

# Towards a Model for Egocentric Interaction with Physical and Virtual Objects

Thomas Pederson<sup>1</sup>, Lars-Erik Janlert<sup>2</sup>, and Dipak Surie<sup>2</sup>

<sup>1</sup>IT University of Copenhagen  
Rued Langgaards Vej 7  
2300 Copenhagen S, Denmark  
tped@itu.dk

<sup>2</sup>Dept. of Computing Science  
Umeå University  
90187 Umeå, Sweden  
{lej, dipak}@cs.umu.se

## ABSTRACT

Designers of mobile context-aware systems are struggling with the problem of conceptually incorporating the real world into the system design. We present a body-centric modeling framework (as opposed to device-centric) that incorporates physical and virtual objects of interest on the basis of proximity and human perception, framed in the context of an emerging “egocentric” interaction paradigm.

## Author Keywords

Interaction paradigm, user interface design.

## ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

## AN EMERGING PARADIGM: EGOCENTRIC INTERACTION

Important new advances in interaction and sensor technology, multimodal Human–Computer Interaction (HCI) [11], mixed initiative user interfaces [2], attention aware systems [6], new approaches for incorporating real-world entities [1], new theories on human activity [8][3] are working towards a new framing of interaction that we call *the egocentric interaction paradigm*. It extends and modifies the classical user-centered approach in HCI [4] on several points, including:

- 1. Situatedness.** Acknowledges the primacy of the agent’s current bodily situation at each point in time in guiding and constraining the agent’s behavior. The situation is the agent’s natural primary vantage point: selecting what can be perceived and attended to, limiting what can be performed.
- 2. Attention to the complete local environment.** Makes it a point to take the whole environment into consideration, not just a single targeted artifact or system.
- 3. The proximity principle.** Makes the assumption that proximity plays a fundamental role in determining what can be done, what events signify, and what agents are up to.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

NordiCHI 2010, October 16–20, 2010, Reykjavik, Iceland.  
Copyright 2010 ACM ISBN: 978-1-60558-934-3...\$5.00.

**4. Changeability of environment and agent–environment relationship.** Takes into account agents’ more or less constant movements of head, hands, sense organs, and body, locally and through the environment, as well as agents’ constant rearrangements and modifications of various parts of the environment.

**5. The physical-virtual equity principle.** It is neither biased towards interaction with “virtual,” immaterial data objects (classical HCI), nor towards interaction with physical objects and machines (classical ergonomics and HMI, Human–Machine Interaction): it pays equal attention to virtual and physical objects, circumstances, and agents.

We have chosen the term “egocentric” to signal that it is the human body and mind of a specific human individual that (literally) acts as centre of reference to which all modeling is anchored in this interaction paradigm. In the context of this article, the term should not be taken as a synonym for “selfish” or other similar higher-level personality traits but instead as the lower-level approach which human agents in general are forced to adopt to perceive and act in the world based on their senses and cognitive abilities, even when working in groups and with shared goals.

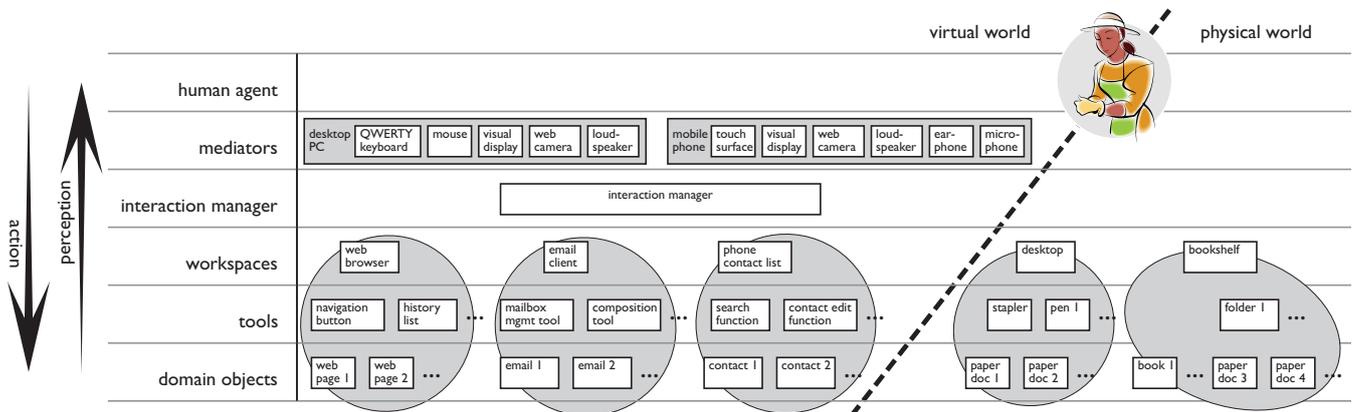
In the remaining part of this paper, we present how the characteristics above have influenced our modeling efforts.

## ACHIEVING PHYSICAL-VIRTUAL EQUITY

We believe the current physical-virtual gap could be made easier to cross for human agents by introducing a mixed-reality infrastructure having an “interaction manager” as central component, responsible (in collaboration with the human agent) for channeling communication between human and system through the currently best available devices and modalities [9]. Fig. 1 illustrates, in principle, how the interaction manager mediates information between virtual objects (a distinction between “workspaces”, “tools”, and “domain objects” is made in this particular model of an office) on the lower three levels of abstraction and the human agent on the top.

## Virtual Objects and Mediators instead of Interactive Devices

Input and output devices can be viewed as *mediators* through which virtual objects are accessed. The purpose and function of mediators is that of expanding the action space and perception space of a given human agent (Fig. 2).



**Fig. 1.** Parts of an office environment modeled with physical-virtual equity in mind and with an interaction manager component handling the interaction between human agent and software services offering access to virtual objects (workspaces, tools, and domain objects).

In the situation pictured in Fig. 1, the interaction manager component can choose among mediators embedded in a desktop PC and a mobile phone to enable the human agent to manipulate and observe virtual objects. Physical objects, to the right in the figure, are of course accessed directly.

#### Action and Perception instead of Input and Output

In the egocentric interaction paradigm, the modeled human individual needs to be viewed as an agent that can move about in a mixed-reality environment, not as a “user” performing a dialogue with a computer. If we take the physical-virtual equity principle seriously, also the classical HCI concepts of input and output need to be reconsidered. We suggest substituting the concepts of (device) “input” and “output” with (human agent) “action” and “perception”. Note that we see object manipulation and perception as processes that can take place in any modality: tactile, visual, aural, etc.

#### A SITUATIVE SPACE MODEL

The situative space model (SSM) is intended to capture what a specific human agent can perceive and not perceive, reach and not reach, at any given moment in time. This model is for the emerging egocentric interaction paradigm what the virtual desktop is for the PC/WIMP (Window, Icon, Menu, Pointing device) interaction paradigm: more or less everything of interest to a specific human agent is assumed to, and supposed to, happen here.

#### Main Components of the Model

The following definitions are agent centered but not subjective; they are principally aimed at allowing objective determination for automated tracking purposes. A more extensive description of the model can be found in [5].

**World Space (WS):** A space containing the set of all physical and virtual objects to be part of a specific model.

**Perception Space (PS):** The part of the space around the agent that can be perceived at each moment. Like all the spaces and sets defined below, it is agent-centered, varying continuously with the agent’s movements of body and body parts. Different senses have differently shaped PS, with different operating requirements, range, and spatial and

directional resolution with regard to the perceived sources of the sense data. Compare vision and hearing, e.g.

Within PS, an object may be too far away to be possible to recognize and identify. As the agent and the object come closer to each other (either by object movement, agent movement, or both) the agent will be able to identify it as X, where X is a certain *type* of object, or possibly a unique individual. For each type X, the predicate “perceptible-as-X” will cut out a sector of PS, the distance to the farthest part of which will be called *recognition distance*.

**Recognizable Set (RS):** The set of objects currently within PS that are within their recognition distances.

The kind of object types we are particularly interested in here are object types that can be directly associated with activities of the agent – ongoing activities, and activities potentially interesting to start up – which is related to what in folk-taxonomy studies is known as the basic level [7].

To perceive the status of a designed object with regard to its relevant (perceivable) states (operations and functions as defined by the designer of the artifact) it will often have to be closer to the agent than its recognition distance: the outer limit will be called *examination distance*.

**Examinable Set (ES):** The set of objects currently within PS that are within their examination distances.

**Action Space (AS):** The part of the space around the agent that is currently accessible to the agent’s physical actions. Objects within this space can be directly acted on. The outer range limit is less dependent on object type than PS, RS and ES, and is basically determined by the physical reach of the agent, but obviously depends qualitatively also on the type of action and the physical properties of objects involved; e.g., an object may be too heavy to handle with outstretched arms. Since many actions require perception to be efficient or even effective at all, AS is qualitatively affected also by the current shape of PS.

From the point of view of what can be relatively easily automatically tracked on a finer time scale, it will be useful to introduce a couple of narrowly focused and highly dynamic sets within AS (real and mediated).

**Selected Set (SdS):** The set of objects currently being physically or virtually handled (touched, gripped; or selected in the virtual sense) by the agent.

**Manipulated Set (MdS):** The set of objects whose states (external as well as internal) are currently in the process of being changed by the agent.

All these spaces and sets, with the obvious exception of the SdS and the MdS, primarily provide data on what is *potentially* involved in the agent’s current activities. Cf. the virtual desktop in the WIMP interaction paradigm.

**EXAMPLE SITUATION: HAVING BREAKFAST**

This section demonstrates the use of the SSM as a tool for analysis of a given mixed-reality situation with the aim of identifying mediators suitable for HCI.

The human agent O sits down at the kitchen table (P18) in order to have breakfast. The kitchen table is fitted with a visual display (M1) in the centre of the tabletop. In his pocket he has a cellular phone (P30) and on his right ear a wireless headset (P31). A wall calendar (P28) two meters away has an embedded touch screen (M10, M11). Various software applications (V1-V10) are running on a server ready to interact with agent O through these mediators. Fig. 2 illustrates this scenario with the mediators and a few objects highlighted.

The glass of milk (P1) is in the right hand of agent O and thus a member of the SdS. It is also in the process of changing its state due to actions performed by O: It is changing its position in physical space as well as becoming emptier of milk as part of the current drinking action. Thus, the glass of milk is not only member of the SdS but also the MdS.

The kitchen table (P18) is not only clearly recognizable as a table by agent O (and therefore a member of the RS) but also regarded as examinable. This because a table’s status by and large is determined by the objects placed on it and O can easily identify most (if not all) objects on P18 and their spatial relationships. In the centre of the tabletop, P18 has an embedded visual display M1 currently providing information from the diet application V1. M1 is examinable because O is within its examination distance. V1 inherits examinability from M1.

The piece of bread P2 is within the field of view of O and close enough to be regarded as examinable: O could easily determine whether the piece of bread has cheese on it or not, or whether it is half-eaten.

With respect to spaces, the table P18 and the piece of bread P2 can be immediately manipulated by O and are thus inside AS (and PS). The display M1 and the diet application are not manipulable, having no mediator associated to them that would allow O to change their status: they remain outside AS (but within PS).

The wall calendar P28, physically made up of a visual display M10 and a touch sensitive surface M11, is too far away from O to be examinable. O cannot determine the days, weeks, months, or any reminders. O can most probably, however, determine that P28, its display M10, and the calendar app V2 make up a calendar by their visual layout and placement.



Fig. 2. Human agent O having breakfast in a mixed-reality environment.

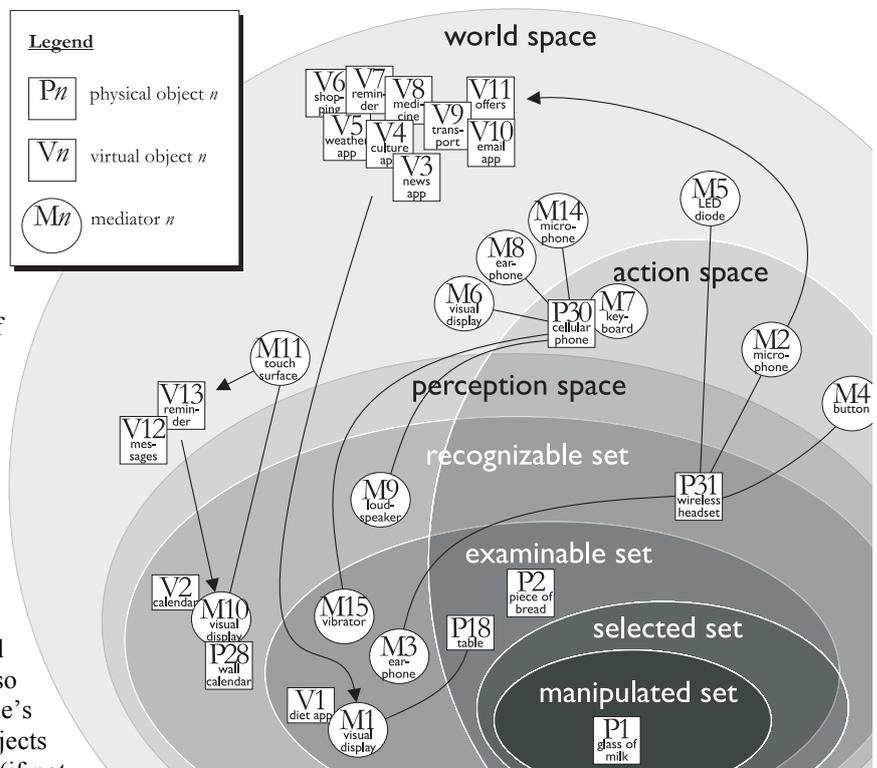


Fig. 3. The breakfast situation of Fig. 2 as viewed through the proposed situative space model. Some virtual objects (V3-V13) not visible in Fig. 2 are shown here in the world space, ready to be made accessible to agent O through mediators in the perception and action spaces. Some potential flows of interaction – specifically, manipulation of virtual objects and perception of the results – are illustrated by arrows. [5]

Thus, the calendar (P28, M10, and V2) is regarded to belong to the RS within PS. It is not within AS because O cannot from his current position change its status. The touch surface M11 belonging to P28, and the currently inactive messaging app V12 and reminder app V13 are not perceivable by O and therefore outside PS.

The wireless headset P31 carried on the ear by O is regarded as recognizable because we assume O to be acquainted to it to the extent that it can be uniquely distinguished from other headsets and earphones just by how it feels to carry it. It is not examinable because its status needs to be determined by visual inspection (of its power/connection LED M5), something not possible to do in the current situation with the LED M5 outside of O's PS. The headset's button M4 is reachable and therefore within AS. Its microphone M2 is also within AS: any utterance from O is immediately captured by it and potentially communicated to virtual applications (V3-V11) running on a server. Output from these applications is, in the current situation, either mediated visually through the tabletop display M1 or aurally through the M3 earphone embedded in the headset P31. Like with most mediators, the perceivability of M3 is decided based on the potential perceivability of an object mediated through it. Assuming that the headset is on and configured adequately, a sound or phrase originating from a virtual object (e.g. V3-V11) would be perceivable. Thus, M3 itself is placed in the perceivable space. Depending on contextual factors such as the audio noise level in the surrounding environment, virtual objects and their attributes mediated through M3 might be examinable (understood in detail) or just recognizable in the sense that agent O can identify which virtual object the information belongs to or originates from. Assuming a silent environment, M3 is member of ES.

Carried in the trouser pocket, the cellular phone P30 and its keyboard M7 is accessible to O's hands and therefore within the AS. Apart from two of the mediators embedded in it (the loudspeaker M9, the vibrator M15), the cellular phone is by and large not perceivable. Assuming that the cellular phone's position in the pocket is such that it is pressing against O's leg, we can assume that any information mediated through M15 will be perceived in detail, however limited grammar the vibration might offer. Thus, M15 belongs to the ES. M9 on the other hand, can be assumed to be a loudspeaker of limited capacity and not being able to deliver message details at its current position (in the pocket). We assume however that agent O would notice if M9 becomes active, as well as the virtual object source of the activation. Thus, it is included in the RS.

#### USE OF THE FRAMEWORK AND MODEL

The egocentric perspective on HCI has enabled us to approach a recent design task (prototyping a wearable device offering everyday activity support for people suffering early dementia) in a very different way compared to if we would have taken a device-centric approach [9]. The situative space model (SSM) has proven both to be

useful for activity recognition [10] and for providing data to a multimodal interaction manager (as pictured in Fig. 1) in an intelligent home application [9]. As real-world sensing technology matures (e.g. allowing for accurate capturing of eye-gaze, body posture, and detailed object manipulation) we believe the SSM to become an increasingly powerful conceptual tool and system component in the future.

#### CONCLUSIONS AND FUTURE WORK

We have briefly presented some characteristics of an emerging egocentric interaction paradigm and our approach in designing for it, coming up with several interaction related concepts as alternatives to more traditional ones, including the situative space model which we believe can contribute in framing mobile mixed-reality interaction both for designers and systems. Future practical efforts include the improvement of our current real-world sensor infrastructure (for details, see [5]) while theoretical work includes the definition of an ontology covering physical and virtual (in the WIMP paradigm) object manipulation.

#### REFERENCES

1. Fitzmaurice, G., Ishii, H., & Buxton, W. 1995. Bricks: Laying the foundations for graspable user interfaces. Proc. of ACM CHI'95. pp. 432-449. New York: ACM.
2. Horvitz, E., Kadie, C., Paek, T., & Hovel, D. 2003. Models of Attention in Computing and Communication: From Principles to Applications. In *Communications of the ACM*, Vol 46, No. 3, pp. 52-59.
3. Hutchins, E. 1995. *Cognition in the Wild*. MIT Press.
4. Norman, D. & Draper, S. (Eds.) 1986. *User centered system design*. Erlbaum, Hillsdale, NJ.
5. Pederson, T., Janlert, L-E., Surie, D.: "Setting the Stage for Mobile Mixed-Reality Computing - A Situative Space Model based on Human Perception". In *IEEE Pervasive Computing Magazine* (to appear).
6. Roda, C. and Thomas, J. (Eds.). 2006. Attention Aware Systems. Special Issue of Journal of Computers in Human Behaviour, Vol. 22(4): Elsevier.
7. Rosch, E. 1978. Principles of categorization. In *Cognition and categorization*. E. Rosch and B.B. Lloyd, eds., Erlbaum, Hillsdale, NJ, 1978, 27-48.
8. Suchman, L. 1987. Plans and situated actions: the problem of human machine interaction. Cambridge: Cambridge University Press.
9. Surie, D., Pederson, T., Janlert, L-E.: "The easy ADL home: A physical-virtual approach to domestic living". *Journal of Ambient Intelligence and Smart Environments*, IOS Press, (2010).
10. Surie, D., Pederson, T., Lagriffoul, F., Janlert, L.-E., & Sjölie, D. 2007. Activity Recognition using an Egocentric Perspective of Everyday Objects. In Proceedings of IFIP UIC 2007, Springer LNCS 4611, pp. 246-257.
11. Turk, M., and Robertson, G. 2000. Perceptual User Interfaces. *Communications of the ACM*, 2000.