

Providing multimodal context-sensitive services to mobile users

◦ Carmelo Ardito, ◦* Thomas Pederson, ◦ Maria Francesca Costabile, ◦ Rosa Lanzilotti

◦ Dipartimento di Informatica, Università di Bari, 70125 Bari, Italy

* Department of Computing Science, Umeå University, SE-90187 Umeå, Sweden
{ardito, pederson, costabile, lanzilotti}@di.uniba.it
top@cs.umu.se

Abstract. In this paper, we describe a framework designed to provide multimodal context-sensitive services to mobile users. This work is part of the CHAT project, which aims at developing a general-purpose infrastructure for multimodal situation-adaptive user assistance. We specifically describe two conceptual corner stones of the project: a) a multimodal interaction framework targeted at providing service access through different modalities for different real-world situations and for improving interaction with mobile devices in general, b) an “egocentric interaction” model for framing interaction with objects in the vicinity of a mobile user, including also other real-world and/or computational entities than the mobile device itself, ranging from computationally “stupid” everyday objects to more advanced interactive devices such as desktop PCs. The final section of the paper is devoted to open issues in the design of the CHAT infrastructure related to the topic of user assistance in intelligent environments.

Keywords: Multimodal interfaces, mobile human-computer interaction.

1. Introduction

The CHAT project (“Cultural Heritage fruition & e-learning applications of new Advanced (multimodal) Technologies”) aims at developing a software infrastructure that provides services accessible through thin clients such as cellular phones or PDAs to be: a) adaptable to personal preferences of the user, with focus on the choice of interaction modalities; and b) adaptive to the physical-virtual context of the human actor carrying the device. In both cases, the proposed architecture should be open both for channeling interaction between services and user through the mobile device itself, as well as through available input and output facilities in the vicinity. Furthermore, real-world phenomena sensed by the device itself or indirectly through external sensor pools will be made available through the CHAT infrastructure as a resource for service developers to effectively design “intelligent” environments.

Users are allowed to interact with the system using several input channel simultaneously, classified by W3C as simultaneous co-ordinated multimodality [7]. Empirical studies proposal targeting this kind of multimodality are described in [1, 2]. The architecture for supporting these kind of multimodal systems is more complex than traditional interactive systems, because we have to consider: a) parallel recognition

modules for each input channel, i.e., every module produces fragments of the overall input that must be combined to become meaningful; b) a general methodology to interpret the meaning of the input fragments; c) a time-sensitive analysis process to determine which fragments must be combined to become meaningful; d) a module to manage the overall user/system dialogue; e) criteria to adapt the input/output modalities to the users' needs and the environment in which they actually are.

The infrastructure is as general as possible because multimodal adaptability and adaptive features are beneficial independently from the area of application. For the purpose of evaluation however, the CHAT project will develop mobile multimodal prototype applications related to two particular activities: e-learning and exploration of cultural heritage.

System development is guided by a multimodal design framework for ensuring state-of-the-art support for multimodal interaction, and an "egocentric interaction" model for guiding the analysis and design of context-aware services.

2. Multimodal interaction framework

The framework designed for the CHAT project is compatible with the W3C Multimodal Interaction Framework, which is not an architecture, but a modeling framework one step above in abstraction [6]. The W3C multimodal interaction framework describes neither how components are allocated to hardware devices, nor how the communication system enables the hardware devices to communicate.

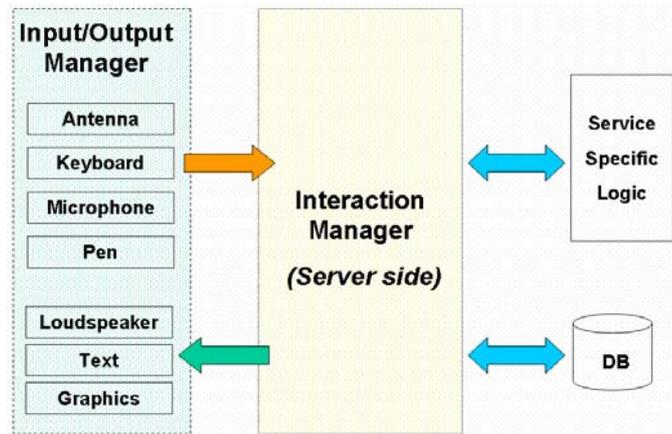


Fig. 1. CHAT Multimodal Interaction Framework.

In the CHAT multimodal interaction framework (Fig. 1), the *Input/Output Manager* is a "lightweight" software running on the user device with the responsibility of managing the input and output channels. I/O Manager captures the input fragments coming from several channels and transmits them to the Interaction Manager running on a server using suitable standard protocols.

The *Antenna* in the framework symbolizes high-level position information received through for instance GPS, GSM, or RFID technologies.

The *Interaction Manager (IM)* receives the multimodal input fragments from the user’s device and processes them to obtain a meaningful input.

The *Service Specific Logic* acts on the input received and potentially produces an output transmitted to the IM.

Finally, the IM generates a multimodal representation of the required service that will be presented to the user by output channels suited to the user’s preferences and needs as well as current environmental context and device type.

3. Egocentric interaction framework

The user interface design in the CHAT project will be guided by the *egocentric interaction* framework [3], inspired by the currently popular view within cognitive science that human individual actions are to a large degree influenced by what the specific individual can perceive of the surrounding environment. Based on an integrated view on physical and virtual space [4] objects in the proximity of a particular human actor can be categorized as being situated in one out of four spaces, at any given point in time (see Fig. 2 left).

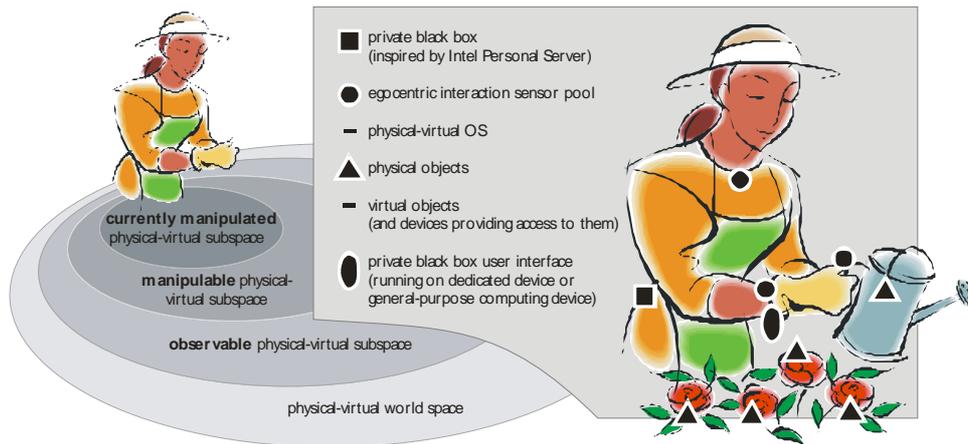


Fig. 2. Left: A situative model of physical-virtual space from the perspective of a specific human actor at a specific point in time. Adapted from [4]. **Right:** General components of an egocentric interaction system.

The borders between the different subspaces are based on assumptions of the specific human actor’s perceptive and cognitive experience of the current physical-virtual environment, e.g. real-world objects and/or virtual objects (presented by, and accessed through computing devices) in the immediate vicinity of the specific human actor. The egocentric interaction framework is based on the belief that computer system models should be closely tied to the cognitive and perceptive models that human actors construct and maintain as part of everyday life. The term ‘egocentric’ has been chosen to signal that the human body and mind of a specific human individual act as a centre of reference to which all interaction modeling and activity assistance is anchored.

The general components of an egocentric interaction system are illustrated in Fig. 2 right. The system aims at permitting smooth transitions between performing actions in the physical world and in the virtual world. It is based on a wearable computing/sensing hardware configuration consisting of a *private black box* offering computing power and storage space for data generated by an *egocentric interaction sensor pool* monitoring object-centric phenomena within the observable physical-virtual subspace of a specific human actor. Furthermore, the private black box runs a *physical-virtual operating system* hosting both advanced physical-virtual applications developed by software developers as well as simpler programs designed by the user her/himself.

Such systems can incorporate the manipulation of both physical objects (e.g. a sculpture at a museum) and virtual objects (e.g. a web page describing the same sculpture). Explicit interaction with the physical-virtual operating system is performed through a *private black box user interface*, either fitted onto the private black box itself, or running on a general-purpose device like a PC. Implicit interaction [5] with the physical-virtual operating system emerges whenever the user interacts with a physical or virtual object inside the manipulable physical-virtual subspace (see Fig. 2 right) monitored by the private black box.

Due to limitations in computation capability of current mobile devices, the *black box* in the egocentric interaction system model (Fig. 2 right) will in the CHAT project be distributed over both client and server illustrated in Fig. 1, allowing the Interaction Manager on the server side to model and compute necessary actions based on the sensor and input device data delivered by the Input/Output Manager on the client side. As mobile computation capability increases, a migration of functionality from server to client is expected. The *private black box user interface* will be based on the existing I/O capabilities of the chosen mobile device with a certain emphasis on modalities that allows for hands-free interaction. The *egocentric interaction sensor pool* will be based on sensors inbuilt into the client device (primarily GPS) but complemented by sensors instrumented in the physical environment. The *physical-virtual operating system* will be implemented as an application running on top of the existing device specific operating system with the primary function of creating a cohesive user experience of physical (in the environment) and virtual (in the device) events.

4. A scenario in CHAT

A scenario that can merge e-learning and cultural heritage is the exploration of an archaeological park to learn about an ancient city. In the CHAT project we have chosen the city of Egnazia, in the Puglia region, as application domain. The city walls date back to the messapian phase, from the end of the 5th century B.C.. Pre-Roman tombs have been found within the city walls, but the true Messapian necropolis are situated outside the city. The Via Traiana which runs through the centre of ancient Egnazia, was part of a major new road build by Emperor Trajan in 109 A.D. The destruction of the city traditionally linked to the invasion of Totila, King of the Goths, in 545 A.D.

The scenario refers to a class of middle schools pupils that use the CHAT system. Before visiting the Egnazia archaeological park, the pupils study its history through an e-learning platform. Then, pupils can visit the park either in a traditional way, guided by an expert, or using a mobile device. The pupils can walk across the excavations and, when

they arrive near a building ruins, the mobile device receives a signal from sensors located in the vicinity, and provides pupils with a 3D reconstruction of the original aspect of that building, with navigation possibilities and access to, together with related information.

At the end of the visit, the pupils may also perform a test to evaluate their learning. The test is organized as a treasure hunt: for example, the system asks the pupil to reach a specific place of the park. The system, through the sensory placed in the environment, automatically determines if the pupil has reached the right place and proposes him/her the next question. The system keeps track of the pupils' movements in the park, thus it is possible to virtually recreate the test execution.

5. Conclusion

The CHAT project is in a preliminary phase. More questions about the multimodal framework and the egocentric interaction model proposed in the project are currently open issues. We find the following questions particularly relevant:

- Which user activities and tasks require assistance? What method or heuristic should we use to identify the most useful and practically achievable assistance for e-learning and exploration of cultural heritage?
- How should the designer choose the best sensing and interaction technologies for a scenario? Having a small mobile device as computation and communication hub among user, “intelligent” environment and server: how should the service logic (see Fig. 1) be developed in order to seamlessly cope with ad-hoc appearance and disappearance of sensors and actuators external to the device itself?

Hopefully, the research activities carried out in CHAT will provide answers to the above questions.

References

1. Oviatt, S., De Angeli, A. and Kuhn, K. (1997). Integration and synchronization of input modes during multimodal human-computer interaction. In Proceedings of CHI'97, Atlanta, Georgia, USA, 18-23 Apr, 1997, pp. 415-422.
2. Paternò, F., & Giammarino, F. (2006). Authoring interfaces with combined use of graphics and voice for both stationary and mobile devices. Proceedings of the International Conference on Advanced Visual Interface 2006 (AVI 2006), Venice, Italy, May 23-26, 2006, pp. 329-335.
3. Pederson, T. (2006). Egocentric Interaction. Workshop on What is the Next Generation of Human-Computer Interaction?, CHI2006, April 22-23, Montréal, Canada.
4. Pederson, T. (2003). *From Conceptual Links to Causal Relations — Physical-Virtual Artefacts in Mixed-Reality Space*. PhD thesis, Dept. of Computing Science, Umeå university, report UMINF-03.14, ISSN 0348-0542, ISBN 91-7305-556-5. Permanent URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-137>
5. Schmidt, A. (2002). *Ubiquitous Computing – Computing in Context*. PhD thesis, Computing Department, Lancaster university, U.K.
6. W3C Multimodal Interaction Framework. <http://www.w3.org/TR/mmi-framework/>
7. W3C Multimodal Interaction Activity. <http://www.w3.org/2002/mmi/>