# A Situative Space Model for Distributed Multimodal Interaction

Thomas Pederson, Dipak Surie
Dept. of Computing Science
Umeå University
SE-90187 Umeå, Sweden
+46 90 786 {6548,9672}
{top, dipak}@cs.umu.se

## ABSTRACT
This paper presents a body-centered model of human-environment interaction based on proximal and perceptional relationships between the human actor and objects of interest (physical objects, virtual objects, and mediators). The model is applied to a real world scenario and also discussed as a tool for designing a distributed multimodal interaction infrastructure.

## Keywords
Ubiquitous Computing, Human-Computer Interaction

## 1. INTRODUCTION
New paradigms for computer use evolve as an effect of continuous technology development and the society's way of embracing it. Lars-Erik Janlert distinguishes between the classical virtuality paradigm, the new mobility paradigm, and the emerging ubiquity paradigm. The virtuality paradigm is about seeing and using computers as windows to a virtual world; the mobility paradigm puts emphasis on an improvised and situational use of computing power somewhat independent from physical space; the ubiquity paradigm depicts the use of computers embedded into everyday real-world objects and environments [1].

Coincidently, three out of the four themes listed in the call for papers for this workshop on future mobile experiences can be 1-1 mapped to the three computer use paradigms: "device interaction" is dealt with in virtuality paradigm research; "location/proximity interaction" is a hot subject within the mobility paradigm; and "physical interaction" is at the core of the ubiquity paradigm. This paper attempts to cover important aspects of human-computer interaction in all three computer use paradigms.

Is it possible to create a model that can capture all this? We hope to contribute in this direction and have started to construct a human body-and-mind centric "egocentric" framework for formalizing aspects of device interaction, location/proximity interaction, and physical interaction all in the same model. In this paper we present our ongoing efforts in expanding the framework to encompass interaction modalities beyond vision and tactility.

### 1.1 The real world is more than "context"
Within the mobility paradigm, device-centric models dominate. One challenge often addressed is the adaptation of the interaction dialogue and modalities inherited from the virtuality paradigm to the new more limited platform. The mobility paradigm has brought with it a new element to the dialogue between user and computer however: physical context [13]. By making mobile devices context aware, the idea is that the mobile computing system can prevent inadequate requests for user action as well as optimizing the modalities in the dialogue.

However useful this approach for bringing real world aspects into the mobile computing paradigm might be, we believe, as researchers of the ubiquity paradigm, that it can be taken much further. We believe that by treating the real-world phenomena as merely context to events in the virtual world (the world accessed through the mobile device), important aspects of human activity are lost. A further reason for system designers to give the real world greater attention is the fact that the humans using their systems will. (This is of course especially true in the cases of the mobility and the ubiquity paradigms.)

### 1.2 Goal
As an inseparable part of our work in going beyond treating the real world as context only, we also address other known open issues within the mobility and ubiquity paradigms, including:

- **The framing problem:** what real-world and digital objects are parts of the computing "application" at any given time? The common desktop metaphor in the virtuality paradigm, defining what is part of the dialogue and what is not, has no correspondence in the ubiquity paradigm.
- **Implicit HCI:** How can we create a clear and useful distinction between explicit and implicit interaction [10] within intelligent environments that simplify both for designers and humans acting in these environments?
- **Human attention:** When and how should a mobile human actor be interrupted?

## 2. THE EGOCENTRIC INTERACTION FRAMEWORK
The proposed view on human-computer interaction (HCI), or better: *human activity*, is based on a set of conceptual cornerstones. We highlight some of them below, contrasting them to concepts currently more widely used in HCI modeling.

### 2.1 A human in the world instead of a user in front of a computer
In the egocentric interaction framework [14], the modeled human individual is viewed as an actor moving about in a physical-virtual world, not as a "user" performing a dialogue with a computer. The human body and mind of a specific human individual (sometimes literally, as will be shown later) acts as centre of reference to which all interaction modeling is anchored.

## 2.2 Virtual objects and mediators instead of interactive devices

The physical-virtual world mentioned in the previous paragraph is made up of physical objects, virtual objects, and mediators. Human activity is considered a process in which physical and virtual objects change states as an effect of being manipulated by the human actor performing the activity. There are no computing devices in the model. Instead, any artifact providing access to virtual objects are seen as a mediator able to transform the unperceivable bits of the virtual world into something humans can sense (in the case of system output) and events in the physical world into events in the virtual world (in the case of input from the user to the system). Examples of mediators commonly used within the virtuality paradigm are visual displays, loudspeakers, keyboards and mice.

The motivation for playing down the role of input and output devices is twofold: 1) if we choose to ignore the fact that virtual objects have to be mediated, we can model interaction with physical and virtual objects in a uniform way and 2) it aligns well with our picture of expert users who are well acquainted with the computing devices they use and mentally forget about them as they direct their attention to the (virtual) object of interest instead.

## 2.3 Manipulation and observation instead of input and output

Manipulation is often inseparable from observation in everyday physical activities. You cannot change the state of an object (e.g. clicking out the point of a ball pen in order to prepare it for writing on paper) without also observing the result (you hear the click sound, you feel the latch locking the ballpoint into its "ready" position, and you might even choose to visually confirm that the ballpoint is out). This tight coupling between manipulation and observation is due to the combination of the way the objects are designed, how we choose to manipulate them, and how the manipulation process interplays with the laws of physics.

The most common interaction mechanism for changing the state of objects within the virtuality paradigm, i.e. that of direct manipulation [11], makes this strong relationship between manipulation and observation prevail also in many parts of the virtual world. We suggest to substitute the concepts of (device) "input" and "output" common in many HCI models with (object) "manipulation" and "observation" in order to enable the modeling of interaction with real and virtual objects alike. Note that we see object manipulation and observation as processes that can take place in any modality including tactile, visual, audio, etc.

## 2.4 The situative space model

A *situative space model* is developed on the basis of what a specific human actor can perceive and not perceive (observable space), manipulate and not manipulate (manipulable space) at any given moment in time as illustrated in Fig. 1.

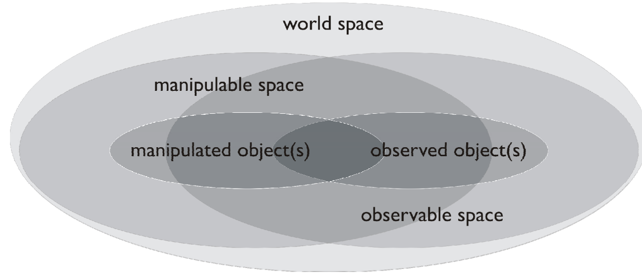


**Fig. 1.** The situative space model, adapted from [2].

Human actors in general situate themselves closer to the objects of interest relevant for their current activity. Mediators that are situated closer to the human actor's body avoid forcing the actor to make changes to his/her location while interacting with physical objects and virtual objects. By capturing the physical objects, virtual objects and mediators within the actor's observable space and manipulable space, we are indirectly filtering the set of possible interaction channels for interaction between the human actor and a set of physical and virtual objects at a particular moment in time.

- **Observable Space (OS):** The space containing a set of physical and virtual objects that are potentially perceivable by a human actor with minimal efforts at a particular moment in time.[1]
- **Observed Objects (OO):** A subset of the observable space containing a set of physical and virtual objects with their state changes being perceived by a human actor by coupling with mediators at a particular moment in time.
- **Manipulable Space (MS):** The space containing a set of physical and virtual objects that are potentially manipulable by a human actor with minimal efforts at a particular moment in time.
- **Manipulated Objects (MO):** A subset of the manipulable space containing a set of physical and virtual objects with their state changes being caused by a human actor by coupling with mediators at a particular moment in time.
- **World Space (WS):** The space containing a set of all the physical and virtual objects known to a system.

# 3. APPLYING THE FRAMEWORK

## 3.1 Application Scenario

The framework is applied to a smart home environment (under development) containing physical objects (including the wearable outfit worn by a human actor); virtual objects running on a personal server [12][3] and a set of thin-clients embedded onto the physical objects; and a set of mediators embedded onto the physical objects. Refer to Fig. 2 and Fig. 3 for more information about how physical objects, virtual objects and mediators are placed within the proposed situative space model. The user (the person drinking a glass of milk in Fig. 2) is performing his/her everyday activities (in this instance, having his breakfast with a friend) in a mixed-reality environment.

---

[1] Note that the term Observable Space is used to refer to all human perceivable modalities including vision, audio and tactile.

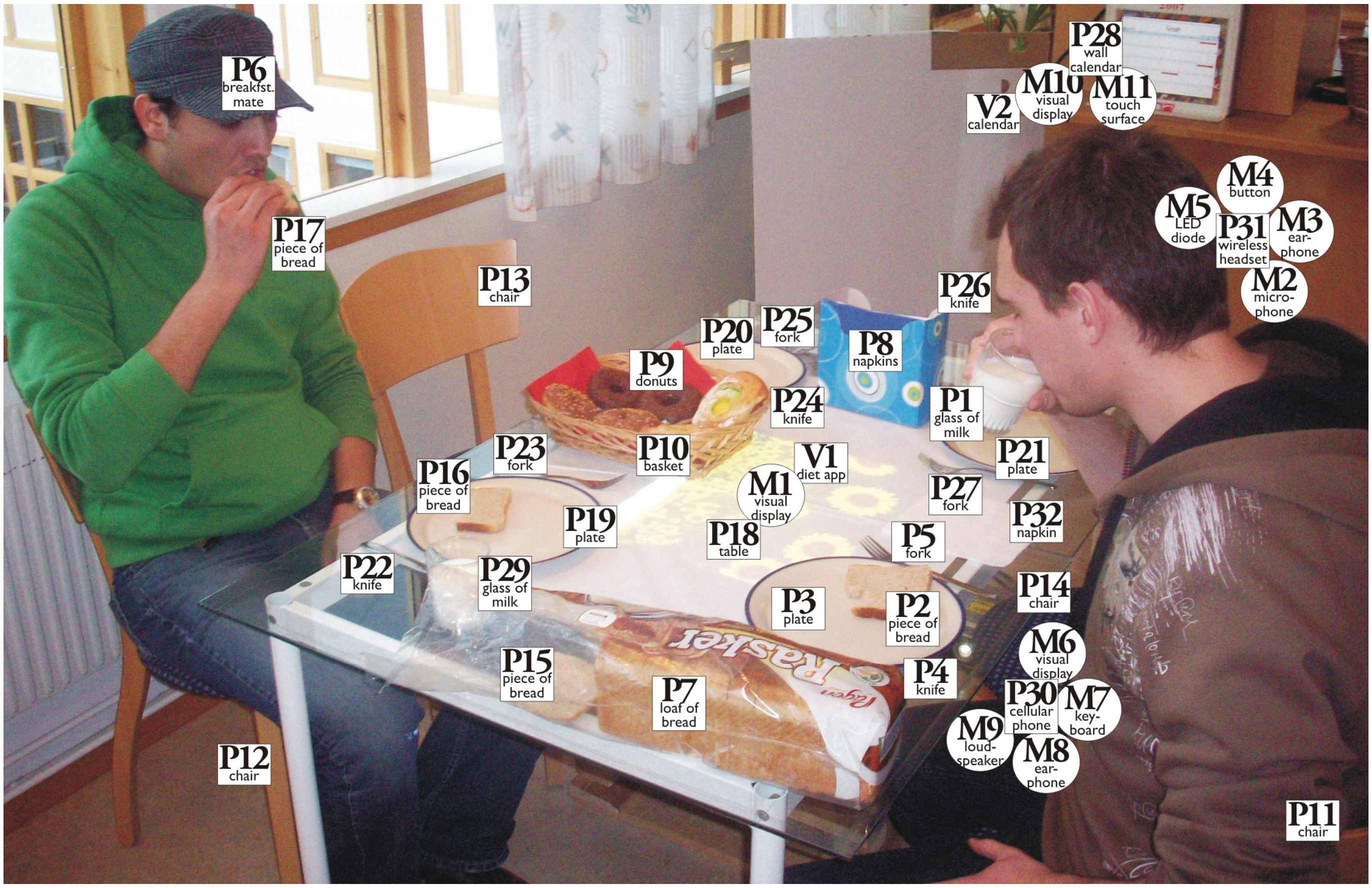


**Fig. 2.** Physical objects, virtual objects, and mediators in a breakfast scenario.

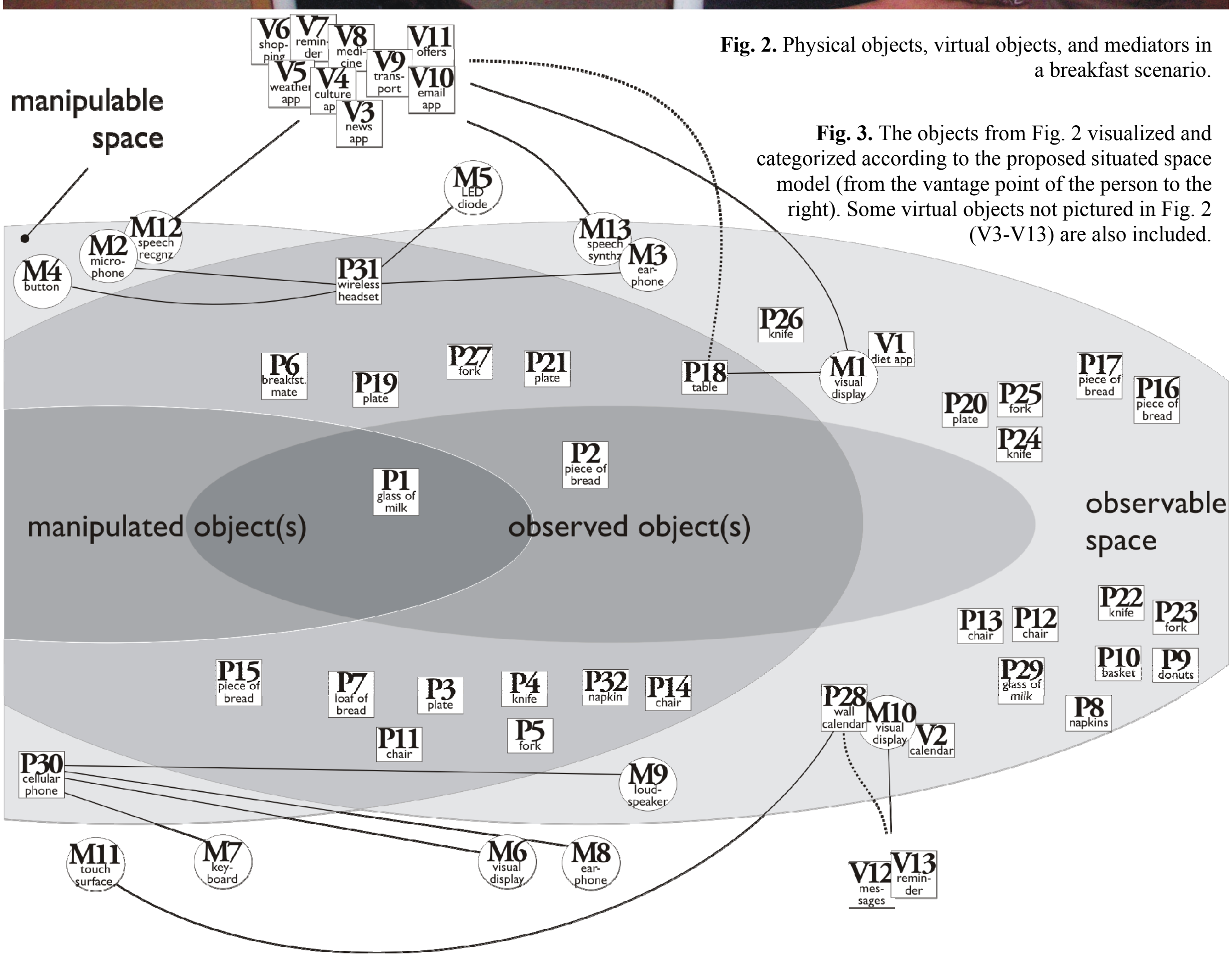


**Fig. 3.** The objects from Fig. 2 visualized and categorized according to the proposed situated space model (from the vantage point of the person to the right). Some virtual objects not pictured in Fig. 2 (V3-V13) are also included.

### 3.2 Capturing the Situative Spaces

**Using Proximity measure:** User's proximity to physical objects, mediators and virtual objects are considered important in framing the objects within individual spaces.

- *Case 1 (Objects connected to embedded thin-clients running WLAN connection):* The proximity measurement between an object (physical or mediator) and the user is calculated using the WLAN signal strength measures (Java Wireless Research API (JWRAPI) developed at Luleå University of Technology, Sweden [4] is used) combined with the objects' properties like physical size, no. of pixels (for visual displays), maximum loudness (for audio displays), etc. All objects (physical and mediator) within *Case 1* are connected to embedded thin-clients that provide their WLAN signal strength measures to a personal server [3] running on a wearable computing platform and connected to a WLAN access point. The proximity measure between virtual objects and the user is calculated based on the object's association with relevant mediators. For instance if a mediator is within the manipulable space and the thin-client to which the mediator is connected is running a set of virtual objects, then all those virtual objects are also within the user's manipulable space.
- *Case 2 (Objects passively tagged and using RFID Technology):* The proximity measurement between an object (physical or mediator) and the user is calculated using passive RFID Technology [5]. Embedded RFID readers are attached to the user's wrist watch or bracelet for both the hands (Skyetek's M1-mini is used considering its range of 5.8 cms to 8.5 cms [6]). Such wrist worn readers provide information about the objects that are grabbed or released by a user. Another RFID reader is embedded onto the user's chest to capture the set of objects within the user's manipulable space (Skyetek's M7 is used considering its range of 1m to 2m [6]). Objects within the user's observable space have a larger proximity value in general and is not captured using passive RFID technology. The proximity measure between virtual objects and the user is calculated based on the object's association with relevant mediators, similar to case 1.

**Using Orientation measure:** User's orientation to physical objects and mediators are also considered important in framing the objects within individual spaces.

- *Case 3 (Objects embedded with an array of Infrared LEDs):* The orientation measurement between an object (physical or mediator) and the user is calculated using the Infrared signal strength of the array of LEDs embedded onto respective objects and an Infrared camera worn by the user (Nintendo wiiremote [8] is used as an infrared camera connected to the personal server using Bluetooth communication) [7]. The individual arrays of infrared LEDs are turned on-and-off to create unique patterns that distinguish a particular object from the rest of the objects. Since infrared technology requires line-of-sight for communication, the objects that are outside the user's field of view could be discarded. Case 3 is mainly restricted to capturing observable space and observed objects within the visual modality.

### 3.3 Distributed Multimodal Interaction within the Situative Spaces

The objects present within the individual spaces play an important role in managing distributed multimodal interaction.

- *Case 4 (Physical Object manipulation as input):* The physical objects manipulated by a user are considered as implicit input to the system [13]. Manipulations of physical objects are captured using simple state-change sensors embedded onto physical objects. Wireless sensor networking (WSN) of physical objects is performed using ZigBee 802.15.4 OEM RF modules considering its low power requirements and out-of-the-box communication advantages [9]. Examples of implicit input include *turning ON the stove*, *opening the fridge door*, etc.
- *Case 5 (Virtual Object manipulation using input mediators):* The virtual objects are designed to be manipulated by a user using input mediators like microphone (BTH-8 Bluetooth headset as a microphone), accelerometer (Phidgets 1059 3-axis accelerometer), speech recognizer (Microsoft Speech SDK 5.1 API for speech recognition), gesture recognizer, etc. for providing explicit input to the system. For a user to manipulate virtual objects the user should first select the concerned mediator(s) and then an automatic association of his behavior to virtual object manipulation is created by the system. For instance, the user should say, "Select Speech User Interface" before manipulating virtual objects concerned using microphone and speech recognizer as mediators.
- *Case 6 (Virtual Object observation using output mediators):* The virtual objects are designed to be explicitly observed by a user using output mediators like ear-phone (BTH-8 Bluetooth headset as a ear-phone), speech synthesizer (Microsoft Speech SDK 5.1 API for speech synthesis), visual display, loud speaker, tactile display (Nokia E70 worn on the wrist), etc. We are currently working on a system component (Egocentric Interaction Manager [3]) that takes care of the management of the user's interaction with virtual objects using distributed multimodal mediators.

## 4. DISCUSSION

We have presented a model which links human actors' perceptional and motoric abilities to measurable proximity. By proximity we refer to the relationships between a specific human actor and the surrounding objects. In some cases, the relationship is constituted by plain Eucledian distance (e.g. the set of physical objects within the manipulable space) while in other cases it is defined by more complex associations (e.g. the set of virtual objects within the manipulable space).

## 5. ACKNOWLEDGMENTS

Our thanks to Prof. Lars-Erik Janlert, Olivier Laguionie, Thomas Johansson, Erik Lövbom, Florian Jäckel, and Dilip Roy.

## 6. REFERENCES


[1] Janlert, L.-E. The Evasive Interface – The Changing Concept of Interface and the Varying Role of Symbols in Human–Computer Interaction. HCII 2007, Springer LNCS 4550, pp. 117–126, 2007.

[2] Pederson, T. 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, ISBN 91-7305-556-5, (2003).

[3] Surie, D., Pederson, T. An Activity-Centered Wearable Computing Infrastructure for Intelligent Environment Applications. In Proceedings of IFIP EUC 2007 Conference on Embedded and Ubiquitous Computing, Springer LNCS 4808, (2007) 456-465.

[4] Java Wireless Research API. Luleå University of Technology. http://www.sm.luth.se/~johank/javawrapi/ (last visited on the 8th Sept. 2008).

[5] Finkenzeller, K. RFID Handbook. John Wiley and Sons, New York, NY, USA, Second edition, (2003).

[6] Skyetek Embedded RFID readers. http://www.skyetek.com/ProductsServices/EmbeddedRFIDReaders/EmbeddedRFIDReaderOverview/tabid/314/Default.aspx (last visited on the $8^{th}$ Sept. 2008).

[7] Jäckel, F. An Infrastructure for Human Interaction with Multiple Displays based on Spatial Alignment. Bachelor Thesis, Department of Computing Science, Umeå University, Sweden (2008).

[8] J. Lee. Wiimote Project. http://www.wiimoteproject.com/ (last visited on the $8^{th}$ Sept. 2008).

[9] Zigbee Alliance. http://www.zigbee.org/en/index.asp (last visited on the $8^{th}$ Sept. 2008).

[10] Schmidt, A. Implicit Human Computer Interaction through Context. Personal Technologies, Springer-Verlag, 4 (2&3), 191-199 (2000).

[11] Shneiderman, B. Direct Manipulation: A Step Beyond Programming Languages. In *IEEE Computer*, 16(8), 57-69 (1983).

[12] Want, R., Pering, T. Danneels, G., Kumar, M., Sundar, M., Light, J. The Personal Server: Changing the way we think about Ubiquitous Computing. In: Borriello, G., Holmquist, L.E. (eds.) Ubicomp 2002, LNCS vol. 2498, Springer, Heidelberg (2002).

[13] Schmidt, A. Ubiquitous Computing – Computing in Context, PhD thesis, Computing Department, Lancaster University, UK (2002).

[14] Pederson, T., Surie, D.: Towards an Activity-Aware Wearable Computing Platform based on an Egocentric Interaction Model. In: UCS 2007. Proceedings of the 4th International Symposium on Ubiquitous Computing Systems. LNCS, Springer, Heidelberg (2007).