Human Cognition as a Foundation for the Emerging Egocentric Interaction Paradigm

Dipak Surie, Thomas Pederson, and Lars-Erik Janlert

Abstract. This chapter presents an "egocentric interaction paradigm" (EIP) centered on human agents rather than on the notion of user. More specifically, this paradigm is based on perception, action, intention and attention capabilities and limitations of human agents. Traditional and emerging interaction paradigms are typically related to a specific computing environment, devices or human capabilities. The novelty of the proposed approach stems from aiming at developing a comprehensive and integrated theoretical approach, centered on individual human agent. Development in Human-Computer Interaction (HCI) has been closely related to the understanding and utilization of natural human skills and abilities. This work attempts to understand and model a human agent, and in particular their cognitive capabilities in facilitating HCI. The EIP is based on principles like situatedness and embodiment, the physical-virtual equity principle, and the proximity principle. A situative space model built upon our understanding of human cognition is described in detail, followed by our experience in exploring the egocentric interaction paradigm in the easy ADL home.

1 Introduction

The term Human-Computer Interaction (HCI) refers to the study of interaction between humans and computers. The scientific field of HCI has seen considerable advancements in the last 30 years, especially since the publication of Card, Moran,

Dipak Surie · Lars-Erik Janlert
Dept. of Computing Science, Umeå University, Sweden
e-mail: {dipak,lej}@cs.umu.se

Thomas Pederson
IT University of Copenhagen, Denmark
e-mail: tped@itu.dk

M. Zacarias & J.V. de Oliveira (Eds.): Human-Computer Interaction, SCI 396, pp. 349–374.
springerlink.com

and Newell's *Psychology of Human Computer Interaction* [9] in 1983. Most of the research efforts in HCI attempt to make use of human skills and abilities that are naturally developed by leading a life in this physical world.

Initial interfaces were command-line interfaces where a user sits in front of a terminal screen, and enters a line specific command to perform specific tasks and wait for a reply. Only a part of the terminal screen real estate was used. Textual interfaces were an extension of the command-line interfaces where the entire terminal screen was used. Both these type of interfaces were developed based on the typewriter metaphor. The notion of *interactive loop* and *dialogue* were introduced and the interfaces made use of human skills and abilities in linguistics.

Graphical user interfaces (GUIs) [56] were introduced with an attempt to visualize abstract computational entities as virtual objects that could be directly manipulated. The attempt was not to simply replace words by icons, but to introduce a two-dimensionally space that mirrors a simplified model of the physical world inhabited by human agents. All that is important for a human agent was assumed to happen within the two-dimensional screen. Mouse was introduced to facilitate direct manipulation of virtual objects. GUIs lead the way towards the popular WIMP (Windows, Icons, Menus and Pointers) interaction paradigm, an inherent part of desktop computing developed based on an office or desktop metaphor. Since the WIMP interaction paradigm is based on a simplified model of the physical world, it does not support the wide range of human abilities and skills exhibited by human agents in interactions within the physical world. The WIMP interaction paradigm is also restricted to office type of activities.

With the advancements in computer technology like processing power, memory, storage, communication and with the advent of novel sensors and actuators, research efforts began in pushing computers out of its desktop shell. At this juncture, there were two types of research efforts that began: 1) exploring the next-generation computer; and 2) exploring the next-generation human-computer interaction (HCI). The goal of this chapter is to address the next-generation human-computer interaction (post-WIMP interaction paradigms and approaches) by presenting a comprehensive and integrated theoretical approach, centered on individual human agent rather than on the notion of user. More specifically, the proposed egocentric interaction paradigm is based on perception, action, intention and attention capabilities and limitations of human agents. This chapter is structured as follows: section X.2 briefly describes the emerging interaction paradigms and approaches; section X.3 presents theories and architectures related to human cognition and focuses on a human agent's perception, action, intention and attention capabilities; section X.4 described the egocentric interaction paradigm in detail with its basic principles and assumptions, a situative space model that acts as a corner stone for the EIP, and an agent-centric architecture; section X.5 describes our experience in exploring the EIP within a smart home named as the easy ADL home; and section X.6 presents the conclusion.

2 Post-WIMP Interaction Paradigms and Approaches

According to the concept of reality-virtuality (RV) continuum [44], there exists a continuous scale between completely physical environments populated with purely physical objects to completely virtual environments populated with purely virtual objects. The environments between these two extremes are populated with both physical and virtual objects with varying levels of augmentation. Mixed-reality environments are in the middle of the RV continuum containing both physical and virtual objects with equal augmentation. Physical objects for instance usually provide rich affordance and can be manipulated using natural human skills, while virtual objects being light-weight could for instance be instantly transported across large distances with ease. Mixed-reality environments attempt to reap the best of both the physical and the virtual world.

Exploring post-WIMP interaction paradigms fall between the reality-virtuality continuum. At one end of the continuum, a human agent is fully immersed into a computer-simulated artificial environment researched in the field of Virtual Reality [7]; and at the other end, a human agent is fully immersed within a physical environment augmented with many small, invisible computers embedded within physical objects researched in the field of ubiquitous computing [68]. Even though different computing environments introduce different HCI-related challenges, there are some HCI challenges that are more fundamental, centered on a human agent and probably are applicable to different computing environments. The egocentric interaction paradigm to be described in this chapter attempts to address such fundamental HCI challenges allowing room for complementary and more environment and device-specific research efforts to exist in parallel.

Research efforts in exploring post-WIMP HCI began by focusing on various human capabilities and skills commonly observable in the physical world. Tangible User Interfaces (TUIs) [29] allows a human agent to directly interact with physical artifacts, instead of indirectly manipulating virtual objects using a mouse. The human agent is not restricted to two-dimensional virtual object manipulations, but could use their entire body in performing activities using physical artifacts (representing abstract computational entities) arranged in a three-dimensional space. TUIs are based on human abilities to grab and manipulate everyday objects in the physical world with ease.

Human agents communicate with other human agents using several modalities, often in parallel and in a natural manner. Modalities like speech, hand gestures, facial expressions, vocal emotions, etc. are more natural to a human agent in comparison to modalities like mouse and keyboard. Taking inspiration from such human-human interaction in the physical world, research efforts have focused on using gestures, hand posture and gaze [37]; facial expressions [17], vocal emotions [49], speech, and multiple modalities [31] in facilitating human interaction with digital information. Human emotions are used in affective user interfaces [28].

Perceptual user interfaces (PUIs) [66] are based on a human agent's ability to perceive information. Even though human perception is common of many of the earlier mentioned HCI models, PUIs consider human expectations, limitations and possibilities of human perception and the semantic nature of human interactions. PUIs take advantage of both human and machine capabilities to sense, perceive and reason. The WIMP interaction paradigm assumes that a human agent spends all their attention on interaction with the desktop computer. However, human agents in a physical environment tend to attend to objects and events with varied levels of attention. This is because, human attention capabilities are limited in nature and this resource requires careful management. Attentive user interfaces (AUIs) [67] refer to computing interfaces that are sensitive to a human agent's attention. AUIs attempt to measure and model human attention and use this information to structure their communication with the human agent.

Human agents interact within physical environments by considering their surrounding context, i.e. information that characterizes the human agents' situations. The WIMP interaction paradigm ignores a human agent's contextual information, assuming their situation to be always constant. However, such an assumption is not suitable for post-desktop computing environments. Context-aware computing [55] attempt to sense, recognize and infer contextual information, and use such information to enhance the human agent's interaction with computing systems. Context includes a human agent's location, activity, proximity to objects, etc. Human agents communicate rich amount of information implicitly in their interaction with other human agents. Contextual information could be considered as implicit input, i.e. input that is not directly intended to be provided to a computing system, thereby increasing the bandwidth of communication between humans and computers.

Human agents' interaction within a physical environment is largely driven by their activities. The WIMP interaction paradigm is designed based on an application-centered or document-centered model suitable for specific tasks, but does not provide support for higher-level activities. Activity-based computing (ABC) [2] consider human activities as a starting point in facilitating human interaction with computing systems. ABC attempts to provide computational support at the activity level by aggregating resources required for individual activities. ABC supports parallel and interleaved activities, and their activity support is context-aware.

Human agents are embodied beings in nature, i.e. they have a physical body and are manifested in and of the everyday world. Embodiment is beyond physical embodiment. Human agents perform activities within a setting where the setting is not a mere background, but a fundamental and constituent component of the activities providing meaning and value to their activities. Embodied interaction [12] is based on the fundamental principle of embodiment, and considers the physical and social aspects of interaction.

Reality-based interaction (RBI) [30] attempts to provide a conceptual framework for the emerging trends in HCI. In particular, RBI attempts to focus on the following four themes relating to human agents: 1) Human agents possess common sense knowledge about the physical world in which they live; 2) Human agents are aware of their body and possess skills to control and coordinate it in the physical world; 3) Human agents are aware of their surroundings and possess skills for negotiating, manipulating and navigating within their environment; and 4) Human agents are aware of other human agents in their environment and possess social skills for interacting with them. Reality-based interaction is quite general and perhaps somewhat vague on details.

Egocentric Interaction Paradigm (EIP) attempts to unify existing research findings relating to a human agent into a single theoretical framework with the possibility to evaluate and refine it. This is in contrast to similar efforts like in tangible user interfaces, perceptual user interfaces, attentive user interfaces, speech user interfaces, gesture user interfaces, etc. that attempts to focus on specific micro-aspects of a human agent. We are aware of the need to exchange research finding between the two contrasting styles of exploring suitable post-WIMP interaction paradigms. Also, the EIP addresses issues and challenges at the interaction level, and needs to be complemented by research efforts at the interface level. RBI and EI have a broader scope in comparison to the EIP, but the EIP is sharper and shares concrete models like the situative space model to be described in section.3.

EIP considers the understanding of a human agent, their cognition, capabilities and limitations from a cognitive science perspective and use it as a starting point for facilitating human interaction with computing systems. EIP is not technologically driven, however technological advancements are important both to improve the understanding of human agents, and also to use it for facilitating HCI. By taking a human-centered perspective on HCI, EIP is expected to withstand the series of technological advancements over time. EIP is based on a human agent's perception and action capabilities, considered to be a starting point for a human agent's behavior in the physical world. In this sense, the EIP is a by-product of basic research on human agents and is valid along the entire RV continuum.

3 Human Cognition

Understanding human cognition plays an important role in facilitating computer-mediated human activities. This introduces a need to develop models of human cognition for a domain like HCI. Such models are useful for two reasons: 1) to utilize the wide range of human cognitive capabilities and to support for their limitations; and 2) to create artificial agents that complement human agents.

3.1 Related Theories

According to the theory of Situated Cognition [6], human cognition cannot be separated from their context. Here the term context could be physical, activity, social, or cultural. Human thoughts and actions occur *in situ*, i.e. adapted to their situation. Human actions to a large extent are improvisatory and are executed by directly coupling perception and action. Situated cognition theory could be considered as a complement to the traditional theories in cognition weighting upon the inferential processes of deliberation and planning. Situated cognition assumes that human knowledge has a dynamic aspect to it, and is formed based on a human agent's interaction within an environment. Human knowledge is constructed in performing actions instead of viewing it as a stored artifact. Human agents are social beings by nature, whose knowledge is created by their active participation in this world and the meaning is created by their ability to actively participate in a meaningful way.

Embodied Cognition (EC) [70], similar to the theory of situated cognition emphasizes that cognition is more a situated activity that is actively constructed. The importance of environment and its role in the development of cognitive processes is also acknowledged by embodied cognition. According to EC, embodiment is a necessary condition for cognition, and it could be understood as a human agent's perception-action capabilities that enable them to interact and survive in an environment. According to [10], human cognition cannot be viewed separately from their body and the surrounding environment. Human agents are embodied beings with a body that constraints the way the world appear to them and the way they interact in it. The world would not appear the same to all human agents and is dependent on the human agent's perspective, which we call their *egocentric perspective* and their interaction in this world as *egocentric interaction*. A human agent's egocentric perspective is dependent on their current activity, perception and action possibilities, and the environment in which the human agent is situated in. The point that human cognition is grounded on their physical and social characteristics and their embodied experience in the world differentiates embodied cognition from research in situated cognition, even though the two theories are closely related and complementary in nature.

According to the theory of Distributed Cognition (DC) [23], human cognition is not confined within an individual but is distributed among objects, tools and human agents in an environment. DC extends the boundary of analyzing cognition from an individual human agent to the environment as a whole, making it an interesting theory to consider in facilitating HCI within ambient intelligent environments. Since cognition is distributed, it may be distributed among a group of human agents resulting in emergent social interactions that are important to consider in understanding cognition. It is discussed as collective intelligence in [41]. Similarly, cognition could be distributed among physical objects, resulting in the "offloading" of information onto the environment and thereby creating a medium for social collaboration. Also, such objects in the environment could be seen as an extension of a human agent's mind, playing an active role in driving the human agent's cognitive process described by the concept of the extended mind [11]. Similar to embodied cognition, DC views the embodied nature of cognition, thereby considering the human body and the material environment around them to take a central role in understanding cognition. Human agents live in complex cultural environments resulting in strong relationship between culture and cognition. Human activities in the historic past has shaped human cultures of the present, which influences the way human agents think and act leading to the emergence of future human cultures.

Activity Theory [39] is widely used as an analytical and evaluation framework in HCI on aspects relevant to EIP such as regarding activities as their unit of analysis, regarding the motives and needs of human agents, as well as the psychological and social contexts of their activities. It also shares premises such as shifting from the notion of user to the notion of agency, and the situatedness of human actions. Furthermore, activity theory could provide valuable insights into extending EIP by including social and cultural aspects, which is not addressed in this chapter and is left for future work. According to Activity Theory [39], a human agent's activity could be modeled using a hierarchy of activity, actions and operations wherein the hierarchy changes dynamically according to the context. Human activities have an objective and could be split into goal directed actions that are achieved by low-level operations.

According to ecological psychology [20], understanding human cognition involves studying a human agent in their everyday environment. Human agents do not exist or act in isolation. Instead, they adapt themselves to establish a harmonious working relationship with their environment. Ecological psychology emphasizes the influences of a human agent's situation on their behavior similar to earlier mentioned theories like situated cognition and embodied cognition. To summarize, most of the above mentioned theories on understanding human cognition emphasizes the following: 1) the importance of a human agent in their everyday environment; 2) the importance of a human agent's situation and embodiment; and 3) the importance of viewing a human agent's environment as a whole including the social and the cultural aspects integrated within it.

3.2 Cognitive Architectures

Cognitive architectures attempt to describe the structure of human cognition in a comprehensive manner. The term was coined by Newell based on his work on computer architecture [3]. Cognitive architectures in general are not developed to address existing problems within HCI, but to establish a wide range of theoretical concepts about human intelligence, their abilities and limitations. Research findings in cognitive psychology usually drive cognitive architectures. Traditionally, human cognition was viewed to be independent of a human body and its surrounding environment. According to the Model Human Processor [9], the human cognitive processor is composed of three modules namely perception, cognition and action. It views cognitive processing to be a cyclic and sequential process from stimulus perception to cognitive problem solving to response execution. Traditionally, both cognitive psychology and artificial intelligence have focused on the cognitive problem solving and performance aspects of the human agent. However, human perceptual and motor systems are complex in their own rights needing appropriate research attention. Also, it is likely that human cognition is constrained by human perception and motor capabilities, and the vice-versa. However, modern view of understanding human cognition focuses on the perceptual and the motor aspects of a human agent. This is in line with the above mentioned cognitive science theories that stress the need to consider human agents as part of a complex system that includes their body and the surrounding environment.

SOAR [47] and ACT-R [1] incorporated some aspects of human perception and motor action within their cognitive architecture. ACT-R is based on a serial model of perception, cognition and action, mainly suitable for human interaction with single computer displays and action possibilities that are limited to using a mouse and a keyboard. ICARUS [40] considers cognition to occur in physical contexts strongly grounded by human perception and action, there by giving importance to physical interaction. ICARUS describes cognitive behavior to occur in recognize-act cycles where a set of perceived objects including their type, names and description are deposited in the perceptual buffer for matching against long-term conceptual definitions before executing actions. ICARUS stores knowledge in the form of concepts that describe environmental situations, and skills that describe the mechanism to achieve goals by decomposing them into sub goals. According to [38], human concepts in general are relational in nature and are described by their interactions among objects and events. ICARUS considers such a relational view of objects in describing concepts grounded in perception and skills grounded in executable actions.

EPIC [35] is another cognitive architecture that contains a comprehensive set of perceptual-motor mechanisms, but considers the cognitive mechanisms to play a peripheral role. EPIC is based on parallel processing of perception, cognition and action. EPIC is based on a simple production rule system that contains a set of production rules that translate preconditions into actions to attain certain post-conditions. Since EPIC does not handle conflict resolution, it allows all the productions that meet the preconditions to execute actions in parallel which is a limitation. ACT-R/PM [8] is an extension of ACT-R that considers the perceptual and motor mechanisms to be an integral part of its cognitive architecture, giving them an equal status with the cognitive aspects. ACT-R/PM comprises of features, especially its parallelism that allows for cognition in dynamic environments. ACT-R/PM is composed of a set of perceptual-motor modules that operate in parallel, and can handle situations that demand for high-performance. However, ACT-R/PM is restricted to accurately modeling human latency delays. HiTEC [22] is a cognitive architecture that includes not only perceptual and action-related information processing but also the interplay between perception and action. This allows for covering some known phenomena in cognitive psychology like Simon effect [57], action planning affects perceptual process [16] and the perceptual effect of action influences action execution [24] to be discussed in section 2.3.

Human emotions, their attitudes, and motivation play an important role in human actions. The importance of human affective experiences is considered in some cognitive architecture like CogAff [58] that attempts to support interaction between cognition and affect; and Emile [21] which considers an agent's emotions in its action selection, while infers other agents' emotions in selecting its dialogue choices. To summarize, modern cognitive architectures have acknowledged the importance of a human agent's perception and action possibilities in modeling human cognition. Further, the interplay between perception and action is also considered important in building a cognitive architecture.

3.3 Perception and Action

Human perception is a process that enables human agents to interpret and organize sensation to produce meaningful experience of the world [42]. Here, sensation refers to the unprocessed result of stimulation of sensory receptors in eyes, ears, nose, tongue, or skin. Many research efforts focus on visual perception for understanding human perception as a whole since 70% of human sensory receptors are visual receptors and vision engages 50% of human cortex. Human action is a process that enables human agents to execute their goals in the world.

As mentioned earlier, human perception and action are closely coupled. According to the theory of event coding [25], both the perceptual and the action processes share a common representational medium with bi-directional interaction between them. According to [57], Simon effect describes human tendency to react towards the location of the stimulus in performing response action even if the location cue is irrelevant to that action. According to [16], action planning can affect perceptual processes, which challenges the traditional view of sequential information flow from stimulus perception to action planning and execution. According to [16], action affects perception by creating a bias on the objects that are perceivable depending on their relationship to the ongoing or intended action. According to [25], object representation and action plan representation overlap to an extend depending on the features that are common between objects and action plans. Even in the absence of action execution, subjects scanned using fMRI showed highest activation in the premotor areas (areas involved during action) of the brain during the anticipation of perceptual events, thereby indicating it to be an action-relevant stimulus [54].

Human agents are capable of learning the perceptual effects of actions and use this knowledge in action execution [24]. Further evidence of it was found in a brain-imaging study [15]. According to the "ideomotor principle" [32], human agents build up the desired effects of an action in mind before executing the goal-directed action. This also suggests that actions are goal-directed to an extent instead of considering them as mere responses to stimulus perception. Otherwise, it would also be difficult for a human agent to know if the performed action has actually satisfied their goal or not; and a stimulus-response agent that behaves purely based on perception might not be capable of making "intelligent" decisions which require higher-level cognition. Human agents realize their goals by actually performing actions in varied contextual conditions.

However, human actions are not fully pre-specified before the actual execution of the action. Human environment and the object upon which actions are directed play an important role in specifying certain parameters for the execution of an action. According to Gibson's concept of affordance [20], an environment or an object affords the possibility of performing certain actions while not allowing for other actions in relation to the actor. Thus, action representations are not completely internal within a human agent and some of the representations are actually distributed in the environment. Predictable action plans are usually internal and are complemented by environmental cues for a human agent to execute actions in varied contextual conditions. Apart from the physical effects of action, there is also an affective aspect to action. According to [65], actions yielding positive affective consequences are more likely to be executed in comparison to those with negative affective consequences.

For a human agent to achieve their goals by executing actions, the human agent should understand their current situation. Human agents not only perceive objects and events that are part of a situation, but also recognize them. Recognition is closely related to categorization and could be considered as the process of mapping objects, events, etc. to concepts or categories already known to a human agent. Recognition is in some sense being able to categorize objects and events into their respective type. Recognition is also closely related to perception in the sense that the output from perception is usually considered as an input for recognition, however some cognitive systems view them alike. Based on the recognition of objects and events, human agents could potentially examine their situation. The process of examination includes determining the states of objects and events; and in the formation of new concepts by conscious observation and learning.

Recognition and examination of the situation, along with the perceptual effects of action in that situation influences a human agent's decision to execute an action among alternative actions. Human agents execute action in an environment by selecting the objects that could potentially be acted-upon considering the environmental conditions and the objects' relevance to their action goals. The selection process is followed by the manipulation process where the objects' states are modified, even though some cognitive systems might view the two processes alike. The situative space model to be described in section X.4 includes processes like perception, recognition, examination, action, selection, and manipulation.

According to [20], human agents do not perceive and act in the entire world, but frame the world into an environment where they can survive through perception and action, constrained by their body. Human perception and action capabilities are framed by their body, which is used as a started point within the egocentric interaction paradigm. Human agents use multiple modalities to sense and act in an environment. Perception could be considered as an integration of the results from different modalities into a unified description of the environmental situation. Similarly a unified description of an action could be executed through several human modalities. Some models do not view perception and action as separate entities unlike the situative space model to be described in section X.4 where the division is considered pragmatic.

3.4 Attention and Intention

Human agents have limited processing resources for perceiving the world and acting within it. Human attention comes in as a filter that preferentially allocates processing resources to objects and events that are important for perception and action depending on the human agent's intentions and the activity context. According to some theories [13], human visual perception can take place without attention by extracting perceptual primitives from the environment. Experimental evidence of grouping under conditions of inattention was shown in [45]. However, according to [43] there is limited perceptual organization under conditions of inattention.

Divided or peripheral attention in human agents allows them to be aware of stimuli within their visual field without focusing their attention on the stimuli. Such stimuli might be distracting stimuli as in a visual-search scenario, or the stimuli might be secondary task relevant stimuli in a dual-task scenario. Full-attention or central attention in human agents allows them to be aware of stimuli within their visual field by focusing their attention on the stimuli. Such stimuli might be primary task relevant stimuli in a dual-task scenario.

Allocation of human attention to a large extent is dependent on their current action [71]. Within an action context, human attention considers a top-down approach in search of objects and events relevant to their action [26]. Outside the action context, human attention considers a bottom-up approach driven by the features of objects in the environment [50]. Thus, human attention could be modeled at two levels: 1) low-level object recognition; and 2) high-level human intention.

Human intention refers to a human agent's mental determination to act in a certain way. Human agents possess the ability to infer other human agent's intentions by observing their actions [4]. According to the simulation theory [18], a human agent A observing another human agent B's action simulates the observed action as their own action and estimate the intentions for that action. Others consider intention recognition to be an inferential process by applying a "theory of mind" [5]. According to [34], action simulation and inferential mechanisms are complementary in understanding action intentions. Irrespective of the actual mechanisms behind understanding action intentions which has created a lot of debate in the last few years, human intentions play an important influential role in human perception, action and allocating attention.

4 Egocentric Interaction Paradigm

Egocentric Interaction Paradigm is a human-centered interaction paradigm based on a human agent's perception, action, attention and intention capabilities (and limitations). The term "egocentric" has been chosen to signal that it is the human body and mind of a specific human individual that acts as centre of reference to which all interaction modeling is anchored. In the context of this article, the term should not be taken to refer to selfishness or other personality traits but to the basic human condition of dealing with the world based on their own senses, actions and cognitive abilities, even when working in groups and with shared goals.

4.1 Basic Principles and Assumptions

Within the egocentric interaction paradigm (EIP), the term "human agent" is used instead of the term "user" for several reasons. The EIP models interaction occurring between a specific human being and potentially several interactive devices at the same time, as well as the interaction between the human being and the physical environment. In such a wide-ranging variety of interactive situations, it is more natural to regard the modeled individual as an *agent* in a physical-virtual environment (created by augmenting a physical environment with virtual objects) rather than a *user* of it. This shift in terminology, from "user" to "human agent" is also a consequence of letting go of "the application" as the tacit focus: applications are "used" by someone and the "user" is defined by the artifact, whereas human agents do actions and are engaged in activities that typically involve a number of objects.

4.1.1 Different from Classical User-Centered Approaches

The EIP is very much centered on the human agent not only in the sense that interaction design should bow to the human agent's needs and preferences, should ultimately be a servant of the human agent, but in the sense that the human agent should be understood as being in the (moving) center of the world. This distinguishes the egocentric interaction paradigm from earlier "user-centered" approaches in HCI, such as, user-centered design [48], which largely ignored a human agent's current bodily situation (because it would have been fixed by the stationary computer system anyway), the egocentric interaction paradigm acknowledges the primacy of the human agent's current bodily situation in the environment at each point in time in guiding and constraining the human agent's behavior (which is precisely what "egocentric" refers to). It assumes that the whole environment is taken into consideration, not just a single targeted artifact or system. It makes the assumption that proximity plays a fundamental role in determining what can be done, what events signify, and what the human agent is up to. It seriously considers mobility and takes into account the human agent's more or less constant movements of head, arms, hands, and body, locally and through the environment, as well as the human agent's constant rearrangements and modifications of various parts of the environment. It is neither mainly oriented towards the interaction with "virtual," immaterial data objects (classical HCI), nor predominantly towards the interaction with physical objects and machines (classical ergonomics and HMI), but pays equal attention to virtual and physical objects, circumstances, and human agents, and their interrelations. It recognizes that the human agent typically will have multiple ongoing activities at the same time, some of which may have little relation to each other, and that activities are started, put on hold, interrupted, resumed, and finished in a never ceasing flow. It makes no pretense and no tacit assumption that the system or the human agent has full information of the situation, because situations are in principle open-ended and ever changing. It should be noted that there were and still are user-centered approaches to interaction design that do take some (but never quite all) of these factors into account, a prime example being participatory design [14]. However, these approaches are focused on methods and ideals for the design and development *process*, and on basically work-oriented and fixed, well-defined and stable tailor-made systems. Circumstances have changed.

We also note that whereas interaction paradigms that are centered on interaction devices will have to change frequently to keep pace with new technological advancements, interaction paradigms that are human-centric like the egocentric interaction paradigm are more stable over time: basic human characteristics change only very slowly.

4.1.2 Situatedness: The View from Here

The following is a commonsensical summary of what we take to be the essence of the *situatedness* of human agents with regard to interactions and activities. It is obvious that human beings have a self-centered perspective on their environment before any other perspective, such as second person perspective or some very-hard-to-acquire unlocated subspecie-aeternitatis perspective that science aspires to ("the view from nowhere" [46]). The world is so large and so rich in details that any agent with limited cognitive capacity must necessarily narrow its focus in some manner. Human beings have physical bodies that are located at a single particular place and oriented together with their limbs and sense organs in particular directions at any particular time. That gives them a natural primary vantage point for selecting which details and aspects to attend to: *the view from here*, what a particular embodied agent can perceive given its current bodily situation in the world. It also gives them a focus for actions: anything they can do is limited by what is within their reach given their current bodily situation in the world.

To control your actions and the effects of your actions it is particularly helpful to be able to perceive the part of the environment in which you are acting and affecting. Vice-versa, to perceive, to pick up information from a certain part of the environment it is helpful to be able to do actions that change your angle of perception and the parts you currently perceive because that can assist and improve the information extraction. *Natural* agents are usually concerned about aspects and details of the world that have relevance for themselves, and naturally perform actions that are relevant for their own existence and relation to the environment.

Human agents are where they are, and that determines what they can do. If they want to do something that cannot be done right here right now, they will have to do something about it, and they will always have to do *that* starting from their very situation here and now. By doing things with what is within their reach, by moving themselves in the world, and by moving other objects in the world, they can change what they can do. The further removed from the current situation what they want to do is, the longer chains of action they need to be able to contemplate and follow to reach their goal. To their advantage, at each point in a long chain of actions, their current situation in the world, what they currently perceive, can help them recall and guide their next action; they may also be able to specifically arrange the environment around them and their relation to it so that their situational guidance through the process is improved. This being so, it also follows that the trajectory of a human body through space-time and the trail of environmental events and changes in its close vicinity may reveal a lot about which goals and intentions are on this human's mind, and which activities are being carried out. *Epistemic operations.*, i.e., physical operations performed to facilitate cognition rather than to further physical progress towards some external goal [36], may offer particularly transparent access into the mind of the human agent. Epistemic operations make cognition transparent by instantiating (partially) external cognition in the sense of the extended-mind hypothesis [11]. They are transparent also in the practical sense of being within the scope of current tracking technology, just as other physical operations are.

The view-from-here perspective means that all information about the environment that is picked up by the human agent derives from information converging on the body and modulated by the point of view of the body wherever it is situated; and that all actions, all effects on the environment produced by the human agent have their starting point in the body and are modulated by the angle of approach of the body wherever it is situated. Most human activities involve the handling of a number of objects and actions to make them available (locomotion, object transportation) in certain sequences and patterns that are appropriate for the objective of the activity. Human agents involved in a particular activity are at each moment guided in their actions by the array of objects in their immediate proximity, at the same time as some of their actions serve to or has as a side effect to change the array of objects in their immediate proximity. We may count on a general human tendency to minimize the effort spent, physical as well as cognitive. At any point in time, objects that are close are likely to be or become relevant for an ongoing or near future activity and objects that are relevant for an ongoing activity are likely to be or become close.

4.1.3 The Physical-Virtual Equity Principle

The EIP differs from traditional interaction paradigms in explicitly ignoring the input and output devices of interactive computers such as keyboards and displays of PCs and cellular phones, considering them as more or less transparent *mediators* for accessing virtual objects. Mediators include sensors, actuators, input devices, output devices, user interface software, context recognizer, etc., that act as tools for providing human agents the access to virtual objects. Taking such a stance permits the modeling of physical and virtual objects as if they were situated in the same *physical space*, which is advantageous when modeling applications for ambient intelligent environments where the interaction complexity vastly surpasses what can be sufficiently described using a classical human-computer interaction dialogue model. We do recognize the inherent differences between physical and virtual objects [51]: our goal is *not* to make them resemble each other as much as possible. Designing virtual objects and their environments as an exact copy of the physical objects and their environments removes the inherent advantages of being virtual (or digital). For instance, it is easier and cheaper to transport a virtual object across space compared to transporting physical objects. Within the EIP, we propose an approach where physical objects and virtual objects are co-located and complement each other with their inherent properties. Traditional interaction paradigms associate virtual objects to specific devices, while within EIP such an association is more dynamic and is often avoided in modeling human–environment interaction. The point with physical-virtual equity is to handle physical and virtual objects uniformly on a high level of abstraction in order to enable better modeling of mixed-reality situations. Furthermore, such a view aligns well with findings in psychology indicating that expert users of tools (whether a tennis racket or a computing device) tend to focus on the domain object they are working on (whether it is a tennis ball or an email) and become less aware of the details of the tool (such as the handle of the tennis racket or the workings of a particular input device) itself.

4.1.4 Perception and Action Instead of Input and Output

One of the strengths of the direct manipulation mechanism [56] widely deployed in user interfaces for changing the state of *virtual* objects, is the fact that it makes the strong relationship between manipulation and perception prevail also in many parts of the virtual world. The classical HCI concepts of input and output need to be substituted with something that works both for physical and virtual object manipulation. Also, the concepts of input and output are device-centric and refer to device interfaces, while the EIP needs to use concepts that refer to a human agent's interface to the environment (i.e., a human *agent's* input and output). Hence, the concepts of perception and action replace the traditional concepts of input and output. Action is often inseparable from or intertwined with perception in physical everyday activities. One cannot change the state of an object (e.g., open a refrigerator door) without perceiving feedback (e.g., tactile feedback while opening the refrigerator door). Conversely, perception needs support of action (e.g., to see what is in the refrigerator, the door must be opened). This tight coupling between action and perception is the result of the way objects are designed, how we manipulate them, and how the manipulation process interplays with the laws of physics. The traditional concepts of input and output, stemming from a time when interaction typically was in terms of exchanges of language expressions, still have a bias towards a turn-taking approach where first an input is provided to a device that processes it and provides an output, whereas perception and action are inter-coupled and take place in parallel. One activity may typically involve several distinct sources (for perception) and targets (for action) at the same time, and multiple activities are often going on in parallel, more or less independent of each other.

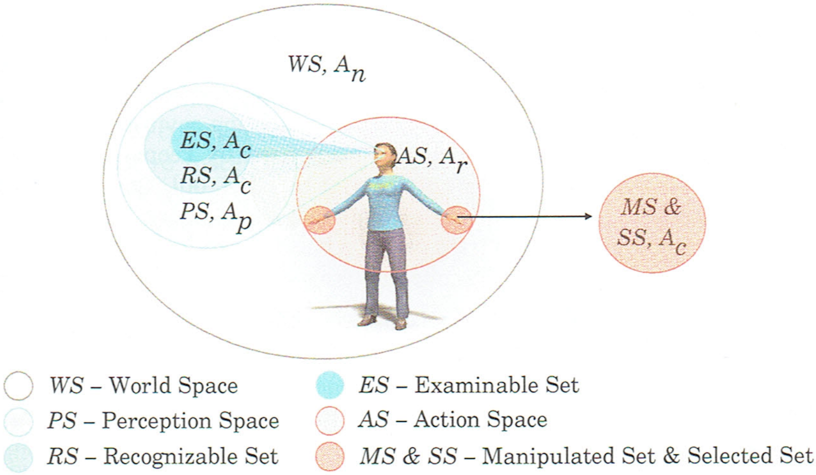
4.2 Situative Space Model

The situative space model (SSM) [52] is intended to capture what a specific human agent can perceive and not perceive, affect and not affect at a 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, and Pointing device) interaction paradigm: more or less everything of interest to a specific human agent is assumed to, and supposed to, happen here.

The situative space model is inspired by the proximity principle which makes the assumption that proximity plays a fundamental role in determining what can be done, what events signify, and what human agents are up to. The proximity principle [33] states that things that are close tend to matter; and things that matter tend to be(come) close. Objects (and other human agents, circumstances) that are close to the human agent tend to matter in the sense that they have a fair chance of getting the agent's attention and figure in the human agent's current cognitive processes and activities.

The proposed SSM is human agent-centered and varies continuously with the human agent's movement of body and body parts. Operational definition of the components of this model is as follows (also refer to Fig. 1):

- *World Space (WS):* A space containing the set of all physical and virtual objects to be part of a specific situative space model.
- *Perception Space (PS):* The part of the space around the human agent that can be perceived at each moment. Perception Space can be given a simple geometrical interpretation like a cone in the case of vision (refer to Fig. 1) as a rough approximation. Objects may occlude other objects and thus create (temporary) holes in the space.
- *Recognizable Set (RS):* The set of objects currently within Perception Space that are within their recognition distances, i.e. it is possible for a human agent to recognize what type of objects they are.
- *Examinable Set (ES):* The set of objects currently within Perception Space that are within their examination distances, i.e. it is possible for the human agent to not only recognize the objects but also their states. Normally, we expect the Examinable Set to be a proper subset of the Recognizable Set.
- *Action Space (AS):* The part of the space around the human agent that is currently accessible through the human agent's actions. Objects within this space can be directly acted on. The outer range limit is basically determined by the physical or virtual reach of the human agent, but obviously depends qualitatively also on the type of action and the properties of objects involved; e.g., a reachable object may be too heavy to handle with outstretched arms. Since many actions require perception to be efficient or even effective at all, Action Space is qualitatively affected also by the current shape of Perception Space.
- *Selected Set (SS):* The set of objects currently being physically or virtually handled (touched, gripped; or selected in the virtual sense) by the human agent.
- *Manipulated Set (MS):* The set of objects whose states (external as well as internal) are currently in the process of being changed by the human agent. Normally, we expect the Manipulated Set to be a subset of the Selected Set.



No Attention $A_n = WS \setminus (PS \cup AS)$

Peripheral Attention $A_p = PS \setminus (RS \cup ES)$

Central Attention $A_c = (RS \cup ES) \cup (SS \cup MS)$

Demand Agent Response $A_r = AS \setminus (SS \cup MS)$

Outside Activity and Action Context $A_{out}a_{out} = A_n$

Inside Activity Context, but outside Action Context $A_{in}a_{out} = (A_p \cup A_r)$

Inside Activity and Action Context $A_{in}a_{in} = (A_c \cup A_r)$

Fig. 1 A geometrical representation of the situative space model, and its association to human attention and intention (in terms of activity and action context) represented in set theory notations.

Human agents perceive and act not based on a single modality but through multiple modalities that interact with each other to create a unified representation of perception and action. This introduces a need to let the situative space model span across multiple modalities, but also to operationalize mechanisms for multimodal fusion.

4.3 Associating the Situative Space Model to Human Intention and Attention

Associating the situative space model to human intention and attention (refer to Fig. 1) is useful in facilitating a human agent's interaction with post-desktop computing systems. Keeping track of all the objects (physical objects and computer-generated virtual objects) within a human agent's perception and action space can be cognitively expensive for a human agent. We assume that a human agent scans the perception and action space to find objects that are relevant for their intentions.

The term intention might have several interpretations, but we focus on human intentions that are possible to infer from recognition of the accompanying activities. An important feature of the egocentric interaction paradigm is to unobtrusively sense and infer human intentions (activity and action context) and adapt a human agent's interaction experience with post-desktop computing environments. Human activities and actions can be modeled and recognized by keeping track of the changes in the content of the situative space over time [64, 60], while their operations could be tracked by capturing the events that occur within the selected set and the manipulated set [59] which are part of a human agent's action space.

Since the situative spaces are described in terms of physical and virtual objects, relationships between individual objects and their association to individual human activities and actions can be established. Such an association can be achieved in at least two ways. One approach is to empirically collect data by observing human activities and their relationship to objects. Activity recognition systems based on a human agent's interaction with objects usually possess techniques for automatically collecting empirical data about an object's association to individual activities. Another approach is more a top-down approach based on the artifact designers' assumption of the activities for which individual objects are designed. With the initiatives like the Internet of Things [19], everyday object manufacturers (linked to physical objects) and computing application developers (linked to virtual objects) are more likely to maintain online information about their manufactured or developed objects' association to human activities and their usage context. Challenges like standardization, maintaining up-to-date information, etc. are to be solved as part of this approach.

By modeling and recognizing human activities and actions, one could distinguish intentional operations from unintentional operations (which could be regarded as noise) that are common in real-world scenarios. Also the higher-level knowledge about the human agent's current intentions enable filtering out activity irrelevant virtual objects from being present within the human agent's perception and action spaces, to reduce their attention and cognitive requirements. Such higher-level knowledge could also be considered as implicit input [53] to improve a human agent's communication bandwidth with computing applications.

Human attention is an important and scarce resource. Post-desktop model of computing might include many computers trying to interact with a human agent in parallel as with Mark Weiser's vision of ubiquitous computing [68]. In such situations it is important not to overload a human agent's attentive and cognitive capabilities. Human agents use different levels of attention within the perception and action spaces. The different levels of attention include no attention, peripheral attention, central attention and agent response. The benefit of considering a human agent's peripheral attention is due to the fact that human agents could perceive a lot of information through peripheral attention without disturbing their limited central attention capabilities. Such a pragmatic view of a human agent's attention provides the possibilities of seamlessly moving virtual objects between the center and periphery of human attention [69].

In a smart home context (one potential example of a post-desktop computing environment) to be described in section X.5, virtual objects could be positioned in a human agent's situative spaces using a mixed-initiative approach [27]. Human agents could directly manipulate virtual objects and position them within the situative spaces and to complement it, virtual objects could be automatically positioned within the situative spaces through artificial software agents like the interaction manager to be described in the next section. Physical objects usually exist in a physical environment and are usually not required to be artificially positioned in the situative space with the exception of having artificial physical agents like robots that might bring and position physical objects. Instead, the physical objects are usually sensed to determine their position in the situative spaces which are used for managing a human agent's interaction with computing applications that make use of the physical objects and their context. Human attention to an extent is dependent on the objects in the situative spaces. Some object might grab human attention, while other might reside peacefully in the periphery of human attention. An attempt to pragmatically associate human attention to the situative spaces is as follows (bottom-up approach). The association is more of a rule of thumb for cognitive systems.

In Fig. 1, the various levels of human attention are mapped to different sets and spaces within the situative space model.

- *Objects in the world space,* but outside the perception and action space demand no attention.
- *Objects in the perception space,* but outside the recognizable and examinable set demand peripheral attention.
- *Objects in the recognizable and examinable set* demand central attention.
- *Objects in the action space,* but outside the selected and manipulated set demand agent response, i.e. to manipulate them (physically and/or virtually).
- *Objects in the selected and manipulated set* are already being acted upon and is assumed to demand central attention.

Human attention is also dependent on human intentions. An object that is currently associated with the human agent's current activity has a better chance of occupying a human agent's peripheral attention than objects that are outside the activity context. Similarly, objects associated with the human agent's current action have a better chance of occupying a human agent's central attention. Refer to Fig. 1. An attempt to pragmatically associate human intention to the situative spaces is as follows (top-down approach):

- *Objects outside a human agent's activity and action context* demand no attention and the object is pushed outside the perception space; and no agent response is required and the object is pushed outside the action space.
- *Objects within a human agent's activity context, but outside the action context* demand peripheral attention and the object is pushed inside the perception space, but outside the recognizable and examinable set; and agent response is

required and the object is positioned in the action space but outside the selected and the manipulated set.

- *Objects within the activity and the action context* demand central attention and the object is pushed inside the recognizable and/or the examinable sets; and agent response is required and the object is positioned in the action space but outside the selected and the manipulated set.

5 The Easy ADL Home

The easy ADL home is a smart home that is built to explore the egocentric interaction paradigm. The first prototype of the easy ADL home [62] was simulated in an immersive virtual-reality (VR) setup where sensors, actuators, interactive devices, physical objects and virtual objects were simulated to focus on the conceptual and the algorithmic parts of the easy ADL home. The second prototype of the easy ADL home [63] was physically constructed as a living laboratory to focus on the technological aspects and to explore the EIP within everyday physical and social context. The easy ADL home is a 54 m^2 apartment intended for providing ADL (Activities of Daily Living) support for single occupants, thereby keeping the exploration setup simple and avoids the challenges inherent in multi-occupant environments.

The easy ADL home is based on agent-centric (human agent) architecture that considers both a top-down and a bottom-up approach as shown in Fig. 2. A human agent perceives information in the perception space, and acts in the action space that is bound to physical, social and cultural context. The attention filter is a concept of a filter indicating that human agents use different levels of attention within

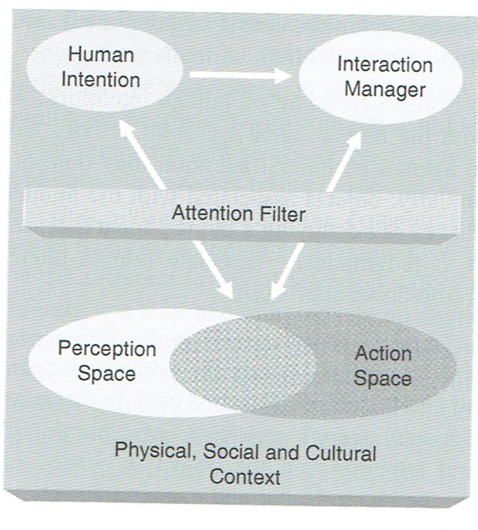


Fig. 2 An agent-centric architecture for facilitating egocentric interaction in the easy ADL home

the perception and action spaces. A human agent's intention and their situation are assumed to drive them in performing activities within the perception and the action spaces. The interaction manager is an artificial agent responsible for efficient management of a human agent's interaction with computing applications. A human agent's perception, action, intention and attention capabilities (and limitations) are considered by the interaction manager in presenting and providing access to virtual objects within the human agent's situative spaces. We have addressed three challenges that are important for facilitating egocentric interaction within the easy ADL home namely: situative space tracking, activity and action recognition, and interaction management.

- *Situative Space Tracking:* Tracking physical objects, mediators and virtual objects in a human agent's situative spaces is achieved using several technologies. A wireless-LAN signal-strength-based situative space tracking system is used to capture physical objects and mediators (that provide access to virtual objects) within the perception space, action space and recognizable set [59]. A wireless sensor network using ZigBee protocol is used to track the physical objects within the selected and manipulated sets [61], while 3-axis accelerometer-based hand gesture recognition and speech recognition is used to track the virtual objects within the selected and manipulated sets [52].
- *Activity and Action Recognition:* Tracking human activities and actions using the perception space, action space and selected set as input information channels is described in [64], while using the selected and manipulated sets as input information channel for activity and action recognition is described in [60]. Both the recognition systems use parallel Hidden-Markov Models to model human activities. Everyday physical activities like *preparing pasta*, *doing the dishes*, etc. are recognized in the immersive virtual-reality simulated easy ADL home.
- *Interaction Management:* A mixed-initiative approach is taken in managing a human agent's interaction with computing applications. Based on tracking the situative spaces and recognizing the human agent's activity and action context, the interaction manager attempts to decide on if, when, where and how to present virtual objects within the situative spaces as requested by the computing applications or the human agent. A rule-based interaction manager is developed following the principles and assumptions of egocentric interaction. A detailed account of the session rules are beyond the scope of this chapter, however a typical example of the rules include pragmatically associating human intention to the situative spaces described in section X.4.3. The interaction manager is also responsible for handling conflicts between human initiated and computing application initiated sessions.

6 Discussion and Conclusion

The proposed egocentric interaction paradigm (EIP) is centered on a human agent and in particular their cognition. This introduces two types of challenges: 1) to build theoretical models of a human agent and their cognition; and 2) to develop technologies that can accurately sense, recognize and model aspects of a human agent and their cognition.

The EIP is intended for everyday environments that could be occupied by multiple human agents. The proposed situative space model should be extended along the following dimensions: a) to consider social and cultural aspects of a human agent; b) to be a part of a network of situative spaces formed by multiple occupants of an everyday environment. Such a network might include situative spaces that are shared among multiple human agents useful in facilitating collaborative activities; and c) to include spatial relationship among objects within the situative spaces as it plays an important role in human behavior. The assumptions upon which the egocentric interaction paradigm is developed might not be valid in some situations, even though they are useful for most situations. For instance, the proximity principle that is used in operationalizing the situative space model might conflict in situations where a far-away object might be more important than the ones that are closer to the human agent. Such exceptions should be explored and handled in the future.

The EIP is dependent on technical and technological advancements. Accurate sensing of the situative spaces in everyday environments subject to noise, sensor failures and dynamic changes to the situative spaces is an important challenge to address. Modeling and recognizing activities in the real world is another important challenge to address. Human activities are usually performed with variations; they change over time, and are often interleaved with other parallel activities. Human activities might take place in a mobile context, and get constantly interrupted and resumed without noticeable events that signify it. Also, multiple human agents might perform group activities, or attempt to share resources in performing their individual activities resulting in conflicts. Accurate and fine-grained activity recognition at the action level is important for facilitating egocentric interaction. Presenting and providing access to virtual objects within the situative spaces is dependent on the available mediators within the respective situative spaces. Different environments might have varied riches in the quantity and quality of available mediators and the interaction manager should handle even the worst situations where limited mediators are available for enabling a human agent's interaction with computing applications. The current rule-based approach is just one solution, while better and more adaptive approaches could be explored in the future. Further work is required not only from a technological perspective, but also in designing experiments so as to ecologically validate the sensor technologies, the recognition algorithms, and the concepts that form a theoretical base for the emerging egocentric interaction paradigm.

To conclude, the egocentric interaction paradigm has its foundations on a human agent and their cognition, and is expected to withstand (and benefit out of) the technological advancements in evolving as a suitable interaction paradigm for post-desktop computing environments.

References

[1] Anderson, J.R.: Rules of the mind. Erlbaum, Hillsdale (1993)
[2] Bardram, J.: From Desktop Task Management to Ubiquitous Activity-Based Computing. In: Integrated Digital Work Environments: Beyond the Desktop Metaphor, pp. 49–78. MIT Press (2007)
[3] Bell, C.G., Newell, A.: Computer Structures: Readings and Examples. McGraw-Hill, New York (1971)
[4] Blakemore, S.-J., Decety, J.: From the perception of action to the understanding of intention. Nature Reviews, Neuroscience 2, 561–567 (2001)
[5] Brass, M., Schmitt, R.M., Spengler, S., Gergely, G.: Investigating action understanding: Inferential processes versus action simulation. Curr. Biol. 17, 2117–2121 (2007)
[6] Brown, J.S., Collins, A., Duguid, P.: Situated cognition and the culture of learning. Educational Researcher 18(1), 32–41 (1989)
[7] Burdea, G., Coffet, P.: Virtual Reality Technology, 2nd edn. Wiley-IEEE Press (2003)
[8] Byrne, M.D., Anderson, J.R.: Perception and action. In: Anderson, J.R., Lebiere, C. (eds.) The Atomic Components of Thought, pp. 167–200. Erlbaum, Hillsdale (1998)
[9] Card, S.K., Moran, T.P., Newell, A.: The psychology of human-computer interaction. Erlbaum, Hillsdale (1983)
[10] Clark, A.: Being There: Putting Brain, Body, and World Together Again. MIT Press, Cambridge (1997)
[11] Clark, A., Chalmers, D.: The Extended Mind. Analysis 58, 7–19 (1997)
[12] Dourish, P.: Where The Action Is: The Foundations of Embodied Interaction. MIT Press, Cambridge (2001)
[13] Duncan, J., Humphreys, G.W.: Visual search and stimulus similarity. Psychological Review 96, 433–458 (1989)
[14] Ehn, P.: Work-oriented design of computer artifacts. Lawrence Erlbaum Associates, Hillsdale (1990)
[15] Elsner, B., Hommel, B., Mentschel, C., Drzezga, A., Prinz, W., Conrad, B., Siebner, H.R.: Linking actions and their perceivable consequences in the human brain. Neuroimage 17, 364–372 (2002)
[16] Fagioli, S., Hommel, B., Schubotz, R.I.: Intentional control of attention: Action planning primes action-related stimulus dimensions. Psychological Research 71 (2005)
[17] Fasel, B., Luettin, J.: Automatic facial expression analysis: A survey. Patt. Recogn. 36, 259–275 (2003)
[18] Gallese, V., Goldman, A.: Mirror neurons and the simulation theory of mind-reading. Trends Cogn. Sci. 2, 493–501 (1998)
[19] Gershenfeld, N., Krikorian, R., Cohen, D.: The Internet of Things. Scientific American 291(4), 76–81 (2004)
[20] Gibson, J.J.: The ecological approach to visual perception. Houghton Mifflin, Boston (1979)
[21] Emile, G.J.: Marshalling passions in training and education. In: Proceedings of the Fourth International Conference on Autonomous Agents, Barcelona, Spain, pp. 325–332 (2000)
[22] Haazebroek, P., Hommel, B.: Towards a Computational Model of Perception and Action in Human Computer Interaction. In: Duffy, V.G. (ed.) ICDHM 2009. LNCS, vol. 5620, pp. 247–256. Springer, Heidelberg (2009)

[23] Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. ACM Trans. Computer-Human Interaction 7(2) (2000)

[24] Hommel, B., Elsner, B.: Acquisition, representation, and control of action. In: Morsella, E., Bargh, J.A., Gollwitzer, P.M. (eds.) Oxford Handbook of Human Action, pp. 371–398. Oxford University Press, New York (2009)

[25] Hommel, B., Muesseler, J., Aschersleben, G., Prinz, W.: The theory of event coding (TEC): A framework for perception and action planning. Behavioral and Brain Sciences 24, 849–937 (2001)

[26] Hopfinger, J.B., Buonocore, M.H., Mangun, G.R.: The neural mechanisms of top-down attentional control. Nature Neuroscience 3(3), 284–291 (2000)

[27] Horvitz, E.: Principles of mixed-initiative user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1999, pp. 159–166. ACM, New York (1999)

[28] Hudlicka, E.: To feel or not to feel: The role of affect in human-computer interaction. International Journal of Human-Computer Interaction 59, 1–32 (2003)

[29] Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 234–241 (1997)

[30] Jacob, R., Girouard, A., Hirshfield, L., Horn, M., Shaer, O., Solovey, E., Zigelbaum, J.: Reality-based interaction: a framework for post-WIMP interfaces. In: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, pp. 201–210 (2008)

[31] Jaimes, A., Sebe, N.: Multimodal human-computer interaction: A sur-vey. Comput. Vis. Image Underst. 108(1-2), 116–134 (2007)

[32] James, W.: The principles of psychology. Dover Publications, New York (1890)

[33] Janlert, L.-E.: Contextual strategies – notes for a theory of context. Technical report UMINF 02.23, Umeå University (2003); ISSN-0348-0542

[34] Keysers, C., Gazzola, V.: Integrating simulation and theory of mind: From self to social cognition. Trends Cogn. Sci. 11, 194–196 (2007)

[35] Kieras, D.E., Meyer, D.E.: An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction 12, 391–438 (1997)

[36] Kirsh, D., Maglio, P.: On Distinguishing Epistemic from Pragmatic Ac-tion. Cognitive Science 18, 513–549 (1994)

[37] Kisacanin, B., Pavlovic, V., Huang, T.S. (eds.): Real-Time Vision for Human-Computer Interaction. Springer, New York (2005)

[38] Kotovsky, L., Gentner, D.: Comparison and categorization in the development of relational similarity. Child Development 67, 2797–2822 (1996)

[39] Kuutti, K.: Activity Theory as a potential framework for human-computer interaction research. In: Nardi, B. (ed.) Context and Consciousness. ch. 2, pp. 17–44 (1995)

[40] Langley, P., Choi, D.: A unified cognitive architecture for physical agents. In: Proceedings of the Twenty-First National Conference on Artificial Intelligence (2006)

[41] Levy, P.: Collective Intelligence: Mankind's Emerging World in Cyber-space. Perseus Books (1997)

[42] Lindsay, P., Norman, D.: Human Information Processing: An Introduction to Psychology (1977)

[43] Mack, A., Tang, B., Tuma, R., Kahn, S., Rock, I.: Perceptual organization and attention. Cognitive Psychology 24 (1992)

4] Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented reality: A class of displays on the reality-virtuality continuum. In: Proc. SPIE Conf. Telemanipulator and Telepresence Technologies, vol. 2351(34), pp. 282–292 (1994)
5] Moore, C.M., Egeth, H.: Perception without Attention: Evidence of Grouping under conditions of Inattention. Journal of Experimental Psychology: Human Perception and Performance 23(2), 339–352 (1997)
6] Nagel, T.: The view from nowhere. Oxford Univ. Press, New York (1986)
7] Newell, A.: Unified theories of cognition. Harvard University Press, Cambridge (1990)
8] Norman, D., Draper, S. (eds.): User centered system design. Erlbaum, Hillsdale (1986)
9] Oudeyer, P.Y.: The production and recognition of emotions in speech: Features and algorithms. Int. J. of Human-Computer Studies 59(1-2), 157–183 (2003)
0] Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision Research 42(1), 107–123 (2002)
51] 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 (2003)
52] 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 (June 2010)
53] Schmidt, A.: Implicit human computer interaction through context. Personal and Ubiquitous Computing 4(2-3), 191–199 (2000)
54] Schubotz, R.I., von Cramon, D.Y.: Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: An fMRI study. Neuroimage 15, 787–796 (2002)
55] Selkar, T., Burlesson, W.: Context-aware design and interaction in com-puter systems. IBM Syst. J. 39(3-4), 880–891 (2000)
[56] Shneiderman, B.: Direct manipulation: a step beyond programming languages. Computer 16, 57–69 (1983)
[57] Simon, J., Rudell, A.: Auditory s-r compatibility: The effect of an irrelevant cue on information processing. Journal of Applied Psychology 51, 300–304 (1967)
[58] Sloman, A.: Varieties of affect and the CogAff architecture schema. In: Proceedings of the AISB 2001 Symposium on Emotion, Cognition, and Affective Computing, York, UK (2001)
[59] Surie, D., Jäckel, F., Janlert, L.-E., Pederson, T.: Situative Space Tracking within Smart Environments. In: 6th International Conference on Intelligent Environments, Malaysia (July 2010)
[60] Surie, D., Lagriffoul, F., Pederson, T., Sjölie, D.: Activity Recognition Based on Intra and Extra Manipulation of Everyday Objects. In: Ichikawa, H., Cho, W.-D., Chen, Y., Youn, H.Y. (eds.) UCS 2007. LNCS, vol. 4836, pp. 196–210. Springer, Heidelberg (2007)
[61] Surie, D., Laguionie, O., Pederson, T.: Wireless Sensor Networking of Eve-ryday Objects in a Smart Home Environment. In: Proceedings of the 4th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Sydney, Australia (2008)

[62] Surie, D., Pederson, T.: An Activity-Centered Wearable Computing Infrastructure for Intelligent Environment Applications. In: Kuo, T.-W., Sha, E., Guo, M., Yang, L.T., Shao, Z. (eds.) EUC 2007. LNCS, vol. 4808, pp. 456–465. Springer, Heidelberg (2007)

[63] 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 2(3), 287–310 (2010)

[64] Surie, D., Pederson, T., Lagriffoul, F., Janlert, L.-E., Sjölie, D.: Activity Recognition Using an Egocentric Perspective of Everyday Objects. In: Indulska, J., Ma, J., Yang, L.T., Ungerer, T., Cao, J. (eds.) UIC 2007. LNCS, vol. 4611, pp. 246–257. Springer, Heidelberg (2007)

[65] Thorndike, E.L.: The law of effect. American Journal of Psychology 39, 212–222 (1927)

[66] Turk, M., Robertson, G.: Perceptual User Interfaces. Communications of the ACM (2000)

[67] Vertegaal, R.: Attentive user interfaces. Special Issue of Communica-tions of the ACM 46(3) (2003)

[68] Weiser, M.: The Computer for the 21st Century. Scientific American, 933–940 (1991)

[69] Weiser, M., Brown, J.S.: The coming age of calm technology. In: Beyond Calculation, Copernicus, New York, NY, USA, pp. 75–85 (1997)

[70] Wilson, M.: Six views of embodied cognition. Psychonomic Bulletin and Review 9, 625–636 (2002)

[71] Yarbus, A.: Eye Movements and Vision. Plenum Press, New York (1967)