Group 4	SUI	SUI	SUI	SUI	GUI

To analyse the effects of a referent device, 3 discovery situations were compared:

- Users discovered PlanResto with a first device. This situation corresponds to SC1 for groups 1 and 3 which used the GUI and groups 2 and 4 which used the SUI.
- Users discovered PlanResto with a new device after having used it with another device. This situation corresponds to SC2 for group 1 using the SUI and group 2 using the GUI.
- Users discovered PlanResto with a new device after 4 consecutive uses with another device. This situation corresponds to SC5 for group 3 with the SUI and group 4 with the GUI.

2.3. Data

A total of 100 dialogues comprising 1109 users' actions were collected. They were fully transcribed. Manual annotations were achieved for each user's utterance.

Specific annotations were made to capture the system's misunderstandings. A misunderstanding is not understood in this study as a word error but as a concept error, which corresponds to the difference between what the system decides is the right answer, and what the user actually says. Full misunderstandings were distinguished from partial misunderstandings according to the number of misunderstood concepts composing a user's utterance.

Another annotation category identifies the dialog context. Four dialogue contexts were distinguished, depending on whether they were consistent or inconsistent across both devices. Consistent contexts are:

- C1: request phrasing. Requests can be phrased exactly in the same way whatever the device is, except that they are typed with the PC and verbally uttered over the phone.

Inconsistent contexts are:

- C2: Choice selection within a list of actions proposed by the system. With the GUI, users may either type their choice or click on it (each action is displayed as an item with a hyperlink). With the SUI, users can only use the first possibility, designating their choice with voice.
- C3: Solutions browsing. With the GUI, all the solutions are simultaneously displayed and users can freely explore them. Conversely, with the SUI,

only one solution is presented at once and it's necessary to formulate a command - "next" or "previous" - to navigate through the solutions list.

- C4: Further restaurant information. With the GUI, simply clicking on the name of a restaurant allows users to visualise all of the available information about it. With the SUI, few information on the selected restaurant is given at the beginning; users must then ask "more information" to get further information.

To better understand how users may adapt – or fail to adapt – to inter-device inconsistencies, interaction modalities were also annotated. Five interaction modalities were distinguished according to the following criteria:

- *Click*: this coding only concerns the GUI when users click on a solution or command presented by the system.
- *Natural language*: users make a complete sentence with SUBJECT – VERB – OBJECT or composed of one or a few nominal groups linked with connectives like "in the" or "at", etc.
- *Key-words*: users use neither an article nor a connective between nominal groups, which may contain one or more single words.
- *Mixed*: users' utterances contain both natural language phrases and key-words.
- *Proposed choice:* when the system presents several possible actions, users may select one of them by repeating the word(s) used by the system.

Finally, we annotated "out of perimeter" requests. Indeed, some requests may concern data or functions that are accessible with the PC but not over the telephone (the opposite was not possible). Theoretically, these requests should reveal an inappropriate knowledge transfer from the GUI to the SUI.

3. Results

3.1. SUI and GUI inconsistencies causing problems

A total of 253 out of perimeter (OOP) requests were collected. A unique category of OOP comprises more than 100 of them: it consists in phrasing a request with a restaurant name, which the current version of PlanResto cannot process. All users of group 3, except one, made this mistake with the SUI during their fifth scenario, after having used the GUI 4 consecutives times. With the GUI, users can click on a restaurant name within the list of solutions to obtain detailed information about it. Therefore, we can assume that this user experience led them to believe that they can search for a restaurant by giving its name. This would mean than clicking on a name or saying it are two equivalent actions from the user's point of view. Unfortunately, PlanResto was inconsistent as regards this equivalence expected by users, which led them to produce so many OOP requests.

3.2. SUI and GUI inconsistencies causing no problem

Three dialog contexts were identified as inconsistent across the SUI and GUI, namely C2, C3 and C4. We noticed that misunderstanding rates are rather similar whether users have used the GUI 0 or 1 time before using the SUI (0.65 vs. 0.74 respectively). This rate reaches a higher value - 0.80 - for group 3. This result could indicate that successive uses with the GUI make spoken dialogue easier in these contexts. If we now consider the reverse shift from the SUI to the GUI, results are rather similar whether users have used the SUI 0, 1 or 4 times before accessing PlanResto with the GUI. Apparently, the existence of a referent device does not seem to create specific difficulties in these contexts, as we expected. Using the GUI, users massively preferred the "click" modality over the other modalities: 92 % of users' actions were done by clicking on an item proposed by the system. This modality is obviously missing from the SUI, with which users preferred the "keyword" and "proposed choice" modalities. So, while users did not adopt the same interaction procedures with both user interfaces, they did not face difficulties in using PlanResto in these contexts.

We can explain these observations firstly by the appropriateness of the guidance proposed over the telephone: in each context, possible actions were explicated by the system, which could help users to easily adapt to the new interface. Moreover, we can assume that clicking on an item or uttering it are two action modalities which, since they share the same words, are less inconsistent that we expected.

Another type of inconsistency did not cause problems. It concerns the interaction modalities that users spontaneously adopted when they discover each device. During the first scenario, groups 1 and 3 mainly used the click modality (about 70% of actions), while groups 2 and 4 preferentially used natural language (74% of utterances). During the second scenario, this contrast between groups decreased. Clicking remains the main modality but with a lesser significance: it accounts for 55% of actions made with the GUI. And natural language used with the SUI was then associated to 50% of utterances. It is particularly interesting to note that group 1 users, who used the GUI first, adopted natural language less often than groups 2 and 4 users during their first scenario with the SUI. This effect is even stronger for group 3 users when they discovered PlanResto over the telephone after having used it 4 times with the GUI: contrarily to the users of groups 2 and 4, they did not massively adopt natural language and preferred "key-words" and "proposed choice" modalities. These results seem to reflect a procedure transfer when users shifted from the GUI to the SUI, which is probably motivated by effectiveness and reliability goals.

The opposite device shift, from the SUI to the GUI, did not lead to the same type of procedure transfer. We noticed at the 5th scenario that group 4 users, who interacted with the PC after four uses of PlanResto over the telephone, had a distribution of modalities quite similar to the groups 1 and 3 during their first scenario with the GUI (preference for "click"). We may conclude that natural language, which initially appears to be a relevant modality when speaking over the telephone, seems to be hardly transferable to graphical interaction, probably because it competes in this context with other available modalities that are known or judged to be more economic and accurate.

A last observation: we noticed that the "mixed" modality appeared during the 2d scenario in group 2 whereas it was absent at this stage in groups 3 and 4. Thus, it seems that alternating GUI and SUI encouraged users to *invent* this specific language made of keywords and natural language expressions. Hopefully, PlanResto user interfaces were sufficiently flexible to make this unexpected language acceptable.

3.3. Consistent contexts causing problems

PlanResto dialogue is consistent across the GUI and the SUI in C1 (request phrasing). In this context, some users' unexpected difficulties were observed.

With the GUI, some users encountered difficulties in phrasing their written requests. In general, they used shortened or inadequate expressions to designate restaurant locations (e.g. 11th arr or only 11th instead of 11th arrondissement). This difficulty was observed in all groups but more often in the groups 2 and 4. Based on spontaneous comments made by some users, we can assume that such behaviours have their roots in a misunderstanding of the system's skills: since over the telephone the system seems to be "intelligent" (it understands natural language!), some users expected it to be as intelligent once they shift to the GUI. It is interesting to note that users did not use the same wording to make their request on the SUI and the GUI. For instance, they did not say "I want a restaurant in the 11th". Actually, users spontaneously adapted their language to the perceived characteristics of the new device. But they seemed to do it consistently with what they believed to be the common skills of the system. Occasionally, this may cause unexpected and problematic behaviours.

Another specific difficulty in C1 appeared as OOP requests. They were observed in the first scenario with the SUI: some users' questions were phrased with expressions that exceeded the system's comprehension capacities. For example, a user asked: "Then, you confirm there is no Japanese restaurant in the 7e for 20€?" or another asked: "You give me the name of a restaurant closed to Centre Beaubourg. Could you give me again the exact address please?" The interesting fact is that none of these types of questions was recorded with the GUI even if asking questions in natural language was also possible. To account for this discrepancy between SUI and GUI behaviours, we can speculate that, with the SUI, the weak effort associated to spoken expression (compared to typed expression) have encouraged users to speak more freely in asking information.

4. Discussion and conclusion

This study aimed at studying how users may react to inconsistencies between a SUI and a GUI of a multi-device service. Several indications shows that knowledge transfers may occur when shifting from a known to a new device. Our initial view was that cross-device transfers may facilitate the discovery of a new device when inter-device consistency is preserved and can damaged it when users face inter-device inconsistencies. Actually, some observations confirm this view but others lead us to relativize it. Thus, we also observed how users may easily adapt to or even prefer some inconsistencies, especially at the level of interaction procedures. In particular, consistency between devices does not seem necessary if more economic and reliable alternatives procedures are known or can be easily detected by users when they discover a new device.

Also, some characteristics of each user interface seem to *implicitly* determine differentiated usages. In particular, interaction effort varying according to the available input devices, users may be inclined to adopt different interaction modalities with different devices.

These results suggest that multi-device service design should attempt to meet the consistency criteria but without neglecting the need to study users' activity, reasoning and preferences.

Our study also suggests a few recommendations for designing multi-device service with a SUI and a GUI:

- Make any clickable label on the GUI understandable with the SUI. If it is not possible, choose another clickable label (e.g., a rank or a letter) and associate it with an unclickable name.
- Make possible actions explicit during the spoken dialogue and use the same action labels on the GUI.
- Don't force users to use a specific modality available on both device.
- Provide enough flexibility in utterance production (as it is possible with Artimis, the core technology used in our study) to allow users to change their expression modality as they gain experience with each device and also to invent new interaction modalities.
- Prepare to face far more out-of-perimeter requests and commands with the SUI than with the GUI. To limit their number, design the spoken dialogue system so as to process them and make the system transparent during dialogue to help users to understand its real comprehension capabilities [5].

5. References

[1] Denis C. & Karsenty L. (2004) Inter-usability of multi-device systems: a conceptual framework. In: Seffah A. & Javahery H. (eds.) *Multiple User Interfaces: Cross-Platform Applications and Context-Aware Interfaces.* Wiley & Sons.

[2] Grudin, J. (1989). The case against user interface consistency. *Communications of the ACM*, 32, 10, 1164-1173.

[3] Sadek, D., Bretier P., & Panaget, F. (1997) Artimis: Natural dialogue meets rational agency. *Proceedings of IJCAI'97*, Nagoya, Japan, pp. 1030-1035.

[4] Seffah A. & Javahery H. (eds) (2004) *Multiple user Interfaces. Cross-Platform Applications and Context-Aware Interfaces.* Wiley & Sons.

[5] Karsenty L. & Botherel V. (2005) Transparency strategies to help users handle system errors. *Speech Communication*, 45, 305-324