Investigations of the Role of Gaze in Mixed-Reality Personal Computing

Thomas Pederson, Dan Witzner Hansen, and Diako Mardanbegi
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen, Denmark
{tped, witzner, dima}@itu.dk

ABSTRACT
This short paper constitutes our first investigation of how eye tracking and gaze estimation can help create better mixed-reality personal computing systems involving both physical (real world) and virtual (digital) objects. The role of gaze is discussed in the light of the situative space model (SSM) which determines the set of objects a given human agent can perceive, and act on, in any given moment in time. As a result, we propose to extend the SSM in order to better incorporate the role of gaze, and for taking advantage of emerging mobile eye tracking technology.

Author Keywords
Interaction paradigm, gaze tracking.

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

INTRODUCTION
The design of interactive systems that involve more than one computer device and also a range of everyday physical objects, demands us to extend the classical user-centered approach in HCI [3]. One challenge is that both system and human needs to continuously establish an understanding of what parts of the physical and virtual worlds that currently make up the "user interface" as devices and interaction modalities change with context. The egocentric interaction paradigm [5] proposes a change in view of a) the role of digital interactive devices in relation to the information they provide access to, and b) to generalize the HCI input/output concept to make room for multiple parallell interaction channels as well as interaction with objects in the real world (physical objects).

2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction.
February 13, 2011, Palo Alto, California, USA.

Virtual Objects and Mediators Instead of Interactive Devices
Input and output devices embedded in digital appliances are viewed as *mediators* through which virtual objects are accessed. Virtual objects are assumed to be dynamically assigned to mediators by an *interaction manager* software component residing on body-worn hardware. The purpose and function of mediators is that of expanding the *action space* and *perception space* of a human agent (Fig. 2).

Action and Perception Instead of Input and Output
In the egocentric interaction paradigm, the modeled human individual is an agent moving about in a mixed-reality environment, not a "user" interacting with a computer. Also the HCI concepts input and output are reconsidered: (device) "input" and "output" are replaced with (human agent) "action" and "perception". Note that object manipulation and perception are processes that can take place in any modality: tactile, visual, aural, etc. In this paper, we focus on *visual* modalities for perception and action.

HUMAN ACTIVITY AND GAZE
Eye movements are versatile and play an important role in everyday activities [2]. It is well known that human eye movements are governed by our interests and intentions [6], and

humans tend to look at the object that they want to act on prior to any motor control. The sequences of fixations, trackable by emerging mobile tracking technology [1] in some cases provide enough data for making predictions [2].

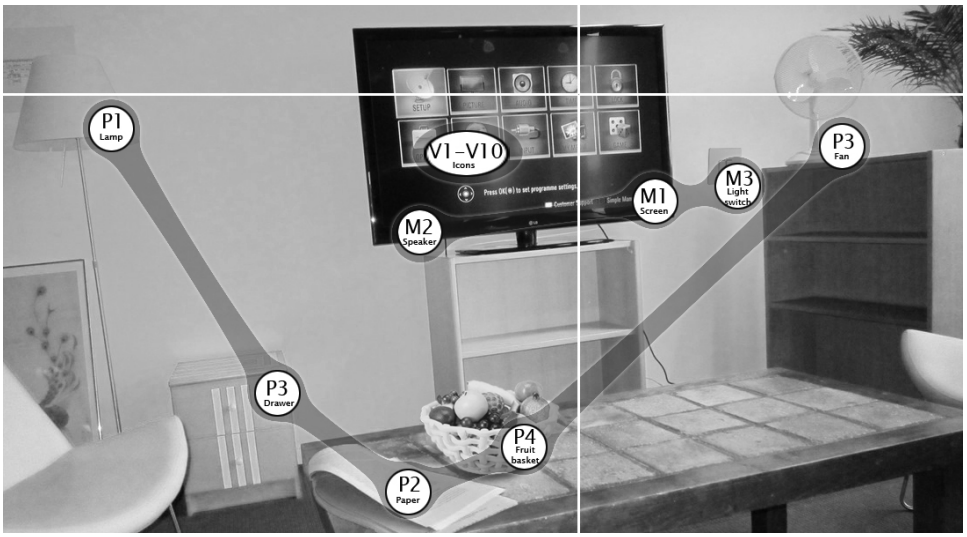

Fig. 1. A living room environment as seen by a human agent. Some physical objects (P1-P5), virtual objects (V1-V10) and mediators (M1 and M2) are labelled for illustrative purposes. The gaze direction of the human agent is indicated by the hair cross.

A SITUATIVE SPACE MODEL

The situative space model (SSM) [4] is intended to model what a specific human agent can perceive, reach and operate, at any given moment in time. This model is intended to be the emerging egocentric interaction paradigm equivalent of 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. Fig. 1. shows a typical situation which the SSM is intended to formalise and capture: a living room environment inhabited by a human agent.

In the following, we will discuss the role of gaze in the light of SSM definition excerpts from [5].

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*. [5]

Naturally, gaze direction plays a fundamental role in defining the visual PS for a given human agent. Any object directly hit by the vector anchored in the fovea and passing through the center of the lense of an eye (that is, the line of sight, LoS) is a top candidate member of the PS since it is only along this vector human agents literally see clearly. However, other components of the human visual perception system "expands" this single vector of visual impression so that visual attention in practice typically is directed to a larger area than just a point in 3D space. Let us call this 2-dimensional expanded area – with the LoS hitting its center – the field of view (FoV). Then, very simplified, the 3D space created by the union of the two eye's FoV, let us call it the 3DFoV, forms the basis for the visual PS (again, with the help of complementary parts of the human perception system, dealing with angular calculations and objects obstructing each other, etc.). All objects in the 3DFoV (not just the object in LoS) should be included in PS for a given agent.

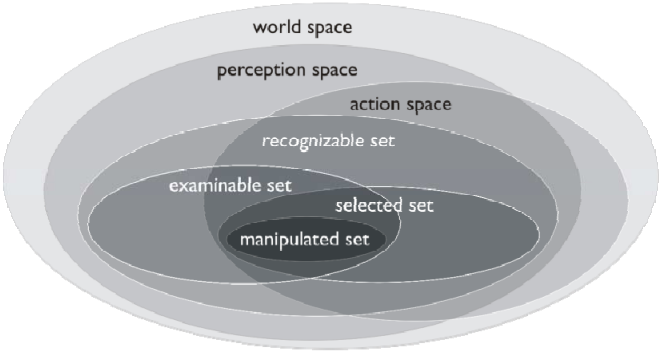

Fig. 2. A Situative Space Model. [4]

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.

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*. [5]

Examinable Set (ES)
The set of objects currently within PS that are within examination distances. [5]

The visual RS and ES in the SSM (motivated by the potential value for an egocentric interaction system to know in what detail objects can be analysed by a human agent) raises gaze tracking questions. Can gaze estimation be used for determining whether an object is examinable, recognizable or just perceivable? Eye movement pattern categorization over time and object types could, potentially, help determining whether a visually perceivable object belongs to RS or ES.

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). [5]

The visual AS is limited: Few actions that change the state of physical or virtual objects can be performed using eyes alone. However, gaze activity is often part of actions executed using other parts of the body such as the hands.

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

Physical selection is almost always preceded by visual selection: before grabbing anything, we visually fixate the object. Without dwelling into the reasons, this fact means that by tracking gaze, computer systems can do heuristical guesses for what object, among all the objects in AS, that is about to get manipulated next.

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. [5]

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 PC/WIMP interaction paradigm.

Like object selection, also object manipulation can involve gaze. While visual feedback is crucial for certain kinds of physical object manipulation (e.g. hand writing), it is probably less important for most. For manipulation of virtual objects, the situation is different. One of the most prevailing critisisms of today's user interfaces is in fact the heavy reliance on visual feedback. Contrary to actions in the real world, most user interfaces *rely* on continuous visual attention also during object manipulation.

EXAMPLE SITUATION
Fig. 1. shows a living room environment. If we assume that the area covered by the photo approximately corresponds to the field of view of a given human agent, objects in the photo can be categorized using the SSM as follows:

Physical objects
The physical object P1 (the floor lamp) belongs to the examinable set since the human agent can determine whether the lamp is on or off. The paper document P2 is not in the examinable set because from this position, the human agent can not likely determine what the document is about, see what page that is on top, let alone read the text of it. P2 is however in the recognizable set because it is indeed clear that the object is a paper document. The drawer P3 belongs to the examinable set because it is possible to see whether it is open or closed. The fruit basket P4 is examinable: it is possible to determine whether it is empty or full and even the kind of fruit that it contains. The desk fan P5 is also

examinable – it is possible to see whether its rotor blades are turning or if they are still.

Mediators

The TV embeds two mediators: The screen (M1) and the speaker (M2). The screen M1 is in the examinable set since the human agent can determine what is shown on it, i.e. the virtual objects that it currently mediates. The TV speaker M2 is not in the visual perception space at all since the case design of the TV hides its presence. (It is true that it is in the aural perception space – virtual objects can be sufficiently sonified from this distance – but we limit our analysis to the visual perception space.) The light switch M3 is in the perception space but not examinable: the human agent cannot determine its state from this distance.

Virtual objects

The icons shown on the screen M1, modeled as virtual objects V1-V10, are all examinable because their state (selected/not selected) can be determined from the position of the h. agent.

Action space

With respect to action space, most of the objects labelled in Fig. 1. are outside of that space. The human agent cannot, from her/his current positon manipulate them. The exception might be the paper document P2 or the fruit basket P4 which might be just about reachable. If we imagine the human agent to hold the TV remote control in her/his hands (a physical object embedding mediator buttons) however, also the 10 icons V1-V10 enter action space since that would allow her/him to manipulate them.

The hair cross in the picture simulates the gaze direction of the human agent, currently examining one of the 10 icons on the TV.

CONCLUSION

In this paper we have taken our initial steps in modeling gaze within the situative space model (SSM). Gaze turns out to be a defining factor for to which space an object belongs, potentially altering an object's location within the model rapidly. To fully exploit the information in eye and gaze movements, the SSM might benefit from the incorporation of something like an "attended-to" set of objects (Fig. 3.), including objects across several existing SSM spaces and sets that the given human agent is attending to. Among many open issues related to gaze and human attention is that a person may attend to objects that they can see but not recognize. At the same time, an object may be recognizable but not really attended to.

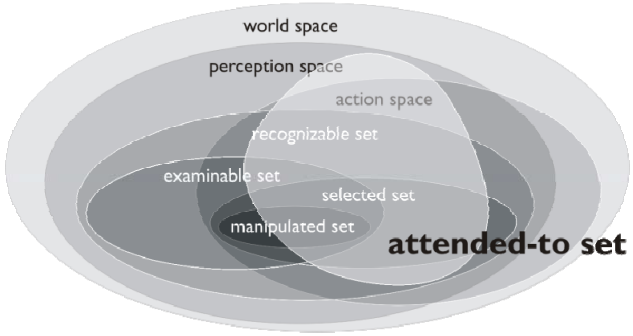

Fig. 3. Future work: extending the situative space model with an "attended-to" set.

REFERENCES

1. Hansen, D. W., Ji, Q., In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 3*, 2010,478–500.
2. Land M.F., Tatler B.W., *Looking and Acting: Vision and eye movements in natural behaviou*r. Oxford; New York: Oxford University Press, 2009.
3. Norman, D. & Draper, S. (Eds.) *User centered system design*. Erlbaum, Hillsdale, NJ, 1986.
4. Pederson, T., Janlert, L-E., Surie, D., Setting the Stage for Mobile Mixed-Reality Computing - A Situative Space Model based on Human Perception. *IEEE Pervasive Computing Magazine* (to appear), 2011.
5. Pederson, T., Janlert, L-E., Surie, D., Towards a Model for Egocentric Interaction with Physical and Virtual Objects. *Proceedings of NordiCHI'10*, ACM Press, 2010, 755-758.
6. Yarbus, A. L., *Eye Movements and Vision*. New York: Plenum Press, 1967.