Chapter 12: 10 major challenges

Lausanne Senior Meeting

12.1. Social interactions

Some recent studies are dedicated to defining patterns for designing conversational or even interfaces, which are somehow still underdeveloped for the moment, despite the progress in related technologies, such as emotion and speech recognition, and Text-To-Speech interfaces. Until recently, the overall interactive cycle between human and computer was ignored.

The specificity of these interfaces is indeed their high level of interactivity with the user(s). Consequently, in addition to be able to control all input and output technologies, they must be as adaptive and reactive as possible, because users' behaviours vary with the individual and circumstances and can adapt to the behaviour of the animated persona that they are facing.

The empirical studies that have been conducted recently are trying to enhance the knowledge related to these changes and, accordingly, to improve the design of multimodal interfaces.

First, people do not automatically use several modalities when given the possibility; they are indeed more likely to interact "unimodally" rather than multimodally when continuing a dialogue that has already been initiated with the system or answering an easy request. The choice among available modalities also
SIMILAR Dreams
depends on several factors: context (mobile use or not), type of request, and individual differences among users. Moreover, when it comes to acting multimodally, people do not all have the same behaviour. Some people prefer to use all modalities together, synchronously, whereas others prefer to use them sequentially. This small difference can have a major impact on the system's interpretation of the inputs. Finally, user category-specific differences also exist. So, elderly users, for example, will use self-speech when confronted with harder requests, which very concretely may require a different configuration of the speech recognition software.

12.2. Software Architecture reference model

Some software architecture reference models have already been proposed for multimodal applications. The authors are nevertheless of the opinion that these existing reference models do not address all the key issues that they have identified, to wit,

- explicit definition of fusion/fission mechanisms for input and output;
- different levels of abstraction where the fusion/fission takes place: lexical, syntactic, semantic, task fusion (could be based on Nielsen's linguistic model of interaction);
- explicit support for reconfigurability for better context-adaptation (especially at runtime as opposed at design time): self-description, exchange protocol; and
- support for multi-* concerns: multi-user, multi-device, multi-platform, multi-media, multi-modal, and so on.

As of now, no single reference model addresses all these requirements.

Standardisation work in this direction is currently being conducted within the W3C working group on multimodal
interfaces, with the collaboration of several SIMILAR members. It is nevertheless far from complete.

12.3. Mix of architectural styles

Several architecture models exist for multimodal interfaces, but so far mixing them in order to improve their characteristics has not been attempted. The authors think this could be a path to explore.

12.4. Software component description

(cfr. CDIL)

a. Services they provide
b. Services they request
c. Resources requested
d. Quality of Service (QoS): stability, latency, and precision

12.5. Polymorphic user interfaces

Interfaces are in principle designed to fit their target users' requirements, skills, etc. But it is not always feasible to have one single interface per user category, e.g., children, adults, people with disabilities, etc. Furthermore, the context also intervenes in the design of the interface, and total control over the context is not possible. Polymorphic UI have several different independent representations. They are thus highly interesting in considering these aspects.

12.6. Genericness of model, methods, and tools

As of now many tools and methods have been and are still being developed, while different models co-exist. Some of them cover
different areas, but others greatly overlap. This competition offers the advantage of leaving room for choice, but it also slows down development because of the lack of genericness. This issue should thus be addressed.

12.7. Scalability

Currently most of the multimodal interfaces that are developed are limited when it comes to the platform(s), number of users, and/or modalities. This lack of capability for extension obviously limits their potential for large adoption.

Scalability should be considered when developing a multimodal interface, in order to enable and facilitate its upgradability.

12.8. Reference Glossary

To bring HCI and DSP closer to each other, which is highly desirable, not to say necessary, for future research in the area of multimodal interfaces, one of the first things to do is indisputably to set up a common vocabulary.

Until now, these two worlds have indeed been using the same words, but their definitions were not always the same, even within the same discipline. With such different meanings, communication and understanding between the two communities effectively remain difficult.

This issue was highlighted at SIMILAR’s Lausanne meeting, where two days of presentations were spent understanding what exactly was covered by the notions of fusion and fission, which are central to multimodal interface development, in each community.

The common vocabulary should ideally cover the most usual areas of Multimodal Applications and User Interfaces.

Such unification work has already been initiated in the CAMELEON Project (Context Aware Modelling for Enabling and
SIMILAR Dreams
Leveraging Effective interaction (ET), a Shared-Costs RTD IST Project, which has given rise to a glossary that can be consulted at:
http://giove.cnucen.cnr.it/cameleon/glossary.html
Nevertheless, the task is far from finished:

- CAMELEON is more HCI oriented, meaning that additional input from the DSP community is needed.
- Some terms are still deemed "dangling", as no unanimous agreement on their definitions could be reached.
- Some terms are still missing.

SIMILAR has thus decided to join CAMELEON in this challenging work.

For this purpose, meetings of senior researchers from HCI and DSP inside the NoE are planned, in constant interaction with the outside world through the SIMILARnet portal, which will reflect the latest state of the definition work, but also through contacts with internationally acknowledged experts in the area of multimodal interfaces.

The hope is that this continuous communication will culminate in the adoption of a set of definitions that is widely accepted both inside and outside the consortium after which this set will gradually be extended to cover all the key words used in multimodal interfaces.

All editions of the SIMILAR Dreams book will also include the latest version of the glossary: this will constitute both a valuable dissemination tool and a simple way to check the work's progress simply by comparing the glossaries of the different editions.
12.9. Generic contactless multimodal interfaces

Contactless multimodal interfaces are considered a valuable choice, notably for edutainment applications. They effectively avoid all physical contact between the user and the system, which greatly lowers the risk of the equipment’s being damaged. Despite this undisputable advantage, to date no generic contactless multimodal interface exists.

12.10. Security/privacy in context-sensitive multimodal applications

In order to interact with users efficiently, multimodal interfaces have to capture a wide range of information concerning them, from their explicit instructions to their location, the emotions they express, etc. As not all this information is consciously emitted by the user, privacy issues may obviously arise. What happens if a user does not want the system to capture what he is talking about, for example? Security is also an important question to be handled, especially in sensitive domains such as military and medical multimodal applications. A multimodal interface should at least make the user aware of what is being captured about him/her, and provide measures to avoid any illegitimate use or processing of this information.
12.11. References


www.iis.ee.ic.ac.uk/~j.pitt/teaching/HCI_AffectiveComputing_2002.ppt


www.w3.org/TR/2005/WD-mmi-arch-20050422/


Constantine Stephanidis, Demosthenes Akoumianakis, Universal design: towards universal access in the information society, CHI '01 extended abstracts on Human factors in computer systems, March 31-April 05, 2001, Seattle, Washington

http://giove.cnuce.cnr.it/cameleon/glossary.html